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

Dissertation

Zur Erlangung des Grades

Doktor der Naturwissenschaften

Am Fachbereich Biologie

Der Johannes Gutenberg-Universität Mainz

Eleftheria Parasyraki

geboren am 05.07.1994 in Amarousio

Mainz, 2024

Dekan:

1. Berichterstatter:

2. Berichterstatterin:

Tag der mündlichen Prüfung:

Summary

Active DNA demethylation proceeds by the sequential oxidation of 5mC to 5hmC, 5fC and 5caC by TET proteins, followed by TDG-mediated removal of 5fC and 5caC to restore unmodified cytosine. While 5mC is an established epigenetic mark with known gene repression function, whether its oxidative derivatives have an instructive, physiologically relevant role of their own, apart from being DNA demethylation intermediates, remains unclear. Previous work from this lab showed that 5fC forms transient chromocenters at the perinucleolar compartment of *Xenopus* embryos during zygotic genome activation (ZGA). Moreover, 5fC was highly enriched on Pol III target genes at ZGA, notably at oocyte-type tandem arrayed *tRNA* genes. In this thesis, I address the question of what is the role of 5fC during *Xenopus* ZGA. I used methylation-assisted bisulfite sequencing to detect 5fC on selected *tRNA* genes at ZGA with base resolution. I found that 5fC marks all CpGs in and around *tRNA-iMet* and *tRNA-Gly* at ZGA at comparable levels. To erase global 5fC from early *Xenopus* embryos, I overexpressed human Tdg mRNA, which led to delayed embryonic development and lethality during gastrulation. Pol III chromatin immunoprecipitation in control and Tdg overexpressing embryos revealed that 5fC is required for Pol III binding at ZGA. By manipulating Tet and Tdg enzymes, I showed that 5fC is required for *tRNA* transcription during ZGA. To complement and extend this finding, I established a synthetic *tRNA-iMet* transgene assay, which showed that 5fC specifically promotes *tRNA* transcription during ZGA *in vivo*. Finally, efforts to identify a 5fC reader did not lead to the discovery of the mechanism by which 5fC exerts its gene activating role, which remains unresolved. Together, these results establish 5fC as an activating epigenetic mark for Pol III during zygotic reprogramming.

Zusamenfassung

Aktive DNA-Demethylierung erfolgt durch die sequentielle Oxidation von 5mC zu 5hmC, 5fC und 5caC durch TET Proteine, gefolgt von der TDG-vermittelten Entfernung von 5fC und 5caC zur Wiederherstellung von unverändertem Cytosin. Während 5mC eine etablierte epigenetische Modifikation mit bekannter Genunterdrückungsfunktion ist, bleibt unklar, ob seine oxidativen Derivate ebenfalls eine physiologisch relevante Rolle spielen. Unpublizierte Arbeiten dieses Labors haben gezeigt, dass 5fC während der zygotischen Genomaktivierung (ZGA) vorübergehende Chromozentren im perinukleolaren Kompartiment von *Xenopus* Embryonen bildet. Darüber hinaus war 5fC bei der ZGA stark für Pol III-Zielgene angereichert, insbesondere für tandem repeat *tRNA* Gene. In dieser Arbeit habe ich methylierungsgestützte Bisulfit-Sequenzierung verwendet, um 5fC auf ausgewählten *tRNA*-Genen bei ZGA mit Basenauflösung nachzuweisen. Ich fand heraus, dass 5fC alle CpGs in und um *tRNA-iMet* und *tRNA-Gly* bei ZGA markiert. Um globales 5fC aus frühen *Xenopus* Embryonen zu entfernen, habe ich menschliche Tdg mRNA überexprimiert, was zu einer verzögerten Embryonalentwicklung und Letalität während der Gastrulation führte. Pol III-Chromatin-Immunpräzipitation in Kontroll- und Tdg-überexprimierenden Embryonen ergab, dass 5fC für die Pol III-Bindung an ZGA erforderlich ist. Durch Manipulation der Tet- und Tdg-Enzyme konnte ich zeigen, dass 5fC für die *tRNA* Transkription während der ZGA erforderlich ist. Um dieses Ergebnis zu erhärten, habe ich einen synthetischen *tRNA-iMet* Transgen Assay durchgeführt, der zeigte, dass 5fC spezifisch die Transkription während der ZGA in vivo fördert. Bemühungen, einen physiologisch relevanten 5fC-Reader zu identifizieren, blieben hingegen erfolglos. Insgesamt belegen diese Ergebnisse, dass 5fC eine aktivierende epigenetische Modifikation für Pol III während der zygotischen Reprogrammierung darstellt.

Acknowledgments

Contents

List of Figures

List of Tables

Abbreviations

α-KG	α-ketoglutarate
2C	2-cell stage
5caC	5-carboxilcytosine
5fC	5-formylcyrosine
5hmC	5-hyroxylmethylcytosine
5hmU	5-hydroxylmethyluracil
5mC	5-methylcytosine
A	adenine
AP	abasic site, apurinic, apyrimidinic
AP	animal pole
BAF	Brg1/Brm associated factor
BDP1	B double primer
BER	base excision repair
BRF1,2	B-related factor
BS	bisulfite
C	cytosine
CGI	CpG Island
Co	control
dC	deoxycytosine
dCTP	deoxycytosine triphosphate
DEG	differentially expressed genes
DHU	dihydrouracil
DIP-seq	DNA immunoprecipitation sequencing
dNTP	deoxynucleotide triphosphate
ds	double strand
E	embryonic day
ETC	extra-TFIIIC
FC	fold change
FITC	fluorescein isothiocyanate
FP	fluorescence polarization
FW	forward
G	guanosine
gDNA	genomic DNA
GO	gene ontology
gRNA	guide RNA
h	hour/hours
HCD	higher-energy collisional dissociation
hCG	human chorionic gonadotropin
HF	high fidelity
hPSC	human pluripotent stem cell
IF	immunofluorescence
IVF	*in vitro* fertilization

KD	knockdown
KO	knockout
LB	lysogeny broth
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LFQ	label-free quantification
LINE	long interspersed nuclear elements
LTR	long terminal repeats
MAB	methylation assisted bisulfite
MBD	methyl-CpG binding domain
mESCs	mouse embryonic stem cells
min	minute
MLL	myeloid/lymphoid leukemia
Mo	Morpholino
MRD	maternal RNA degradation
MRP	mitochondrial RNA processing
MS	mass spectrometer
MZT	maternal-to-zygotic transition
NMR	nuclear magnetic resonance
NSC	neural stem cell
nt	nucleotide
NuRD	nucleosome remodeling and deacetylase complex
O/N	overnight
OE	overexpression
OGT	O-linked N-acetylglucoasamine transferase
oxC	oxidative methylcytosine derivatives
PAGE	polyacrylamide gel electrophoresis
PGC	primordial gem cells
PN	pronucleus
PNC	perinucleolar compartment
PNK	polynucleotide kinase
Pol I	RNA polymerase I
Pol II	RNA polymerase II
Pol III	RNA polymerase III
PSI	pounds per square inch
PTM	post-translational modification
rpm	rounds per minute
rRNA	ribosomal RNA
RT	room temperature or reverse transcription
RV	reverse
SD	standard deviation
SDS	sodium dodecyl sulfate
SE	strand exchange
sec	second

seq	sequencing
SINE	short interspersed nuclear elements
SMN	survival of motor neurons
SNP	single nucleotide polymorphism
snRNA	small nuclear RNA
ss	single strand
st	stage
SWI/SNF	switch/sucrose non-fermentable
T	thymine
TAD	topologically associated domain
TBP	TATA-box binding protein
TDG	thymine DNA glycosylase
TET	ten-eleven translocation
TF	transcription factor
TKD	triple knockdown
TKO	triple knockout
Tm	melting temperature
tRNA	transfer RNA
TSS	transcription start site
U	uracil
VP	vegetal pole
WB	western blot
WGBS	whole genome bisulfite sequencing
WGS	whole genome sequencing
wt	wild type
X.	*Xenopus*

1. Introduction

1.1. DNA methylation & demethylation

1.1.1. 5-methylcytosine (5mC) is a repressive mark

In DNA cytosine can be methylated at its 5th carbon to produce 5-methylcytosine (5mC). DNA methylation was first described in bacteria in 1925 (Johnson & Coghill, 1925). More than 20 years later and shortly before the discovery of the DNA double helix and the Watson-Crick base pairing (Watson & Crick, 1953), 5mC was re-discovered in eukaryotic DNA from mammals, insects and plants (Hotchkiss, 1948; Wyatt, 1950, 1951). Still, some of the most widely used animal models, namely *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Caenorhabditis elegans*, do not have DNA methylation.

Cytosine methylation occurs mainly in CpG dinucleotides, while CpHpG and CpHpH, (H= A, C, or T) methylation has also been described though at lower levels (Doskočil & Šorm, 1962; Gruenbaum et al., 1981; Lister et al., 2009; Sinsheimer, 1954). In mammals, CpG methylation is a frequent event, as 70-80% of all CpGs in somatic tissues is methylated (Li & Zhang, 2014). DNA methylation is mostly symmetric, as virtually all methylated CpGs are methylated on both strands (Lister et al., 2009). High levels of DNA methylation are found on repetitive elements, including transposable elements and satellite repeats, as well as gene bodies and exons. An exception to CpG methylation are CpG islands, which are mostly located on gene promoters and 5' gene regions and are widely unmethylated (Bird et al., 1985).

DNA methylation is established by the *de novo* DNA methyltransferases DNMT3a and DNMT3b (Okano et al., 1998, 1999). These enzymes can methylate cytosine in any context, with highest preference for CpG, followed by CpA (Aoki et al., 2001; Gowher & Jeltsch, 2001; Ramsahoye et al., 2000). Symmetric CpG methylation is maintained upon DNA replication by the maintenance methyltransferase DNMT1 (Bestor et al., 1988), in coordination with UHRF1, which binds hemimethylated CpGs at the replication fork and recruits DNMT1 (Bostick et al., 2007). Other members of the DNMT family have cell-type specific functions, as for example DNMT3L that is required for *de novo* methylation in the germline (Bourc'his et al., 2001; Bourc'his & Bestor, 2004; Ooi et al., 2007).

In bacteria, where it was first discovered, DNA methylation renders a protective mechanism against viral infection: bacterial restriction enzymes digest bacteriophage DNA in a sequence specific manner, while the same sequences on bacterial DNA are protected from digestion by DNA methylation (Arber, 1965; Gold et al., 1963). In eukaryotic DNA, 5mC is a repressive mark:

methylation of gene promoters results in transcriptional silencing (Mohn et al., 2008; Weber et al., 2007). Comparison of the methylation status of differentially expressed genes across different human cell types revealed a negative correlation between methylation levels and gene expression (Kuo et al., 1979; McGhee & Ginder, 1979). In addition, accessible chromatin bound by transcription factors is lowly methylated, if at all (Stadler et al., 2011). In contrast to gene promoters, DNA methylation at the gene body associates with induced gene expression (Aran et al., 2011; Ball et al., 2009; Hellman & Chess, 2007) and actively transcribed genes are enriched for 5mC at their gene bodies (Lister et al., 2009). Intragenic 5mC has also been proposed to promote transcriptional fidelity by preventing spurious transcription (Neri et al., 2017).

The repressive role of 5mC in gene expression is mediated by a complex interplay between 5mC readers, transcription factors (TFs) and histone modifications. 5mC can inhibit transcriptional activation by impeding TF binding to CGI gene promoters, as some TFs are sensitive to CpG methylation (Domcke et al., 2015; Kim et al., 2003; Watt & Molloy, 1988; Yin et al., 2017). Notably, CTCF, an insulator that regulates long distance enhancer-promoter interactions, is DNA methylation sensitive (Phillips & Corces, 2009). Furthermore, several 5mC reader proteins recognize and bind 5mC. Among them, methyl-CpG binding domain (MBD) proteins (Hendrich & Bird, 1998), like MeCP1 (Boyes & Bird, 1991; Meehan et al., 1989) and MeCP2 (Lewis et al., 1992), interact with nucleosome remodeling complexes and histone deacetylases to induce transcriptional silencing (Nan et al., 1998; Ng et al., 1999). Similar to MBDs, zinc-finger proteins (Kaiso, ZBTB4 and ZBTB38) bind 5mC and repress gene expression (Filion et al., 2006; Prokhortchouk et al., 2001; Yoon et al., 2003). A cooperative interplay between H3K9 histone methyltransferases, H3K9me readers and DNMTs mediates heterochromatin formation and maintenance (Chang et al., 2011; Estève et al., 2006; Lehnertz et al., 2003; Li et al., 2006). Conversely, DNA methylation and the activating histone modification H3K4me are mutually exclusive at CGIs (Li & Zhang, 2014). Collectively, the 5mC mark triggers gene silencing and is in turn summoned to silenced loci to maintain and propagate repression.

By employing a combination of the above-mentioned mechanisms, DNA methylation controls gene expression in physiological processes like embryonic development. During cellular differentiation and cell lineage specification, DNA methylation switches off cell-type irrelevant genes (Davies et al., 2012; Laurent et al., 2010; Nagae et al., 2011). DNA methylation is also involved in establishing and maintaining X chromosome inactivation (Lessing & Lee, 2013) and it consists the main defense mechanism against transposable element activity and its genotoxic effects (reviewed in Deniz et al., 2019; Jansz, 2019). Genomic imprinting, that is the monoallelic expression of certain genes based on their parental origin, is also mediated by DNA methylation

(Li et al., 1993). Moreover, CpG methylation functions as a marker of ageing. While age advances, the genome undergoes global DNA hypomethylation but local hypermethylation. Epigenetic clocks have been designed to predict the biological age of organisms, organs or tissues, based on the methylation profile of selected CpGs (Horvath & Raj, 2018). Similar to ageing, DNA methylation pattern changes are observed in cancer. More specifically, cancers are associated with hypomethylation of repetitive elements and oncogenes and *de novo* methylation of CGIs and tumor suppressor genes (reviewed in Klutstein et al., 2016; Skvortsova et al., 2019). It is thus evident that 5mC is central to fundamental cellular processes. Changes in this precisely fine-tuned system can have a range of implications, varying from natural organismal ageing to detrimental cancer development and progression.

As the first, and up to this point the only instructive DNA mark, 5mC sets the standards for a DNA modification to be considered functional: defined sequence context, writers and readers and a carefully validated role in biological processes.

1.1.2. Passive and active DNA demethylation

Despite 5mC being associated with stable gene silencing, DNA methylation is a dynamic process. The discovery that DNA methylation can be reversed through DNA demethylation, broadened the fundamental principles of 5mC. Both genome-wide and locus specific DNA demethylation have been observed. Genome-wide demethylation mediates genomic reprogramming during embryonic development (see section 1.1.7.), while locus specific demethylation occurs during somatic cell reprogramming and in pathology. Two distinct mechanisms of DNA demethylation have been proposed: *passive* and *active* demethylation. *Passive* DNA demethylation is the progressive loss of 5mC during DNA replication in the absence of maintenance DNA methyltransferase activity. *Active* DNA demethylation refers to the enzymatic conversion of 5mC to C, in a DNA replication independent manner.

Several pathways of active DNA demethylation in vertebrates have been proposed (reviewed at Schomacher & Niehrs, 2017; Wu & Zhang, 2010); however, the most prevalent one is TET/TDG mediated DNA demethylation. In this pathway, Ten-Eleven Translocation (TET) enzymes sequentially oxidize 5mC to 5-hydroxylmethylcytosine (5hmC) (Kriaucionis & Heintz, 2009; Tahiliani et al., 2009), 5-formylcytosine (5fC) (Ito et al., 2011) and 5-carboxylcytosine (5caC) (He et al., 2011; Zhang et al., 2012) (Figure 1.1). Thymine DNA Glycosylase (TDG) recognizes 5fC and 5caC and excises them, creating an abasic (AP) site that will be processed by the Base Excision Repair (BER) pathway to restore unmodified cytosine (Cortellino et al., 2011; He et al., 2011; Maiti & Drohat, 2011; Weber et al., 2016) (Figure 1.1). This mechanism is unique to animals,

while in plants 5mC is directly excised by DNA glycosylases (Zhu, 2009). Overall, active DNA demethylation in vertebrates requires the iterative oxidation of 5mC to a series of oxidized derivatives, collectively referred to as oxCs. The higher oxidation state products, 5fC and 5caC, can then be replaced by unmodified cytosine, reversing DNA methylation and its repressive signaling.

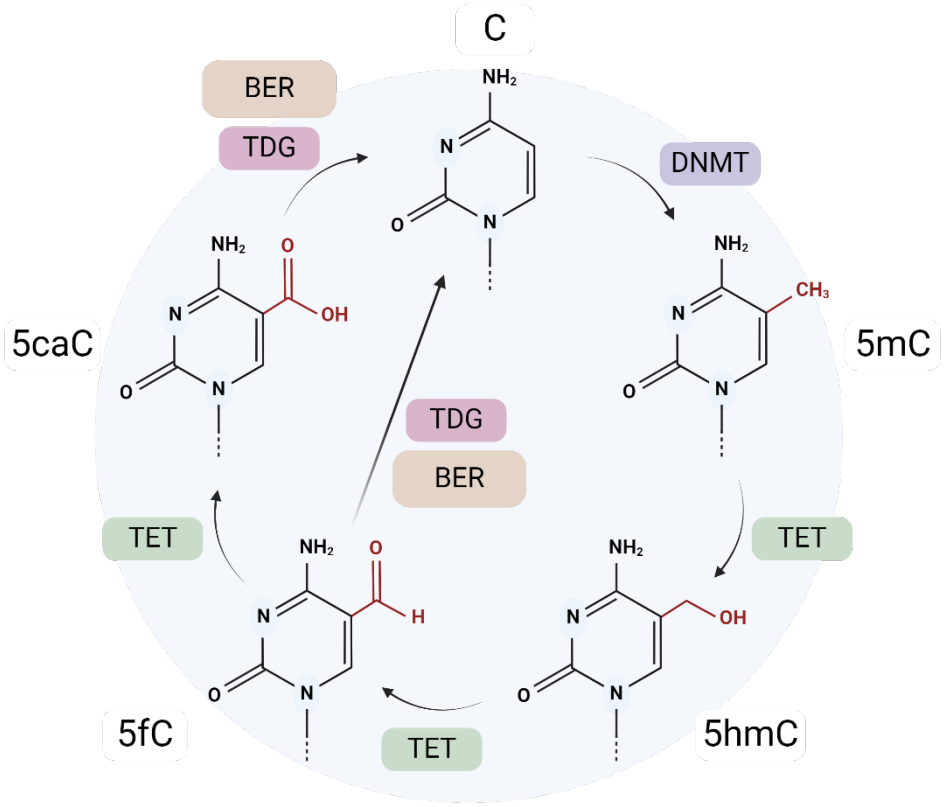


Figure 1.1. Active DNA demethylation pathway.
Deoxycytidine can be methylated at its 5th carbon by DNMTs to introduce 5mC. During active DNA demethylation, TET enzymes sequentially oxidize 5mC to 5hmC, 5fC and finally 5caC. 5fC and 5caC are targeted by TDG and removed through the BER pathway, to reconstitute unmodified cytosine. Adapted from Schomacher & Niehrs, 2017.

1.1.3. TET proteins

TET proteins are Fe^{2+} and α-ketoglutarate (α-KG) dependent dioxygenases. TET was initially described as a fusion with MLL in acute myeloid leukemia (Lorsbach et al., 2003; Ono et al., 2002) and an orthologue of JBP1 and -2, that oxidize T to 5hmU in *Trypanosoma brucei* (Borst & Sabatini, 2008; Cliffe et al., 2009; Yu et al., 2007), before it was associated with DNA demethylation (Tahiliani et al., 2009). Mammals and other vertebrates like zebrafish, have three

Tet ortholoques: Tet1, Tet2 and Tet3 (Almeida et al., 2012; Tahiliani et al., 2009). Among them, Tet1 and Tet3 have multiple isoforms expressed in different cell types (Akahori et al., 2015; Melamed et al., 2018). During mouse development, oocytes and zygotes express high levels of Tet3 and low levels of Tet1/2, while the opposite is true for preimplantation stages (Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011) (Figure 1.2). Accordingly, mouse embryonic stem cells (mESCs) mostly express Tet1 and Tet2 (Ficz et al., 2013; Ito et al., 2010; Koh et al., 2011; Lu et al., 2014). In *Xenopus*, Tet2 and Tet3 are expressed during early development (Xu et al., 2012), while a Tet1 orthologue is absent (Bogdanović et al., 2016; Xu et al., 2012). Despite different TET orthologue combinations being expressed in different species and developmental stages, they seem to have redundant roles in DNA demethylation.

TET genomic targeting and processivity determine DNA demethylation progression, oxC production and 5mC removal. In the genome, TET proteins bind transcription start sites (TSS), promoters, distal enhancers and gene bodies and their affinity for these genomic regions positively correlates with CpG density (Deplus et al., 2013; Hon et al., 2014; Huang et al., 2014; Vella et al., 2013; Williams et al., 2011; Wu et al., 2011; Xu et al., 2011). TET targeting is mediated by transcription factors such as NANOG, PRDM14, TEX10 and others, in a context-specific manner (Costa et al., 2013; Ding et al., 2015). Additionally, TET1 physically interacts with TDG (Weber et al., 2016) and TDG was suggested to be required for TET induced 5mC oxidation at CGIs (Iurlaro et al., 2016). TET-DNA binding does not involve the methyl group of 5mC, allowing the enzymes to bind and process different forms of modified cytosine (Hu et al., 2013). However, TETs do have a preferable substrate, in the order 5mC > 5hmC > 5fC. Kinetics analysis showed that 5mC to 5hmC conversion is 3-5 fold faster than that of 5hmC to 5fC or 5fC to 5caC *in vitro* (Hashimoto et al., 2014; Hu et al., 2015; Ito et al., 2011). To this end, the chemical processivity of TETs, which refers to the ability of the enzyme to catalyze the iterative oxidation of 5mC to 5caC without releasing the substrate, remains unclear as different studies had conflicting findings (Crawford et al., 2016; Tamanaha et al., 2016). Local chromatin environment and transcription factors are also likely to affect TET processivity, as 5mC, 5hmC, 5fC and 5caC have both overlapping and modification-specific genomic signatures, suggesting that 5mC is oxidized to different extend at different genomic regions (Shen et al., 2014; Z. Sun et al., 2015; Wu et al., 2014; Wu & Zhang, 2017). Collectively, TET substrate specificity and the chromatin environment of each methyl-CpG likely determine the final product of TET oxidation.

Several labs generated TET knockout (KO) animals and mESCs to study the role of TETs and DNA demethylation in development. Tet triple-knockout (TKO) mESCs maintained pluripotency marker expression and self-renewal but showed reduced mesodermal and endodermal marker

expression in embryonic bodies (Dawlaty et al., 2014; Lu et al., 2014). Additionally, Tet TKO mESCs showed increased expression of 2-cell stage (2C) like genes, suggesting that Tet proteins control the switch from totipotency to cell specification (Lu et al., 2014). Tet TKO embryos produced by conditional germline TKO mice developed all three germ layers but did not survive past gastrulation (Dai et al., 2016). However, a more recent study used morpholinos (Mo) to acutely inhibit all three Tet proteins in mouse oocytes and found that Tet deficient embryos arrest at 2C (Arand et al., 2022). This conflicting finding suggests that germline Tet deletion may allow for the development of compensation mechanisms, masking the role of Tets in early embryonic development. All in all, mESCs and mouse experiments confirm an important role of Tet proteins in pre- and post-implantation development, and highlight the significance of Tet function at 2-cell stage.

Apart from their role in DNA demethylation, TETs also have catalytic-independent roles in gene expression. Specifically, TET1 interacts with the SIN3A co-repressor complex to mediate transcriptional silencing (Williams et al., 2011), while TET2 can interact with O-linked N-acetylglucoasamine transferase (OGT) to promote histone modification and H3K4me3 deposition (Deplus et al., 2013). Furthermore, TET1 suppresses thermogenic gene expression in the adipose tissue by coordinating HDAC activity, in a DNA demethylation independent manner (Damal Villivalam et al., 2020). Notably, a recent study identified Tet1 as a global regulator of histone modifications and showed that Tet1-mediated transcriptional regulation in mESCs is mainly independent of its catalytic activity (Stolz et al., 2022). TET proteins also have a catalytic activity towards RNA, adding an additional level of TET mediated gene expression regulation (Fu et al., 2014; Huang et al., 2016; Huber et al., 2015; Delatte et al., 2016; Lan et al., 2020; Shen et al., 2018). Overall, TET proteins have diverse, DNA demethylation dependent and independent roles in gene expression. Thus, orthogonal experimental validation is required to differentiate between them.

Currently, oxidative 5mC derivatives produced by TET enzymes are mostly considered DNA demethylation byproducts rather than instructive DNA modifications. Furthermore, the fact that 5fC and 5caC are TDG targets, as well as the strikingly low levels of these modifications in the vertebrate genome, shaped the assumption that the highly oxidized cytosine bases are unlikely to have their own biological roles. Yet, a growing body of evidence suggest that these DNA modifications might have a function on their own, beyond being DNA demethylation intermediates.

1.1.4. 5-hydroxylmethylcytosine (5hmC)

5hmC is the first 5mC oxidative derivative. 5hmC levels greatly vary among different developmental stages and tissues, and it is most abundant in ESCs (0.7% of total deoxicytosine; dC) and the central nervous system, where it consists up to 0.7% and 1.2% dC, in mouse and human respectively (Globisch et al., 2010; Kriaucionis & Heintz, 2009; Tahiliani et al., 2009; Nestor et al., 2012; Ruzov et al., 2011; Song et al., 2011; Szwagierczak et al., 2010; Wagner et al., 2015). Unlike its precursor 5mC, 5hmC is non-symmetrically distributed between the two DNA strands (Bachman et al., 2015; Song et al., 2016). Apart from being further oxidized by TET, 5hmC can be passively diluted during DNA replication (Inoue & Zhang, 2011).

Genome-wide studies of 5hmC distribution correlate it with transcriptional activation (reviewed at Shen & Zhang, 2013; Shi et al., 2017). 5hmC is enriched at TSS of genes with poised promoters (H3K4me3 and H3K27me3 marked) (Pastor et al., 2011) and at active (H3K4me and H3K27ac) and poised (H3K4me) gene enhancers (Sun et al., 2015; Szulwach et al., 2011). Moreover, 5hmC enrichment at gene promoters and gene bodies positively correlates with increased transcription levels (Ficz et al., 2011; Nestor et al., 2012; Song et al., 2011; Wu et al., 2023). Conversely, another study found that 5hmC at intragenic regions correlated with active transcription, but 5hmC enrichment at gene promoters correlates with transcriptional repression in mESCs, suggesting that 5hmC can have opposing effects on transcription depending on its genomic localization (Wu et al., 2011). Overall, it is unclear whether 5hmC has an active role in promoting gene transcription or it simply offers transcriptional de-repression by reversing 5mC induced silencing.

Identifying 5hmC interactors could help delineate its function. The repressor protein MeCP2 has reduced affinity for 5hmC compared to 5mC *in vitro* (Khrapunov et al., 2014), suggesting that 5hmC might attain transcriptional activation by repelling silencing factors attracted by 5mC (Song & Pfeifer, 2016). On the other hand, *in vitro* experiments identified potential 5hmC readers, including DNA glycosylases (Mpg, Neil3) and helicases (Recq1), suggesting that 5hmC might have autonomous roles in gene expression regulation (Bai et al., 2021; Iurlaro et al., 2013; Spruijt et al., 2013). Additionally, the Nucleosome Remodeling and Deacetylase (NuRD) complex subunit Mbd3 has higher affinity for 5hmC than for 5mC *in vitro* (Yildirim et al., 2011). Mbd3 also seems to regulate global 5hmC levels and 5hmC-marked gene expression in mESCs, suggesting a complex interplay between 5hmC, Tets and nucleosome remodeling factors (Yildirim et al., 2011). Moreover, a recent study suggested that 5hmC marks gene bodies to prevent aberrant RNA polymerase II (Pol II) transcription (Wu et al., 2023). Apart from its potential roles in transcription

regulation, 5hmC can also attract DNA repair enzymes, as it specifically colocalizes with γH2AX, 53BP1 and RAD51 in HeLa cells (Kafer et al., 2016).

Despite the great interest of the scientific community, the effort to attribute a gene regulatory role to 5hmC has proven challenging: Since 5mC is a transcriptional repressor, it is difficult to distinguish between the presence of 5hmC and the absence of 5mC as the driving force behind the correlation of 5hmC with transcriptional activation. Additionally, to study the role of 5hmC, researchers often manipulate TET enzymes by TET inactivation or TET overexpression. However, apart from their role in DNA demethylation, TET proteins have catalytic activity-independent roles in transcription. Thus, attributing TET-associated effects to 5hmC is problematic and conclusions drawn by these studies need to be carefully accessed. The difficulties in defining the role of 5hmC in gene expression extend to its derivatives 5fC and 5caC, with the additional obstacle of the profoundly low levels of these modifications.

1.1.5. 5-formylcytosine (5fC)

5fC is the oxidative derivative of 5hmC during active DNA demethylation and, unlike 5hmC, 5fC can be excised by TDG (Ito et al., 2011; Maiti & Drohat, 2011). 5fC is a relatively rare DNA modification: it is up to 100-fold less abundant than 5hmC and it consists ~0.0006% of all dC) in mESCs (Bachman et al., 2015; Han et al., 2019). 5fC is detected in all mouse tissues and it is most abundant in the brain (Bachman et al., 2015; Iurlaro et al., 2016). It is present in human and mouse gametes and it is retained throughout embryonic development, with global 5fC levels to peak at the paternal pronucleus (PN) in mouse and human zygote (Gao et al., 2020; Zhu et al., 2017).

Regarding its genomic distribution, 5fC is enriched at poised and active enhancers and intragenic regions, particularly exons (Iurlaro et al., 2016; Song et al., 2013; Xia et al., 2015). Notably, Long Interspersed Nuclear Elements (LINEs) and Long Terminal Repeats (LTRs) are depleted for 5fC in mESCs but enriched in human embryos, especially at the zygote, showing a developmental stage-specific 5fC signature (Gao et al., 2020; Song et al., 2013). Additionally, 5fC is enriched at developmental enhancers in a tissue-specific manner, suggesting a role of 5fC in cell fate specification (Iurlaro et al., 2016). On a base level, in mESCs 5fC distribution is distinct from that of 5hmC and 5fC marked regions are more enriched for the activating marks H3K4me, H3K27ac, p300 as well as DNase I hypersensitivity regions than 5hmC regions (Song et al., 2013; Xia et al., 2015). Additionally, in mESCs 5fC-rich CGI promoters correspond to transcriptionally active genes and are enriched for the activating histone mark H3K4me3 (marks initiated promoters) and Pol II (Neri et al., 2015; Raiber et al., 2012). During mouse embryonic development, 5fC enrichment at

gene promoters precedes transcriptional activation of developmental genes, while 5fC-marked promoters in human embryos show transcriptional upregulation at the same developmental stage (Gao et al., 2020; Zhu et al., 2017). Overall, based on its genomic distribution and associated chromatin state, 5fC shows characteristics of an activating mark.

To gain insights into its mode of action, *in vitro* studies tested the effect of 5fC on DNA conformation and chromatin configuration. Crystal structure and nuclear magnetic resonance (NMR) experiments showed that 5fC destabilizes the DNA double helix, resulting in helical unwinding (Dubini et al., 2020; Jaisal et al., 2023; Raiber et al., 2015). These findings suggest that 5fC might regulate gene expression by altering the conformation of the DNA, thereby affecting DNA-protein recognition and interaction (Fu et al., 2019). However, another study reported no significant conformational changes on DNA due to 5fC (Hardwick et al., 2017). 5fC was also shown to form reversible covalent bonds with histones, to increase DNA flexibility and to induce nucleosome occupancy and mechanical stability *in vitro,* suggesting that 5fC might have a role in chromatin remodeling and nucleosome positioning (Ji et al., 2017; Li et al., 2017; Ngo et al., 2016; Raiber et al., 2018). In line with these findings, nucleosome-occupied regions are enriched for 5fC compared to nucleosome-depleted regions in human embryos (Gao et al., 2020). Collectively, 5fC alters DNA conformation and nucleosome organization *in vitro*, while it remains unclear whether this is also the case in the complex cell environment.

Other *in vitro* studies sought to characterize the role of 5fC in gene expression. Oligo pulldown experiments found more proteins to bind to 5fC than 5hmC (Iurlaro et al., 2013; Spruijt et al., 2013). Among the 5fC binders were DNA repair proteins as well as transcriptional repressors, such as SIN3A, MBD3, the putative Polycomb group protein L3MBTL2 and all NuRD complex components, implying that 5fC might act as a repressive mark. 5fC was also reported to impede Pol II transcription rate, substrate specificity and transcriptional fidelity *in vitro* (Kellinger et al., 2012; Kitsera et al., 2017; Wang et al., 2015; You et al., 2014). Importantly, the transcriptional repression by 5fC at the promoter of a reporter construct was BER-dependent, suggesting that DNA repair activation or the AP site, rather than 5fC, interfere with transcription (Müller et al., 2021). In sum, *in vitro* evidence support a suppressive role for 5fC in gene expression.

In summary, it appears likely that 5fC has a role in gene expression beyond its function as a DNA demethylation intermediate. However, the genomic distribution of 5fC *in vivo,* and *in vitro* efforts to identify its function yielded largely contradictory outcomes. It remains unclear whether 5fC is an effector DNA modification and if so, whether it mediates transcriptional activation or repression. A

clear causal role of 5fC in transcription regulation in a physiologically relevant context is yet to be identified.

1.1.6. 5-carboxylcytosine (5caC)

5caC represents the highest oxidation state of 5mC and it is the rarest of all oxCs, as it consists ~0.00008-0.0001% of total dC in mESCs (Ito et al., 2011; Ličytė et al., 2020; Wheldon et al., 2014), and it was not detected in most mouse tissues (Bachman et al., 2015). Due to its low abundance, characterizing 5caC and identifying its potential functions has been particularly challenging and currently there is little known about this DNA modification. In mESC genome, 5caC is enriched at poised (H3K4me marked) and active (H3K27ac and H3K4me marked) enhancers, satellite repeats and SINEs (Ličytė et al., 2020). Notably, despite sharing similar genomic signatures, there is <7% overlap between 5fC and 5caC sites (Lu et al., 2015). 5caC is also enriched at various TF binding sites, such as pluripotency factors Sox2, Nanog, Oct4, as well as CTCF and p300 sites in mESCs (Ličytė et al., 2020; Nanan et al., 2019). Some studies proposed that 5caC may be involved in cell differentiation. More specifically, 5caC transiently accumulates at cell-type specific enhancers during lineage specification of mouse neural stem cells (NSCs) and during differentiation of human pluripotent stem cells (hPSCs) towards hepatic endoderm (Lewis et al., 2017; Wheldon et al., 2014). Furthermore, 5caC is enriched at evolutionarily young, transcriptionally active LINE-1 elements in round spermatids but depleted during spermatid maturation, suggesting a role of 5caC in transposable element activation during male gametes maturation (Blythe et al., 2021).

In vitro studies identified candidate 5caC binders in an effort to delineate potential 5caC functions (Golla et al., 2014; Spruijt et al., 2013). Interestingly, there was a limited overlap between 5fC and 5caC binders, while both DNA modifications pulled down more interactors than 5hmC, as both the formyl- and the carboxyl- groups are chemically more reactive than the hydroxymethyl- group (Spruijt et al., 2013). Among the 5caC-specific readers were Dnmt1 and Swi/Snf chromatin remodeling complex subunits, such as BAF170 (Spruijt et al., 2013). Another potential 5caC reader is the mismatch repair complex (MMR), as 5caC was shown to activate DNA polymerase δ exonuclease activity and MMR *in vitro*, supporting that 5caC behaves as a DNA lesion (Shibutani et al., 2014). X-ray crystallography of Pol II – DNA complex showed that 5caC forms hydrogen bonds with Pol II, interfering with transcription by impeding nucleotide incorporation at the newly synthesized RNA (Konovalov et al., 2021; L. Wang et al., 2015). Additionally, 5caC inhibited Pol II transcription form a reporter construct in the absence of BER, while BER activity transiently re-activated gene expression, suggesting that 5caC/BER may act as a molecular switch of Pol II

transcriptional activity (Müller et al., 2021). Overall, 5caC has divergent genomic localization and interactors from 5fC. Similar to 5hmC and 5fC, the function of 5caC in the vertebrate genome is currently unknown.

1.1.7. DNA modifications in development

While 5mC is generally a stable DNA modification, there are two waves of global 5mC reprogramming in mammals: one at early embryonic development and another in primordial germ cells (PGCs) (Howlett & Reik, 1991; Kafri et al., 1992; Rougier et al., 1998; Sanford et al., 1987). DNA methylation reprogramming in preimplantation development is mediated by both passive demethylation, through DNA replications in the absence of DNMT1, and active TET/TDG dependent DNA demethylation. Methylation reprogramming in early development serves a dual function: First, 5mC erasure facilitates the transition from specialized gametes to totipotent zygote. Second, subsequent re-establishment of DNA methylation dictates the exit from pluripotency towards cell fate specification (Lee et al., 2014).

In mammals, the male and female gametes have different methylation levels: the sperm is highly methylated, while the oocyte is relatively hypomethylated (Howlett & Reik, 1991; Monk et al., 1987; Wang et al., 2014). Shortly after fertilization, the paternal genome undergoes active, almost complete DNA demethylation (Mayer et al., 2000; Oswald et al., 2000; Santos et al., 2002). DNA demethylation at the paternal pronucleus (PN) is mediated by TET3 and results in 5hmC and 5fC accumulation (Gao et al., 2020; Gu et al., 2011; Iqbal et al., 2011; C. Zhu et al., 2017; Wossidlo et al., 2011; Yan et al., 2023). Paternal genome demethylation is completed at 1-cell stage in mouse and 2-cell stage in human, consistent with overall slower human embryonic development (Guo et al., 2014; Zhu et al., 2018). The maternal genome mainly undergoes passive, replication-dependent DNA demethylation (Mayer et al., 2000; Rougier et al., 1998). Notably, there is also limited active DNA demethylation in the maternal PN (Guo et al., 2014; Shen et al., 2014; Wang et al., 2014), despite initially thought otherwise. The embryos undergo further 5mC removal until the blastocyst stage, after which DNA methylation is gradually re-established by *de novo* DNA methyltransferases DNMT3A and DNMT3B before implantation (Okano et al., 1998, 1999) (Figure 1.2). The only genomic regions that escape global demethylation during preimplantation development are sex-specific imprinted genes (Tremblay et al., 1997; Xie et al., 2012).

Consistent with the global DNA demethylation pattern, 5hmC and 5fC levels peak at human and mouse zygote, where 5hmC consists up to ~9% of total dC and 5fC up to 0.1% of all cytosines in CpG context (Gao et al., 2020; Yan et al., 2023; Zhu et al., 2017). Importantly, Tdg mRNA is not detected at 2-cell stage mouse embryos (Huang et al., 2017) and Tdg deficiency does not affect

DNA demethylation in the zygote (Guo et al., 2014; Santos et al., 2013). These findings suggest that (a) Tdg is not maternally provided and (b) oxC are passively diluted via DNA replication. Consistent with the general uncertainty about the functional roles of oxCs, it is currently unknown whether these DNA modifications have physiological roles in embryonic development. In case they do, global 5mC reprograming in early embryonic development would have an additional fundamental function, that of oxC establishment.

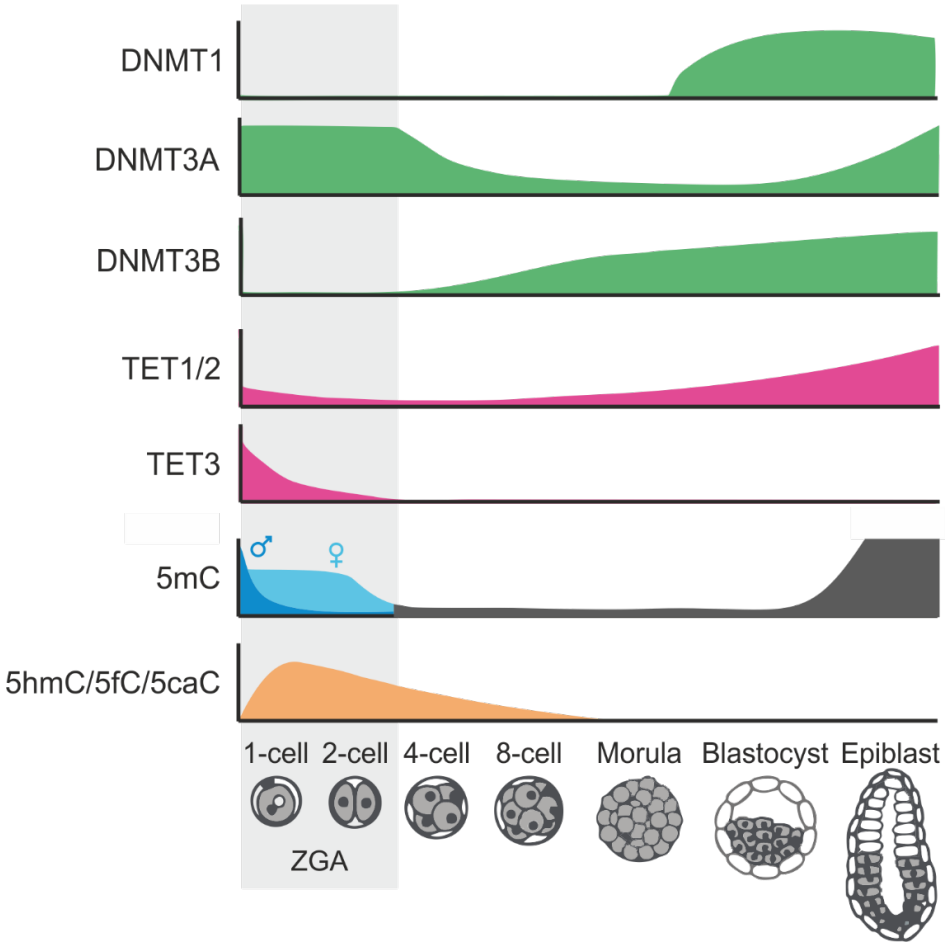


Figure 1.2. DNA methylation and demethylation components dynamics in early mouse development. DNMT1/3A/3B and TET1/2/3 expression patterns, as well as relative 5mC, 5hmC, 5fC and 5caC abundance during early mouse embryo development. ZGA is indicated with grey shadowing. ♂: paternal PN. ♀: maternal PN. Adapted from Ambrosi et al., 2017.

Global 5mC reprogramming is specific to mammals, as it has not been observed in other vertebrates. Zebrafish oocytes are hypomethylated compared to sperm, as is the case for mouse and human gametes (Jiang et al., 2013). However, no global DNA demethylation is detected between gametes and early cleavage stages (Jiang et al., 2013; Macleod et al., 1999). Instead, upon fertilization the paternal methylome remains unchanged, while maternal 5mC patterns reset to mimic sperm methylome patterns by midblastula transition (MBT) (Jiang et al., 2013; Potok et al., 2013). 5hmC is present at very low levels in early zebrafish development (Jiang et al., 2013; Potok et al., 2013). Similar to zebrafish, in *X. laevis* 5mC levels remain stable during early development (Veenstra & Wolffe, 2001), while the status of oxC is currently unknown.

1.2. The *Xenopus* model

Xenopus is one of the main vertebrate animal models in biological research, with a long-standing history in endocrinology (it was used in pregnancy testing in the 1930s and '40s), developmental biology, toxicology and disease modeling (Gao & Shen, 2021; Gurdon & Hopwood, 2000; Tandon et al., 2017). Its main advantage is that it offers an easy-to-work-with, cost-efficient model for *in vivo* experiments. Each animal can produce on demand hundreds of eggs within 16 h after injection with human chorionic gonadotropin at the dorsal lymph sacs. *Xenopus* eggs can be *in vitro* fertilized to produce embryos that will synchronously develop to tadpoles in just a few days. Conveniently, *Xenopus* embryos grow in a standard salt buffer in a petri dish. The large egg size (0.8 – 1.2 mm) and the *ex utero* development allow for various genetic manipulations: Plasmids, mRNA, morpholino modified antisense oligos (Morpholinos, Mo) and proteins can be microinjected into the embryos, while small molecules can be taken up through the growing media. Another advantage of *Xenopus* is that it is evolutionarily closer to human than other aquatic model organisms (Nenni et al., 2019). Overall, *Xenopus* provides a powerful system to study biological processes, such as embryonic development and human disease, and has led to major scientific discoveries, including the Nobel Prize awarded discoveries of cell cycle regulation (Hunt, 2002) and nuclear reprogramming (Gurdon, 1962).

The two main species of *Xenopus* used in biomedical research are *X. laevis and X. tropicalis*. *X. laevis* was the first *Xenopus* species imported from South Africa as a model organism; however, its genome is pseudo-tetraploid (co-existence of two diploid genomes; Bisbee et al., 1977; Cannatella & de Sa, 1993; Graf & Kobel, 1991; Session et al., 2016), which renders it suboptimal for genetic studies. *X. tropicalis* is the only diploid species of the genus, making it a popular alternative to *X. laevis* (Amaya et al., 1998; Hirsch et al., 2002; Kashiwagi et al., 2010). The two species share similar morphology, embryonic development and gene expression profiles (Harland

& Grainger, 2011; Yanai et al., 2011), thus they are commonly used interchangeably in research. The genomes of both species have been annotated (Hellsten et al., 2010; Session et al., 2016) and genome assemblies are constantly updated (Vize & Zorn, 2017). Moreover, RNA-seq (Owens et al., 2016; Session et al., 2016) of every developmental stage and various adult tissues, quantitative proteomics (Lombard-Banek et al., 2016; Peshkin et al., 2015; Sun et al., 2016), ATAC-seq (Bright et al., 2021) and Hi-C (Niu et al., 2021) for several developmental stages, whole genome bisulfite sequencing (WGBS) (Session et al., 2016), and an increasing amount of ChIP-seq experiments, provide a valuable resource to support *Xenopus* research. All this information, along with gene phenotypes, morpholinos, guide RNAs and more, is readily available at www.xenbase.org.

1.2.1. *Xenopus* development

Xenopus is a useful model to study embryogenesis as its morphological changes during development resemble those of human embryos. The *Xenopus* egg is polarized, with a dark colored animal hemisphere (animal pole, AP) and a lightly coloured vegetal hemisphere (vegetal pole, VP). After fertilization, the embryo undergoes a series of rapid, synchronous cleavages without increase in size and the blastocoel cavity is formed internally of the AP (stage (st) 2-6) (Figure 1.3). In morula, (st6.5) cleavages between AP and VP become asynchronous. Morula is followed by blastula (st7-9), marked by the initiation of the specification of the three germ layers (ectoderm, mesoderm and endoderm). Importantly, around st9 embryonic transcription starts, with a process known as Zygotic Genome Activation (ZGA; see section 1.3). Subsequently, during gastrulation (st10-12) the blastopore forms (Figure 1.3), cell movement begins, the three germ layers are rearranged and the anteroposterior and dorsoventral body axes are established. Gastrulation is followed by neurulation (st13-20), where the dorsal ectoderm folds to form the neural tube, the precursor of the central nervous system (Figure 1.3). Next, during organogenesis (st21- 45) cells are further specified to form all major tissues and organs (Figure 1.3). At the last stages of development, *Xenopus* tadpoles undergo metamorphosis (st46-66), a process characteristic to amphibians, insects and other organisms, during which the limbs form and the tail and head shrink.

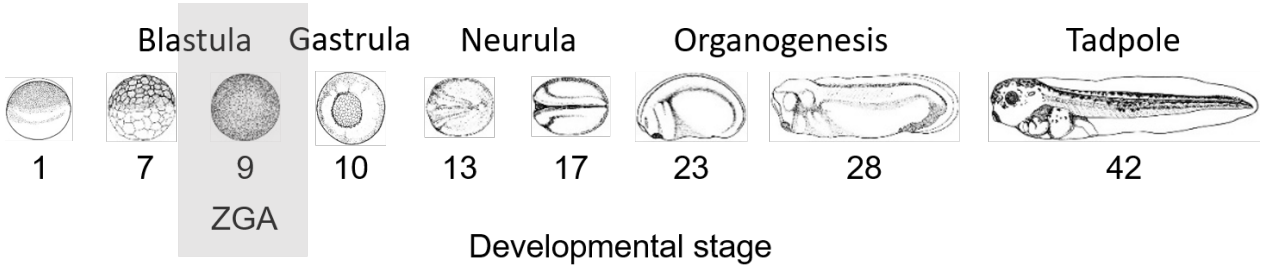


Figure 1.3. *Xenopus* developmental stages.
Representative embryo stages of *Xenopus* 1-cell (st1), blastula (st7, 9), gastrula (st10), neurula (st13, 17), organogenesis (st23, 28) and tadpole (st42). ZGA is indicated with grey shadowing. Drawings from (Nieuwkoop & Faber, 1994).

1.3. ZGA

At the beginning of embryogenesis, development is under the control of maternally provided transcripts and proteins that were loaded to the egg during oogenesis. At this point, the embryonic genome remains quiescent. As development progresses, developmental control passes from maternal to embryonic products, in a process known as Maternal-to-Zygotic transition (MZT). MZT duration varies from a few hours in *Drosophila* and *Xenopus*, to days in mouse and human (reviewed by Tadros & Lipshitz, 2009; Vastenhouw et al., 2019). MZT is characterized by two main processes: the maternal RNA degradation (MRD) followed by the zygotic genome activation (ZGA). MRD is initially mediated by maternally provided RNA binding proteins (RBPs) and is progressively handed over to embryonic RBPs and small non-coding RNAs (Aanes et al., 2011; Bashirullah et al., 1999; Graindorge et al., 2006; Dobson et al., 2004; Hamatani et al., 2004; Sha et al., 2020). Accordingly, ZGA historically consists of two expression waves, minor and major ZGA, that signify the gradual activation of embryonic transcription (Tadros & Lipshitz, 2009). Maternal and embryonic transcripts are often simultaneously present in early embryos, which makes experimentally distinguishing between the two complicated.

The developmental stage in which ZGA takes place differs among species. In mouse, minor ZGA begins in the zygote and major ZGA is reached at 2-cell stage (Abe et al., 2018; Aoki et al., 1997; Bouniol et al., 1995), while in human ZGA also initiates at the zygote but peaks at 8-cell stage (Asami et al., 2022; Braude et al., 1988). In *Xenopus*, major ZGA is reported after the 12th cleavage, conventionally around st9, while low levels of transcription are noted as early as st6 (Bachvabova & Davidson, 1966; Tan et al., 2013; Yanai et al., 2011; Gentsch et al., 2019; Paranjpe et al., 2013). Studying ZGA in mammals is challenging due to several technical limitations: *in utero* development hindering embryo accessibility, small number of embryos per pregnancy, and few (1 to 8) cells per embryo. Moreover, human embryo research poses important

ethical considerations and it is only permitted in a few countries under strict legal regulations. Thus, most of the current knowledge on ZGA derives from research on more accessible vertebrate models, such as *Xenopus* and zebrafish.

The embryonic genes expressed upon ZGA include various TFs and other signaling molecules, implicated in furthering maternal product degradation, awakening the zygotic genome and promoting cell differentiation. Some of the genes expressed upon ZGA are strictly zygotic, while other are already maternally provided and re-expressed from the zygotic genome (De Renzis et al., 2007; Lee et al., 2013). In *X. tropicalis* embryos, directly after ZGA, between st9 and 10, begins the expression of genes involved in translational initiation, elongation and termination (Collart et al., 2014), signifying the handover of protein synthesis from maternal to embryonic control. Notably, in *Xenopus* embryos, Pol II and Pol III transcription is reported at the onset of ZGA, while Pol I transcription does not initiate until late gastrula stage (st12) and only at the AP, showing an RNA polymerase type dependent spatiotemporal transcriptional activation (Knowland, 1970; Newport & Kirschner, 1982a; Woodland & Gurdon, 1969). Concerning the spatial onset of ZGA in *Xenopus*, embryonic transcription initiates at the very center of the AP and gradually spreads towards the VP (Chen et al., 2019). Stochastic ZGA is also observed in other models, such as *Drosophila* and zebrafish (Boettiger & Levine, 2009; Stapel et al., 2017).

1.3.1 Regulation of ZGA

ZGA is an elaborate process involving multiple levels of regulation to ensure successful progression of development. One of the mechanisms mediating repression of the embryonic genome during early development involves maternally loaded transcriptional repressors. For example, xKaiso binds 5mC-modified sequences to repress embryonic gene expression before ZGA in *X. laevis* embryos (Ruzov et al., 2004). Additionally, early vertebrate embryos have a large surplus of histones (Adamson & Woodland, 1974; Anderson & Lengyel, 1980), which function as a general transcriptional repressors by forming nucleosomes that prevent the transcriptional machinery from accessing the DNA (Almouzni & Wolffe, 1995; Amodeo et al., 2015; Lorch et al., 1987; Workman & Kingston, 1998). The nucleus/cytoplasm ratio model (or DNA/cytoplasm model) suggests that maternal inhibitor(s) (e.g. free histones) present at the cytoplasm are titrated down by the progressively reduced amount of cytoplasm (or increased amount of DNA) per cell during early cleavages, resulting in transcriptional de-repression (Newport & Kirschner, 1982a, 1982b; Jevtić & Levy, 2015, 2017; Chen et al., 2019). Indeed, in zebrafish and fruit fly, approaching ZGA, soluble histone levels decline (Joseph et al., 2017). Based on this, it was proposed that histones outcompete TFs for binding to the DNA in early development, and only after histones are diluted,

are the TFs able to bind and initiate transcription (Joseph et al., 2017). The nucleus/cytoplasm ratio model would also explain the delayed ZGA in the *Xenopus* VP cells, which are larger than the AP cells at st9 (Chen et al., 2019). However, it would not apply to mouse and human embryos, where minor ZGA already begins in 1-cell stage.

Many TFs are retained in the cytoplasm during early cleavages and only enter the nucleus at late blastula, around ZGA (Artee Luchman et al., 1999; Dreyer, 1987; Veenstra et al., 1999; Brewer et al., 1995; Miller et al., 1991), suggesting that TF nuclear translocation timing regulates the onset of embryonic gene expression. TF availability is also controlled at the level of translation. After fertilization, the polyA tales of many maternal transcripts are either too short or completely absent, rendering them unstable and incompatible with translation (reviewed in Woodland, 1982). It has been suggested that ZGA relies on the progressive increase of TF abundance, as a consequence of mRNA polyadenylation and subsequent translation (Blitz & Cho, 2021; Collart et al., 2014). Characteristically, the TATA-box binding protein (*tbp*) mRNA undergoes cytoplasmic polyadenylation upon fertilization and Tbp protein levels gradually increase until st9 in *Xenopus* (Veenstra et al., 1999). In contrast to TFs, all three RNA polymerases localize at the nucleus before ZGA in *Xenopus*, and the amounts of each protein remain relatively stable from egg to gastrula (Roeder, 1974). Additionally, Pol II core TF TfIIb and the Tfiif subunit Gtf2f1, are stably expressed through early embryogenesis (Veenstra et al., 1999), and Pol III TF Gtf3c1 (also known as Tfiiica, the DNA binding subunit of Tfiiic) protein levels remain stable from egg to st10 (own unpublished data, not shown). Thus, additional signaling, apart from protein availability and localization, controls the transcriptional activation of RNA polymerases. In the case of Pol II, Polr2a subunit is dephosphorylated upon fertilization and subsequently re-phosphorylated at C- terminal Ser2 and Ser5 upon ZGA, resulting in transcriptional activation (Palancade et al., 2001). Still, since a small amount of embryonic transcription is reported before st9 (Collart et al., 2014), the activation of RNA polymerases is most likely regulated at a sequence- and chromatin- specific rather than a global level.

It is clear that multiple mechanisms work together to fine-tune the timing and progress of zygotic expression. On top of that, there are distinct mechanisms regulating the activity of the three different RNA polymerases (Almouzni & Wolffe, 1995). However most studies around ZGA focus on Pol II gene expression regulation. Moreover, in order to get a clear view of mRNA transcription control, many studies use polyA enrichment or rRNA depletion to exclude Pol I and Pol III transcripts. Since Pol I encodes rRNAs and Pol III encodes tRNAs (see section 1.4.), both indispensable for protein synthesis, studying their activation will allow for a deeper understanding of the embryonic genome awakening.

1.3.2. Epigenetic reprogramming

In early development, chromatin maintains a relatively open state, permissive to global epigenetic reprogramming. Before ZGA, the embryonic genome is largely unstructured, it lacks topologically associated domains (TADs) (Du et al., 2015; Hug et al., 2017; Niu et al., 2021) and heterochromatin (Laue et al., 2019; Mutlu et al., 2018), while it maintains high mobility (Bošković et al., 2014). Hence, transcriptional silencing is majorly due to lack of contact between the TFs and their target genes, rather than global chromatin compaction. Approaching ZGA, chromatin transitions to a higher order structure and becomes globally more compact, while specific genomic regions gain accessibility (Vastenhouw et al., 2019). Changes in nucleosome composition and positioning, as well as establishing of histone modifications, facilitate chromatin remodeling that in turn alters TF accessibility and leads to transcriptional activation.

During zygotic genome reprogramming, specialized histone variants, activating and repressive histone modifications, and nucleosome repositioning orchestrate TF accessibility to the genome. Upon fertilization, alternative histone variants replace several core histones. In early embryogenesis, an embryonic variant of the somatic linker histone H1 contributes to less compact chromatin and allows for chromatin remodeling. Embryonic H1 is later replaced by somatic H1, which leads to chromatin compaction and is required for ZGA (Dimitrov et al., 1993; Pérez-Montero et al., 2013; Saeki et al., 2005; Ura et al., 1996). In parallel, modification of histone tails alters chromatin accessibility, either directly or indirectly. More specifically, histone acetylation alters the charge of histone tails, impairing histone-DNA contact and creating a permissive local environment for TF engagement (Klemm et al., 2019). On the other hand, histone methylation acts through reader proteins that either promote or inhibit transcription. Accordingly, the levels of the activating histone marks H3K4me3, H3K9ac and H3K27ac increase around ZGA (Gupta et al., 2014; Hontelez et al., 2015; van Heeringen et al., 2014). Notably, recruitment of p300, the histone acetyltransferase responsible for H3K27ac, is mostly zygotically regulated, as α-amanitin induced transcription inhibition erased p300 signature by 85% in gastrula stage *Xenopus* embryos (Hontelez et al., 2015). PRC2 deposited H3K27me3 is also established at ZGA and it is selectively enriched at the AP during gastrula stages, where it marks genes that are differentially expressed at the VP, regulating spatial gene expression on the AP-VP axis (Akkers et al., 2009; Hontelez et al., 2015; van Heeringen et al., 2014). Furthermore, chromosome remodeling by specialized complexes can directly alter chromatin accessibility. Importantly, Brg1, the catalytic ATPase subunit of the SWI/SNF chromatin remodeling complex, is required for Pol II transcription in mouse ZGA (Bultman et al., 2006).

Are DNA modifications involved in ZGA? Even though histone modifications are clearly indispensable for epigenetic reprogramming, the role of DNA modifications in ZGA is elusive. Upon fertilization, mouse and human embryos undergo global DNA demethylation; however, this is not the case for other vertebrates (see section 1.1.7). Other than that, DNA methylations does not seem to play a role in ZGA. Studies in *Xenopus* conclude that there is no strong correlation between promoter DNA methylation and transcriptional repression in blastula and gastrula stages (Bogdanović et al., 2011). Likewise, both promoter methylation and hypomethylation are compatible with gene expression during ZGA in zebrafish (Andersen et al., 2012). However, 5mC might prevent early expression of gastrulation and morphogenesis related genes during zebrafish ZGA (Potok et al., 2013). Since no oxC or other DNA modification have been implicated in zygotic reprogramming, it currently remains unclear whether and how DNA modifications might be involved in ZGA.

1.4. RNA Pol III

Pol III is the largest RNA polymerase, consisting of 17 subunits, five of which are shared between all three RNA polymerases (Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12) (Hoffmann et al., 2015; Ramsay et al., 2020). Pol III specializes in high content transcription of short (30 – 400 nt), non-protein coding genes, such as tRNAs, U6 small nuclear RNA (snRNA) and 5S rRNA, and approximately 15% of all cell transcripts are Pol III transcribed (Moir & Willis, 2013).

The Pol III holoenzyme is rather autonomous, which makes Pol III transcription mechanistically simpler compared to Pol II. In contrast to Pol II, Pol III does not require post-transcriptional modifications, such as C-terminal phosphorylation to transition through the different phases of transcription. Hence, Pol III binding is typically considered a direct proxy for active Pol III transcription. In addition, Pol III transcription rate is markedly higher than that of Pol II. Characteristically in yeast, while Pol II occupancy is one enzyme per 12000 bp (Pelechano et al., 2010), Pol III occupancy is one enzyme per 44 bp in 5S rRNA genes (French et al., 2008). The unique ability of Pol III in high rate transcription of short non-coding RNAs is utilized in biochemistry and specifically in the CRISPR-Cas9 system, where Pol III promoters drive the transcription of guide RNAs (gRNAs) used to target selected genes (Kor et al., 2023; Schwartz et al., 2016).

1.4.1. Promoter types

Different Pol III genes have different promoter types, namely type I, II or III promoters (Figure 1.4A). Type I and type II promoters are intragenic and they are located downstream of the TSS of 5S rRNA and tRNA genes, respectively. The internal promoter elements of the 5S rRNA (type I promoter) are an A-box and a C-box, while tRNA genes internal promoters (type II promoter)

consist of an A-box and a B-box, which coincide with the D-loop and the T-loop of the mature tRNA transcript, respectively (Turowski & Tollervey, 2016) (Figure 1.4A). Notably, short interspersed nuclear elements (SINEs), the most abundant class of transposable elements in human, derive from Pol III genes, specifically tRNA, 7SL RNA or 5S rRNA genes, and they are Pol III transcribed under type II promoters (Kramerov & Vassetzky, 2011). Finally, type III promoters closely resemble Pol II promoters, as they are located upstream of the TSS and are used for U6, 7SK, RNase P and mitochondrial RNA processing (MRP) transcription (Dieci et al., 2007; Schramm & Hernandez, 2002) (Figure 1.4A).

1.4.2. Transcription regulation

Transcription takes place in three highly coordinated steps: initiation, elongation and termination. Transcription initiation involves promoter recognition, Pol III recruitment upstream of the TSS and promoter opening, all of which are mediated by specialized TFs. Pol III transcription relies on three conserved TFs: TFIIIA, TFIIIB and TFIIIC. TFIIIA is a single ~40 kDa protein responsible for type I promoter recognition and TFIIIB recruitment to 5S rRNA genes (Camier et al., 1995). It was the first eukaryotic TF purified in 1980 from *X. laevis* oocytes (Engelke et al., 1980). TFIIIB is the only TF common to all promoter types (I, II and III), as it is responsible for Pol III recruitment and promoter opening (Hahn & Roberts, 2000; Kassavetis et al., 1998). It is itself recruited by either TFIIIA (5S rDNA) or TFIIIC (tDNA). TFIIIB consists of three proteins, TATA-binding protein (TBP; also a Pol I and Pol II TF), B-related factor (either BRF1 or BRF2) and B double primer (BDP1), which only stably associate when in complex with the DNA (Schramm et al., 2000; Schramm & Hernandez, 2002). Lastly, TFIIIC is responsible for type II promoter recognition and TFIIIB recruitment (Baker et al., 1987; Deprez et al., 1999) (Figure 1.4B). TFIIIC is the largest Pol III TF with a total mass of 520 kDa and it consists of six subunits organized in two subcomplexes, τA and τB (Marzouki et al., 1986; Schultz et al., 1989). τA binds the A-box and interacts with TFIIIB and Pol III, while τB binds the B-box mediating high affinity interaction with the promoter.

After Pol III is successfully loaded to the target gene, the initiation complex undergoes remodeling to accommodate elongation. Characteristically, TFIIIA and TFIIIC are essential for transcription initiation but they are dispensable after TFIIIB has been recruited to the promoter (Kassavetis et al., 1990; Soragni & Kassavetis, 2008). Notably, in yeast TFIIIC competes with Pol III for DNA binding to the transcribed region (Ciesla et al., 2018; Roberts et al., 2003). These findings suggest that TFIIIA and TFIIIC hinder Pol III transcription, as they occupy internal promoter elements, and they have to be removed for successful elongation by Pol III. On the contrary, TFIIIB remains bound to the target even after Pol III release, facilitating quick transcription re-initiation on the pre-

assembled TFIIIB, which results in the characteristic fast turnover of Pol III. Finally, Pol III transcription termination relies on a stretch of 4-7 adenines on the non-coding strand (Arimbasseri et al., 2013; Braglia et al., 2005; Cozzarelli et al., 1983). Weak interaction between poly-A on the template DNA and poly-U on the nascent transcript destabilize the Pol III-DNA interaction, while poly-T on the coding strand induce Pol III staling and transcript release (Arimbasseri & Maraia, 2015).

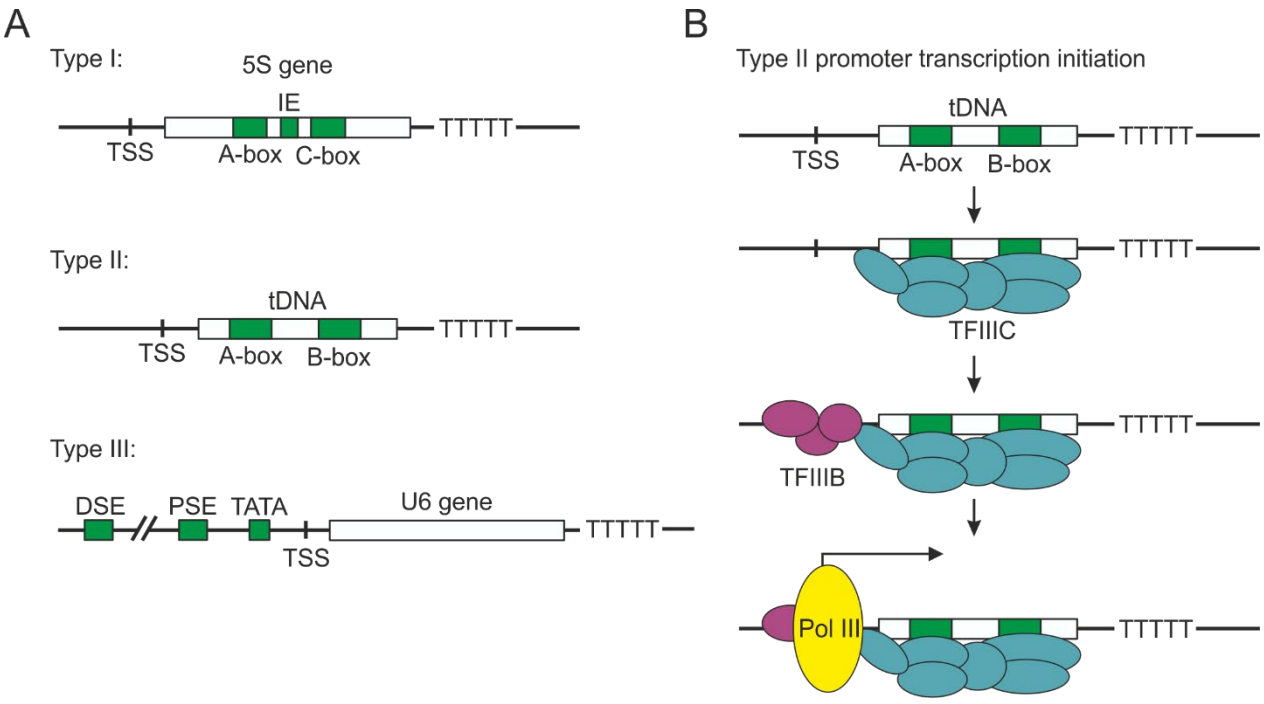


Figure 1.4. Pol III transcription under different promoter types.
A. Type I, II and III Pol III promoters. Top: Type I promoter of *Xenopus 5S rRNA* gene includes intragenic promoter elements A-box, intermediate element (IE) and C-box. Middle: Type II promoter of general tRNA gene includes the intragenic promoter elements A-box and B-box that correspond to the mature tRNA D-loop and T-loop, respectively. Bottom: Type III promoter of *Homo sapiens* U6 gene includes a distal sequence element (DSE), a proximal sequence element (PSE) and a TATA-box, all upstream of the U6 sequence. Based on Schramm & Hernandez 2002. **B.** Type II promoter stepwise initiation complex assembly and Pol III recruitment. Based on Turowski & Tollervey, 2016.

Pol III transcription is highly coordinated in space and it has been associated with the perinucleolar compartment (PNC). The PNC is a membraneless organelle at the periphery of the nucleolus (Huang et al., 1997), mostly present in transformed cells and its structural integrity depends on active Pol III transcription (Kopp & Huang, 2005). In order to understand Pol III transcription regulation in 3D space it is critical to describe the organization of its target genes in the genome. In eukaryotes, 5S rRNA genes cluster in tandem repeats (Cloix et al., 2000; Daniels & Delany, 2003; Kress et al., 2001; Procunier & Tartof, 1975; Sørensen & Frederiksen, 1991; Steffensen et al., 1974). Notably, in *Xenopus*, around 20,000 oocyte-specific *5S* genes, clustered along different chromosomes are expressed in oocytes, while only about 400 somatic *5S* genes from a single cluster on chromosome 9 are expressed in somatic cells (Pardue et al., 1973; Peterson et al., 1980). *In situ* hybridization of *5S* genes showed that they coalesce at the nucleolar periphery where they are transcribed. Importantly, 5S gene clustering at the PNC is conserved in plants (Highett et al., 1993), flies (Ananiev et al., 1981), frogs (Pardue et al., 1973), Chinese hamsters (Amaldi & Buongiorno-Nardelli, 1971) and humans (Matera et al., 1995). tRNA genes are mostly scattered throughout the genome, although a limited number of tRNA gene clusters have been described in the rat (Rosen et al., 1984), mouse (Wood et al., 1991) and human genome (Shortridge et al., 1989). Markedly, *Xenopus* oocyte-specific *tRNA*s cluster in tandem repeats across different chromosomes, while somatic *tRNA*s are low copy number and dispersed (Clarkson et al., 1973; Müller et al., 1987; Stutz et al., 1989). Despite being interspersed, tRNA genes coalesce at the nucleolus where they are transcribed and processed (Hopper et al., 2010; Thompson et al., 2003). Whether other, single copy Pol III target genes also localize at the PNC is less clear, mostly due to technical difficulties of *in situ* visualization of single copy genes. Overall, Pol III transcription is spatially allocated to the PNC, where Pol III targets coalesce from different chromosomes.

1.5. 5fC accumulates at Pol III target genes during *Xenopus* ZGA

Previous work from this lab conducted by Dr. Dandan Han and Dr. Michael Musheev quantified the levels of 5mC and its oxidative derivatives during *X. laevis* development using mass spectrometry (Figure 1.5A). 5mC reaches ~6% of total dC at st9 and it remains stable until tadpole stages. 5hmC is present at very low levels until st23, from whereon it gradually increases. Notably, 5fC levels show a local peak at st9-10, with 5fC values 5-fold higher than 5hmC at the same stages (Figure 1.5A). 5fC levels decrease by st13, to gradually increase again from st23 on, similar to its precursor 5hmC. Lastly, 5caC is undetectable before neurulation and gradually increases from st23 on. This 5fC specific accumulation in early development was intriguing, as st9 corresponds to ZGA in *Xenopus* (see section 1.2.1).

To visualize the various DNA modifications in early *X. laevis* embryos, Dr. Victoria Hatch, this lab, performed 5mC, 5hmC, 5fC and 5caC immunofluorescence (IF) at st9 embryos. While 5mC positive nuclei were detected throughout the entire embryo, 5fC nuclei were exclusively present at the AP (Figure 1.5B). Importantly, within individual nuclei 5mC signal was homogeneously distributed, while 5fC formed distinct foci (Figure 1.5C). We named these foci "5fC chromocenters". 5hmC and 5caC signals were at background levels at st9, in agreement with the mass spectrometry data. 5fC chromocenters are transient, as they start forming at st8 and are detectable at st9 and 10, while st13 (neurula) nuclei homogeneously stain for 5fC (Figure 1.5D). Double IF for 5fC and the nucleolus marker xNopp180 (Schmidt-Zachmann et al., 1984) showed that the 5fC chromocenters localize directly adjacent to the nucleolus, at the PNC (Figure 1.5E, top). Double IF for Pol III and the nucleolus marker xNopp180 showed that Pol III is also located at the PNC (Figure 1.5E, middle). Importantly, double IF for 5fC and Pol III showed that the 5fC chromocenter colocalizes with Pol III at the PNC during *Xenopus* ZGA (Figure 1.5E, bottom). Metaphase chromosome spreads staining revealed that the 5fC chromocenter is composed of genomic loci residing on different chromosomes, which are coalescing in interphase nuclei to form a single compartment (Figure 1.5F).

To study the genomic localization of the different DNA modifications in ZGA, Dr. Victoria Hatch and Dr. Medhavi Mallick, this lab, performed DNA Immunoprecipitation sequencing (DIP-seq) for 5mC, 5hmC, 5fC and 5caC at *X. tropicalis* st9 embryos. DIP-seq showed that, while Pol II target genes were similarly occupied by all DNA modifications and Pol I targets were lowly occupied by any of them, Pol III targets were selectively enriched for 5fC, followed by 5mC (Figure 1.5G). A characteristic of 5fC peaks at Pol III targets was that they cluster along tens of kilobases (Figure 1.5H-I). More specifically, 5fC clustered at oocyte-type *tRNA* gene tandem repeats, which are distinct from the low copy number somatic-type *tRNA* genes (see section 1.4.2). Oocyte *tRNAs* are expressed during oogenesis and upon ZGA, while somatic *tRNA* expression initiates in early neurulation, when oocyte-type *tRNA* expression stops (Andrews et al., 1991). Comparing the occupancy of the different DNA modifications and Pol III (Hontelez et al., 2015) across *tRNA* genes in ZGA showed that while Pol III occupancy is relatively homogeneous, 5fC occupancy was very heterogeneous (Figure 1.5J). More specifically, 5fC peaks were most prevalent on the initiator *tRNA-iMet*, followed by *tRNA-Gly* and *tRNA-His*, while for example elongation *tRNA-eMet* was lowly occupied by 5fC (Figure 1.5J). A small fraction of 5mC peaks overlapped with *tRNA* genes, while 5hmC and 5caC were almost absent. The above observations are visualized in Figure 1.5K, on a characteristic oocyte-type *tRNA* tandem repeat cluster expressed during *X. laevis* ZGA (Clarkson & Kurer, 1976; Müller et al., 1987), which has a 143 kb homologue in *X. tropicalis*. The

cluster is comprised of sequential repeats of the same 3 kb long *tRNA* cassette. Each cassette includes two *tRNA-iMet* genes, followed by *tRNA-Tyr*, *tRNA-Lys*, *tRNA-Leu*, *tRNA-Ala* and *tRNA-Asn* (Figure 1.5K). While Pol III occupied each of the *tRNAs* in the cassette, a prominent 5fC peak occupied the two *tRNA-iMet* genes (Figure 1.5K). In contrast, the most prominent 5mC peak was ~2 kb offset the 5fC peak and coincided with a Pol III occupancy valley (Figure 1.5K).

In summary, previous work from this lab identified for the first time a 5fC accumulation on Pol III target genes during *Xenopus* ZGA. At this crucial developmental stage, 5fC forms transient chromocenters at the PNC, where it colocalizes with Pol III. On a genomic level, 5fC marks oocyte-type *tRNA* gene tandem repeats, which are also occupied by Pol III.

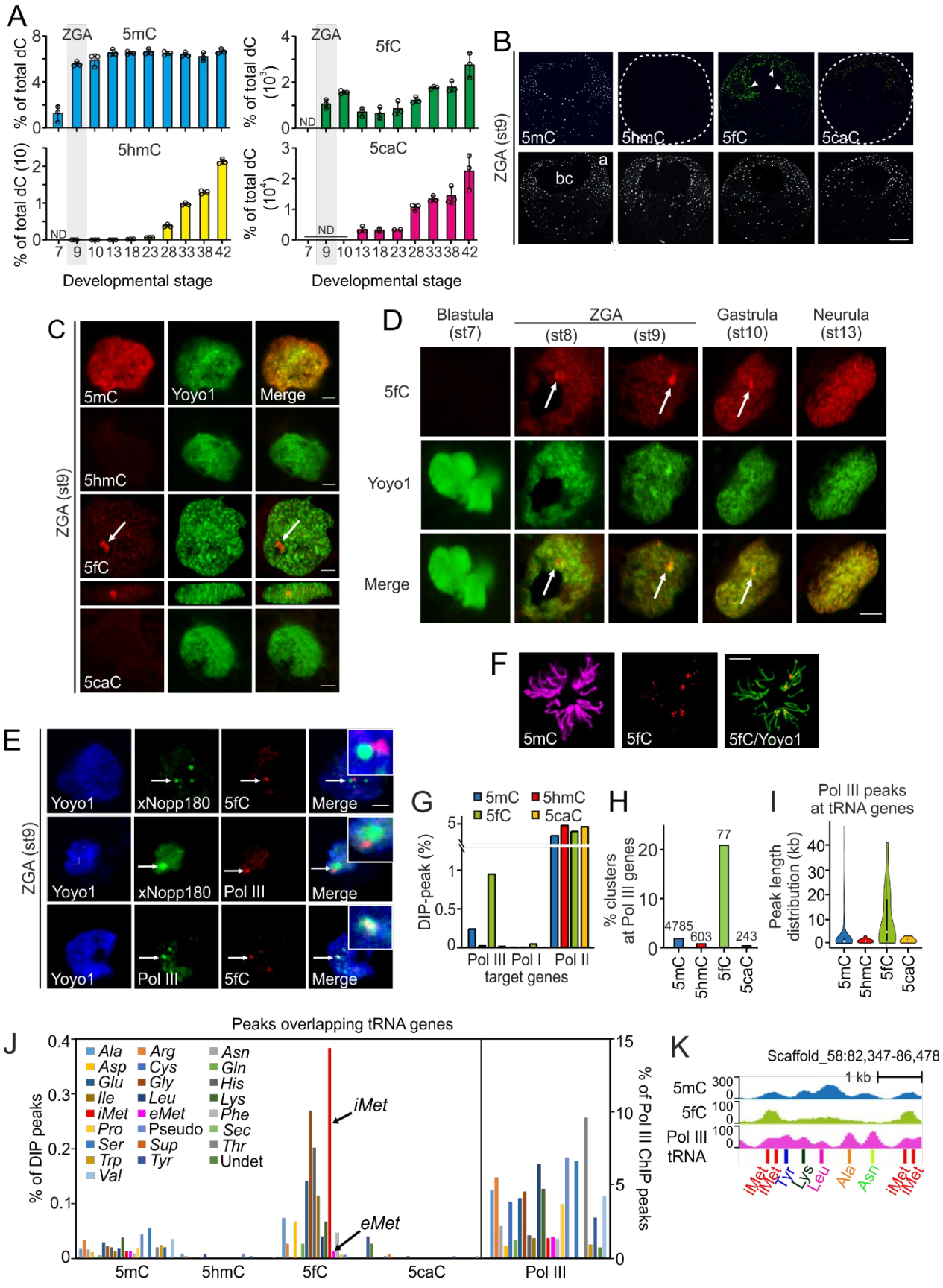

A
ZGA
5mC
5fC
5hmC
5caC
% of total dC
% of total dC (10)
% of total dC (10)
ND
7 9 10 13 18 23 28 33 38 42
Developmental stage
B
ZGA (st9)
5mC
5hmC
5fC
5caC
a
bc
C
ZGA (st9)
5mC
Yoyo1
Merge
5hmC
5fC
5caC
D
Blastula (st7)
ZGA
(st8)
(st9)
Gastrula (st10)
Neurula (st13)
5fC
Yoyo1
Merge
E
ZGA (st9)
Yoyo1
xNopp180
5fC
Merge
Pol III
F
5mC
5fC
5fC/Yoyo1
G
5mC
5hmC
5fC
5caC
DIP-peak (%)
Pol III
Pol I
Pol II
target genes
H
% clusters at Pol III genes
4785
603
77
243
I
Pol III peaks at tRNA genes
Peak length distribution (kb)
J
Peaks overlapping tRNA genes
% of DIP peaks
% of Pol III ChIP peaks
Ala
Asp
Glu
Ile
iMet
Pro
Ser
Trp
Val
Arg
Cys
Gly
Leu
eMet
Pseudo
Sup
Tyr
Asn
Gln
His
Lys
Phe
Sec
Thr
Undet
iMet
eMet
5mC
5hmC
5fC
5caC
Pol III
K
Scaffold_58:82,347-86,478
1 kb
5mC
5fC
Pol III
tRNA
iMet
iMet
Tyr
Lys
Leu
Ala
Asn
iMet
iMet

Figure 1.5. 5fC accumulates in transient chromocenters that represent Pol III target genes during *Xenopus* ZGA.
A. Mass spectrometry analysis of 5mC, 5hmC, 5fC and 5caC levels during *X. laevis* development. n=3 biological replicates +/- SD. ND, not detected. **B.** Immunofluorescence (IF) microscopy of st9 *X. laevis* embryos stained for 5mC, 5hmC, 5fC and 5caC. 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. Arrow indicates the 5fC chromocenter. Yoyo1, DNA stain. Scale bar 5 µm. **D.** IF for 5fC in nuclei of *X. laevis* embryos at st7 (pre-ZGA), 8 (early ZGA), 9 (ZGA), 10 (early gastrula) and 13 (early neurula). Arrows indicate 5fC chromocenter. Yoyo1, DNA stain. Scale bar 5 µm. **E.** Co-IF of 5fC and the nucleolar marker xNopp180 (top), Pol III and xNopp180 (middle) or 5fC and Pol lll (bottom). Arrows indicate adjoining 5fC-xNopp180 and Pol III-xNopp180 spots, and colocalizing 5fC-Pol III. Yoyo1, DNA stain. Scale bar 5 µm. **F.** 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. **G.** Correlation of 5mC, 5hmC, 5fC and 5caC peaks from DIP sequencing with Pol I, Pol II and Pol III target genes. **H.** Percent 5mC, 5hmC, 5fC or 5caC-peak clusters at Pol III genes. Total count of clusters indicated over each bar. **I.** Violin plots showing peak length distribution of Pol III peaks overlapping with DIP peaks at *tRNA* genes. **J.** Percentage of 5mC, 5hmC, 5fC, 5caC DIP-peaks and Pol III-ChIP peaks overlapping with various *tRNA* genes. **K.** UCSC browser screenshot of a 5fC *tRNA* gene cluster with normalized DIP sequencing and ChIP sequencing coverage tracks of st9 (ZGA) *X. tropicalis* embryos.

1.6. Tet proteins promote Pol III binding

To manipulate endogenous 5fC levels, Dr. Victoria Hatch used morpholinos against Tet proteins, the 5fC writers. Individual *Tet2* and *Tet3* Mo KD induced embryonic lethality and severe developmental abnormalities during gastrulation, while *Tet2/3* double knockdown was 100% lethal by late gastrula (Figure 1.6A-B). These findings suggest that Tet proteins are required for normal progression of early embryonic development in frog.

Next, Dr. Victoria Hatch used *Tet2*/3 KD and *Tet2/3* OE st9 *X. tropicalis* embryos to perform Pol III ChIP-qPCR. She found that compared to the control, *Tet2/3* KD reduced Pol III binding to 5fC-rich *tRNA* genes and other 5fC-rich Pol III targets (*5S*, *U6*) in ZGA (Figure 1.6C). Conversely, *Tet2/3* OE induced Pol III binding, selectively on 5fC-rich Pol III targets. Furthermore, Tfiiic ChIP-qPCR in *Tet2/3* KD ZGA (st9) embryos showed reduced Tfiiic binding to 5fC-rich *tRNA* genes compared to the control (Figure 1.6D). Together, these data showed that Tet proteins promote Pol III binding to its targets in ZGA and hint towards a potentially instructive role of 5fC for Pol III transcription during *Xenopus* ZGA. However, since Tet proteins have catalytic activity-independent roles in gene expression (see section 1.1.3), the requirement of 5fC for Pol III binding needs to be carefully validated.

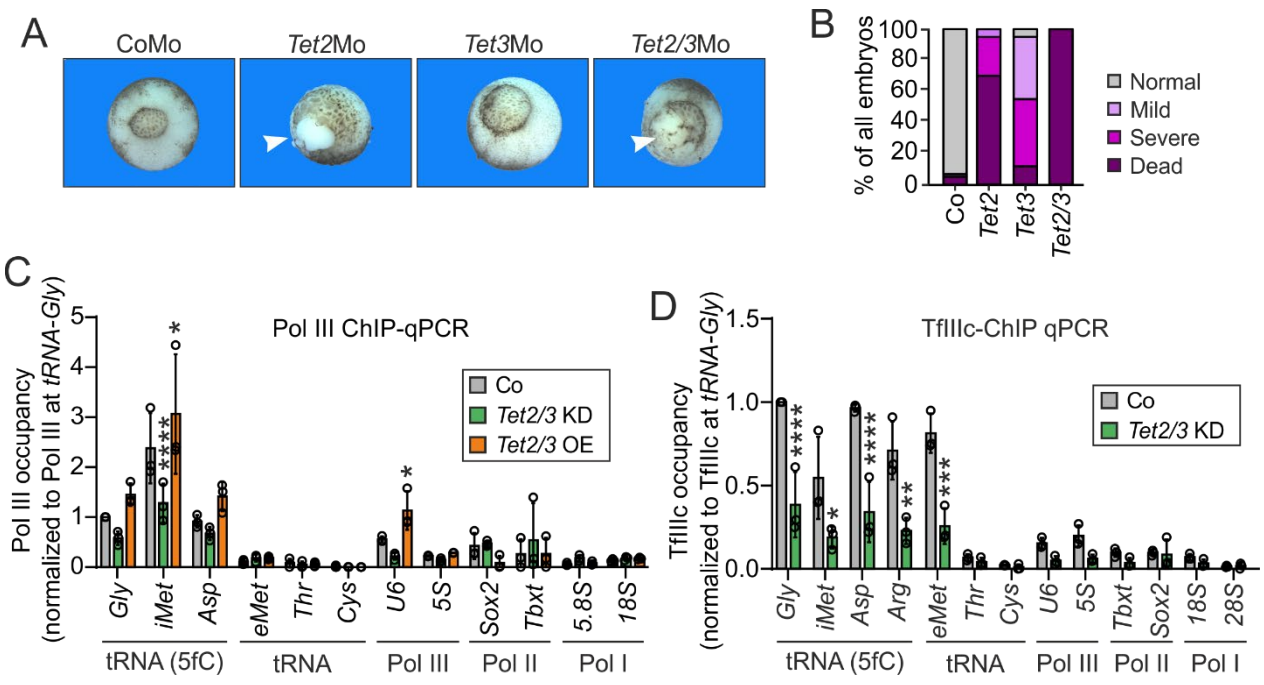


Figure 1.6. Tet promotes Pol III binding at ZGA.

A. *X. tropicalis* gastrula stage embryos injected with the indicated morpholinos (Mo). Arrowheads indicate cell death. **B.** Phenotype scoring of gastrula stage *X. tropicalis* embryos injected with the indicated Mo. N=90 embryos per condition. **C.** Pol III Chip-qPCR in control (Co), *Tet2/3* KD and *Tet2/3* OE st9 (ZGA) *X. tropicalis* embryos. Data are mean of n=3 independent ChIP pulldown samples +/- SD. P-values are calculated with 2-way ANOVA and corrected for multiple comparisons with Dunnett's test. **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.

1.7. Aim

Epigenetic modifications are essential regulators of gene expression involved in numerous physiological and pathological process, from embryonic development to ageing and disease. Although the regulatory roles of various histone and RNA modifications have been extensively studied, the picture is less clear when it comes to DNA modifications. The most prominent DNA modification, DNA methylation or 5mC, is associated with gene promoter silencing and it is known to be reversible through active DNA demethylation. The 5mC oxidative derivatives that are produced during active DNA demethylation, namely 5hmC, 5fC and 5caC, have divided the field over whether they are mere intermediates of 5mC removal, or DNA modifications with their own regulatory function. Even though many studies describe the genomic localization and *in vitro* characteristics of these marks, there is currently no conclusive evidence establishing their regulatory roles in a physiologically relevant context.

As detailed in sections 1.5 and 1.6, previous work from this lab discovered 5fC accumulation during *Xenopus* ZGA, specifically on Pol III targets. Tet manipulations showed Tet-dependent Pol III binding to 5fC-rich targets in ZGA, suggesting that 5fC could potentially function as a positive regulator for Pol III in ZGA.

This study aims to characterize the role of 5fC accumulation on Pol III targets during *Xenopus* ZGA. More specifically, during my PhD I sought to answer three main questions:

1. Does 5fC promote Pol III binding during ZGA?
2. Does 5fC promote Pol III transcription during ZGA?
3. What is the mechanism by which 5fC regulates Pol III in ZGA?

By addressing these questions, this study provides for the first time direct evidence that 5fC is a DNA modification with a defined role in a physiological process, namely tRNA transcription activation during zygotic reprogramming. The present work adds 5fC as a second name in the list of instructive DNA modifications, after the over many years established 5mC.

2. Results

2.1. 5mC and oxC kinetics in early *X. laevis* development

Previous work from this lab showed that during early *Xenopus* development, 5fC selectively accumulates at st9 and 10, with a 5-fold enrichment over 5hmC (see section 1.5 and Figure 1.5A). To gain further insight into 5mC and oxC dynamics around ZGA, in collaboration with Dr. Michael Musheev, we performed a detailed time course of 5mC and oxC dynamics focusing on early development and especially the window around ZGA. I collected *X. laevis* embryos between blastula st6.5 and neurula st13, in 30 min to 1 h intervals (Figure 2.1.A). High quality gDNA from pools of 20 embryos per stage in three biological replicates was analyzed by LC-MS/MS. Our analysis showed that 5mC gradually increases between st6.5 and 9, when it stabilizes at around ~6% of total dC (Figure 2.1B), like previously observed (Figure 1.5A). Importantly, close 5hmC inspection during early development showed a transient 5hmC accumulation between st6.5 and the onset of st8, when ZGA initiates. Half an hour into st8, 5hmC gradually decreased to give rise to 5fC, which started accumulating at st8 + 1 h and peaked at st9 (Figure 2.1B). This observation suggests that 5mC undergoes a one-step Tet mediated oxidation before ZGA, which allows for 5hmC accumulation. Upon ZGA, 5hmC is further oxidized by Tet enzymes to 5fC, resulting in a transient 5fC accumulation during ZGA. Notably, 5fC levels were up to ~6 fold higher than those of its precursor 5hmC during ZGA (Figure 2.1B), reproducing previous findings (Figure 1.5A). The 5fC peak gradually decreased by early gastrula (st10) and continued decreasing in neurula (st12-13). 5caC was not detected before neurula and only appeared in low levels (0.000016% of total dC) at st13 (Figure 2.1B), like previously observed (Figure 1.5A). Collectively, these data capture the stepwise transition from 5mC to 5hmC and 5fC in early *Xenopus* development, delineating the process of transient 5fC accumulation during ZGA.

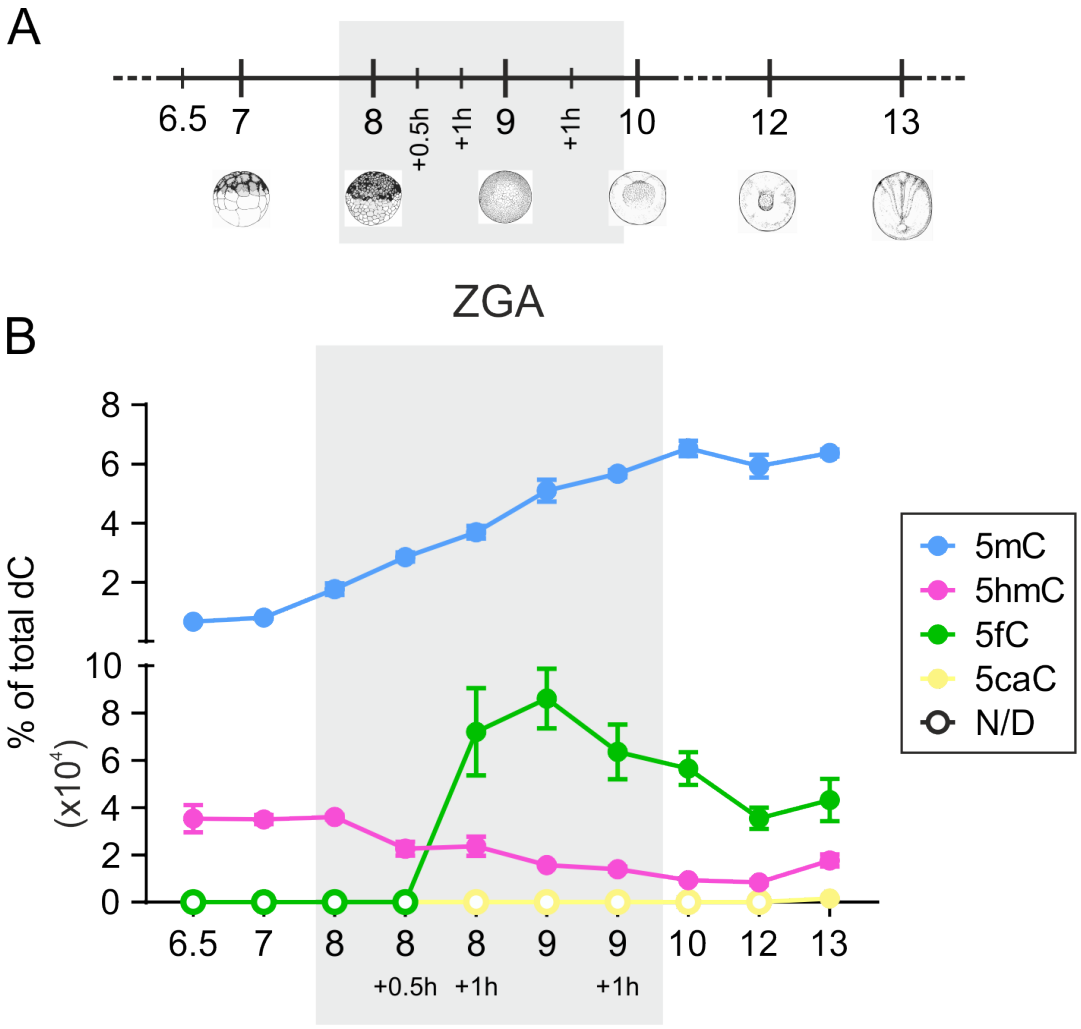


Figure 2.1. 5mC and oxC dynamics in early *Xenopus* development.

A. Time course of *X. laevis* embryo collection for LC-MS/MS analysis of genomic 5mC and oxC levels. **B.** LC-MS/MS analysis of genomic 5mC, 5hmC, 5fC and 5caC of st6.5-13 *X. laevis* embryos. Embryos were harvested at the beginning of the indicated Nieuwkoop & Faber stages or 0.5 or 1 h afterwards. n=3 biological replicates. Error bars show SD. N/D, not detected, signal below detection limit. Grey shadowing indicates ZGA.

2.2. Base resolution 5mC and 5fC mapping in ZGA

To validate and further previously obtained 5mC and oxC DIP-seq data (Figure 1.5K), we sought to obtain base resolution 5mC and 5fC profiles at ZGA. First, Dr. Alexandr Gopanenko, this lab, analyzed publicly available *X. tropicalis* bisulfite sequencing data (BS-seq; Bogdanović et al., 2016) from ZGA (st9), late gastrula (st12), tailbud (st30) and tadpole (st43) stages. Base resolution 5mC profiles matched well 5mC DIP profiles (Figure 2.2). All CpGs at the *tRNA-iMet* locus where heavily methylated at ZGA (st9), with methylation levels varying between 85-95%. Methylation profiles remained stable from ZGA to tadpole stage (Figure 2.2). This finding echoes our previous finding that global 5mC levels are established at ZGA (st9) and remain stable throughout development (Figure 1.5A and 2.1B). However, it is important to note that the observed 5mC profiles are whole embryo averages and the profiles of individual cell populations might defer (e.g. AP vs VP at st9, various cell types at st30 and 43).

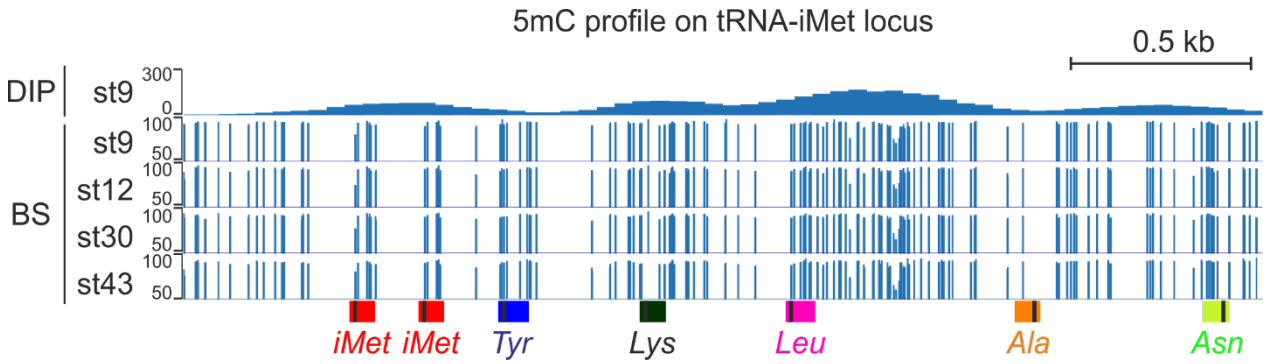


Figure 2.2. *tRNA-iMet* locus CpGs are highly methylated throughout development.
tRNA-iMet locus methylation profile from *X. tropicalis* 5mC DIP sequencing (only for st9) and base resolution BS-seq from Bogdanović et al., 2016 at the indicated stages.

To obtain a 5fC base resolution profile at ZGA, I first applied malononitrile sequencing (Zhu et al., 2017). In this protocol, malononitrile treatment converts 5fC to 5fC-M (Figure 2.3A), which is read as thymine (T) in sequencing. Hence, 5fC signal is identified as C-to-T transitions. Malononitrile conversion was highly efficient as measured by LC-MS/MS. A 5fC oligo carrying two 5fCs in selected positions showed almost complete 5fC conversion, while gDNA conversion efficiency was ~70% (Figure 2.3B). I next tested different commercial DNA polymerases to identify an enzyme that can amplify 5fC-M sequences with high efficiency, using the 5fC oligo as template. According to the published protocol Taq, Q5 (NEB) and DeepVent (exo-) (NEB) DNA polymerases can amplify malononitrile treated DNA in any sequence context (Zhu et al., 2017). However, in my hands, high fidelity DNA polymerase Q5, as well as Phusion U Hotstart and KapaHiFi U+ (Takara) failed to yield any product from the malononitrile-treated 5fC oligo template after 5 PCR cycles (Figure 2.3C). OneTaq and DeepVent (exo-) gave some PCR product with the 5fC substrate after 5 PCR cycles, although 4-fold lower than the unmodified control (Figure 2.3C). In all cases, the PCR signal picked-up at 15 cycles. Since DeepVent (exo-) was used for library preparation in the published protocol, I nevertheless selected it for library preparation for whole genome sequencing (WGS) and 5fC DIP malononitrile-seq with *X. tropicalis* st9 gDNA. 5fC DIP malononitrile-seq was intended to enrich for 5fC in case the endogenous levels were below the limit of detection.

5fC DIP profiles from control and malononitrile treated samples nicely reproduced those of the previous DIP (see section 1.5). In both conditions, 5fC peaks occupied *tRNA-iMet* and *tRNA-Gly* tandem repeat arrays end-to-end, as previously observed (Figure 2.3D-E). Notably, the 5fC mark was more prominent on the forward DNA strand, while peaks at the reverse strand were much less pronounced (Figure 2.3D-E). The major 5fC peak at the *tRNA-iMet* locus was at the two *tRNA-iMet* genes, like previously observed, while secondary peaks were detected over *tRNA-Lys* and intergenically between *tRNA-Leu* and *tRNA-Ala*, in both control and malononitrile-treated DNA (Figure 2.3F). *tRNA-Gly* locus encompassed a single peak over *tRNA-Gly* (Figure 2.3G). We conclude that the 5fC DIP in control and malononitrile conditions was successful, as it yielded the expected enrichment of the oocyte *tRNA* gene tandem repeat array and individual *tRNA* genes previously identified.

We then tested the coverage of a 5fC spike-in in control and malononitrile samples. The 5fC spike-in carried two 5fC on the forward strand, while the reverse strand was unmodified. We confirmed 3-fold enrichment at the coverage of the forward over the reverse strand in control samples, as expected (Figure 2.3H). However, in malononitrile-treated samples forward strand coverage was >10 fold lower than that of the control samples. Reverse strand coverage was almost identical between control and malononitrile samples (Figure 2.3H). Since the 5fC DIP was successful in

both conditions (Figure 2.3D-G), these data suggest that 5fC-M sequences were selectively not amplified during sequencing library preparation due to preference of the DNA polymerase for unmodified sequences (Figure 2.3C). Accordingly, no 5fC signal above noise was detected in WGS and 5fC DIP malononitrile-seq. I conclude that this method is not suitable for 5fC sequencing because DeepVent (exo-) DNA polymerase efficiency on 5fC-M substrate is greatly impaired.

I also tried pyridine borane-assisted sequencing (Liu et al., 2021). Pyridine borane treatment converts 5fC to dihydrouracil (DHU) that is sequenced as T after PCR. I faced similar problems to malononitrile-seq, as neither a uracil-tolerant DNA polymerase nor a Taq DNA polymerase were able to amplify 5fC substrates (data not shown). Collectively, chemical 5fC conversion to a DNA lesion (DHU) or a bulky adduct (5fC-M) selectively impaired the amplification of 5fC-modified sequences and is thus not compatible with sequences of high 5fC frequency at moderate levels.

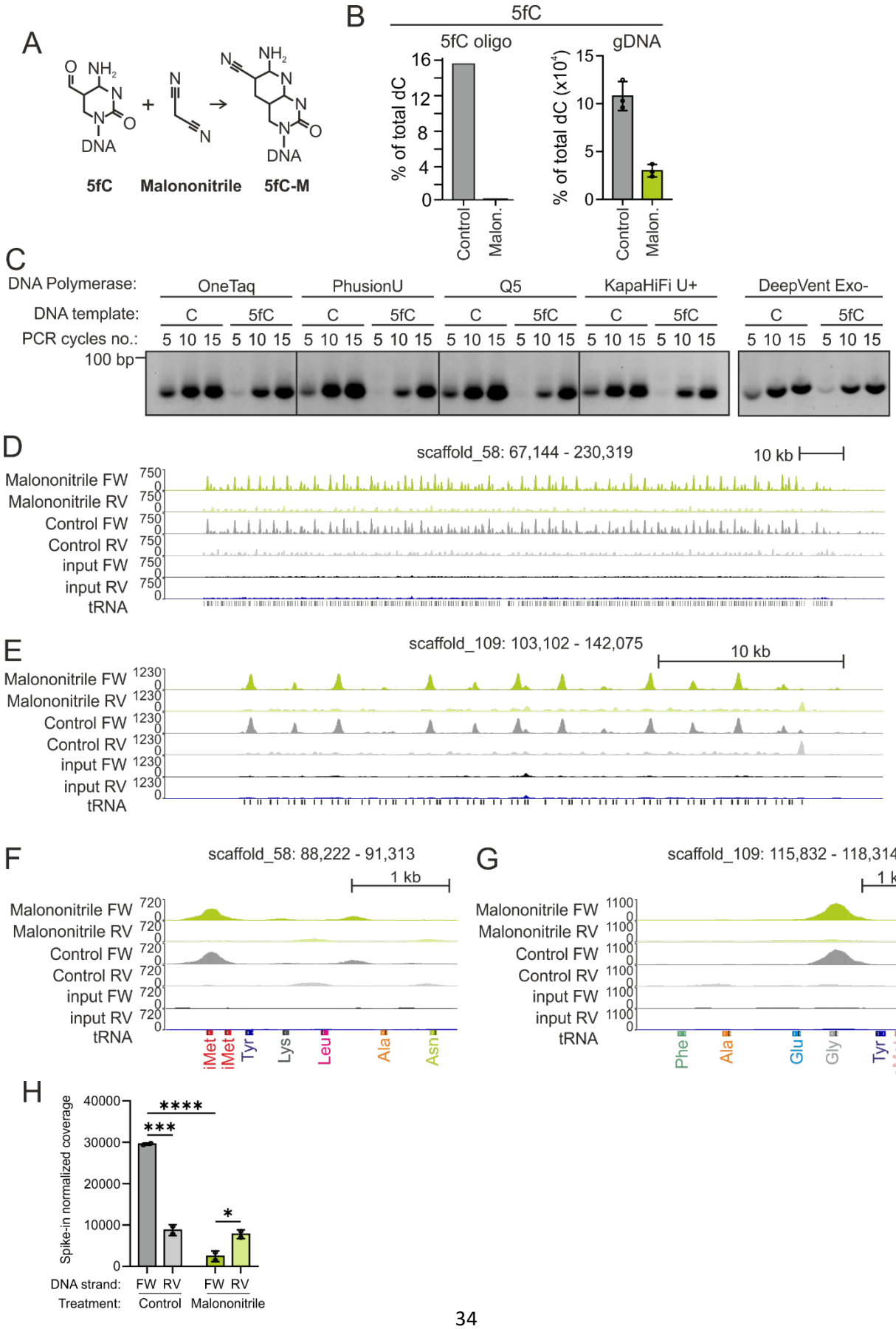

A
5fC
Malononitrile
5fC-M
B
5fC
5fC oligo
gDNA
% of total dC
% of total dC (x10^4)
Control
Malon.
C
DNA Polymerase:
OneTaq
PhusionU
Q5
KapaHiFi U+
DeepVent Exo-
DNA template:
PCR cycles no.:
100 bp
D
scaffold_58: 67,144 - 230,319
10 kb
Malononitrile FW
Malononitrile RV
Control FW
Control RV
input FW
input RV
tRNA
E
scaffold_109: 103,102 - 142,075
F
scaffold_58: 88,222 - 91,313
1 kb
iMet
Tyr
Lys
Leu
Ala
Asn
G
scaffold_109: 115,832 - 118,314
Phe
Glu
Gly
eMet
H
Spike-in normalized coverage
DNA strand:
Treatment:
Control
Malononitrile

Figure 2.3. No 5fC detection with malononitrile-seq.

A. Malononitrile-seq principle. Upon malononitrile treatment 5fC is converted to 5fC-M, which is read as T after PCR. **B.** LC-MS/MS analysis of 5fC at untreated or malononitrile treated synthetic 92 bp DNA oligo bearing two 5fC-modified sites and gDNA, respectively. 5fC oligo: n=1 replicate. gDNA: n=3. Error bars show SD. **C.** Agarose gel of PCR products amplified with OneTaq, Phusion U Hotstart, Q5, KapaHiFi U+ and DeepVent (exo-) DNA polymerases, using unmodified and 5fC-modified synthetic DNA oligo template after malononitrile treatment. Products after 5, 10 and 15 PCR cycles are shown. Expected product size 92 bp. **D-G.** IGV browser screenshots of *tRNA-iMet* **(D, F)** and *tRNA-Gly* **(E, G)** clusters with stranded 5fC DIP-seq coverage tracks of ZGA (st9) *X. tropicalis* malononitrile treated or untreated control gDNA as indicated. Malononitrile and control (untreated) *X. tropicalis* st9 whole genome sequencing (WGS) data are used as input. Individual *tRNA* genes are shown at the bottom of each panel. *tRNA* gene B-box positions are indicated as black lines in each gene. **H.** Normalized read coverage for 5fC spike-in in untreated control and malononitrile treated 5fC DIP samples. Note that the forward (FW) spike-in strand carries two 5fC, while the reverse (RV) strand is unmodified. Adjusted P-values are calculated with 2-way ANOVA and corrected for multiple comparisons with Šídák's test. *=0.046, ***=0.0003, ****<0.0001. n=2 technical replicates.

I finally accomplished high quality base resolution 5fC profiles using methylation-assisted bisulfite sequencing (MAB-seq) on selected loci of *X. tropicalis* st9 gDNA (Neri et al., 2016). During MAB-seq, all unmodified cytosines in CpG context are first methylated by M.SssI CpG methyltransferase, followed by bisulfite treatment. Thus, originally unmodified C, 5mC and 5hmC will be protected from conversion and read as C after PCR, while 5fC and 5caC will be converted to uracil and read as T after PCR (Figure 2.4A). Because 5caC was not detected during *Xenopus* ZGA by LC-MS/MS (Figure 1.5A and 2.1B), we define C-to-T transitions identified during MAB-seq as 5fC sites. The advantage of this method over methods that use chemical 5fC conversion, such as malononitrile- and pyridine borane-seq, is that MAB-seq is based on the well-established and robust BS-seq, which is compatible with any uracil-tolerant DNA polymerase for target amplification and sequencing library preparation.

In collaboration with Dr. Michael Musheev, I performed MAB amplicon sequencing of *tRNA-iMet* and *tRNA-Gly* at *X. tropicalis* st9, which we previously identified as 5fC-positive by 5fC DIP (Figure 2.3D-F). We selectively amplified the forward strand of both *tRNA* loci, which was preferentially enriched for 5fC peaks in DIP (Figure 2.3D-F). Base resolution 5fC profiles matched well with 5fC DIP peaks (Figure 2.4B). All CpGs in and around the tested *tRNA* genes were formylated with 5fC levels varying from 5 to 17% of total C (Figure 2.4B-D), while *gapdh* control locus showed background C-to-T conversion signal (Figure 2.4B). More specifically, both genes in the *tRNA-iMet* locus were formylated to similar levels (Figure 2.4C). Among the five CpGs on *tRNA-iMet*, the two CpGs on and downstream of the B-box showed consistently 10-12% 5fC in both genes. An intergenic CpG among the two *tRNA-iMet* genes was also formylated to ~17% (Figure 2.4C). *tRNA-Gly* locus was very CpG-rich, with all CpGs on- and around *tRNA-Gly* being formylated at levels between 5-15% (Figure 2.4D). Note that once again the 5fC levels are whole embryo averages that may be higher in specific regions, notably in 5fC chromocenter-positive cells, which are restricted to the animal hemisphere (Figure 1.5B).

Overall, whole embryo BS-seq confirmed *tRNA* gene methylation and MAB-seq confirmed *tRNA-iMet* and *tRNA-Gly* formylation at inter- and intragenic CpGs at ZGA. 5mC and 5fC levels at these loci (st9) were 80-95% and 5-17%, respectively. However, since 5fC selectively localizes at the animal pole during ZGA, the ratio of these modifications is expected to differ along the AP-VP axis.

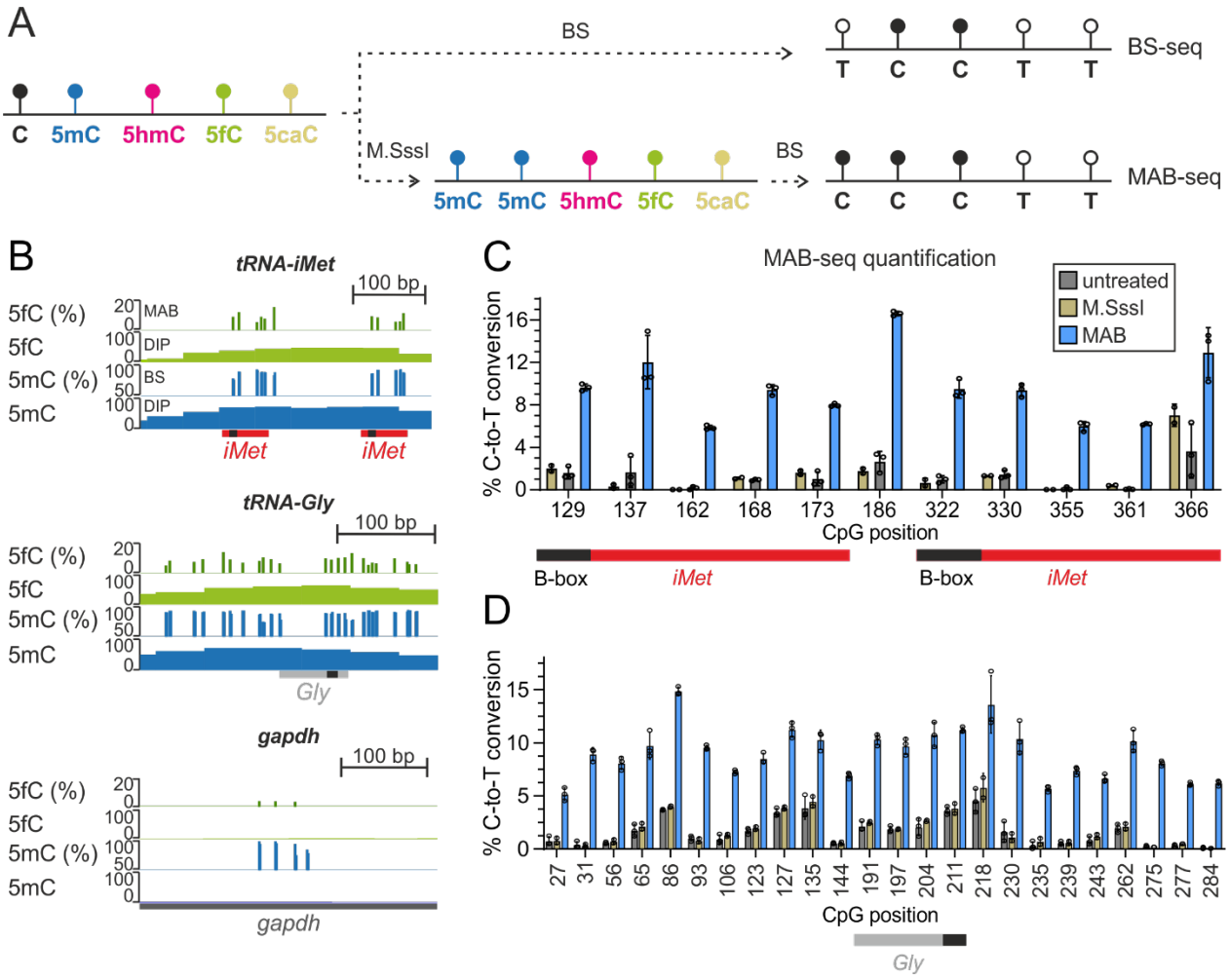


Figure 2.4. Base resolution 5fC sequencing of *tRNA-iMet* and *tRNA-Gly*.
A. Schematic representation of BS-seq versus MAB-seq. In BS-seq, 5mC and 5hmC are protected from bisulfite conversion and read as C after PCR, while unmodified C, 5fC and 5caC are converted to U and read as T after PCR. In MAB-seq, DNA is treated with M.SssI CpG methyltransferase, which methylates all unmodified cytosine in CpG context prior to BS treatment. Sequencing reads (formerly) unmodified C, 5mC and 5hmC as C, while 5fC and 5caC read as T. **B.** 5fC (green) MAB-seq, 5mC (blue) BS-seq and DIP-seq profiles at *X. tropicalis* ZGA (st9) for *tRNA-iMet* (Scaffold_58:76,642-77,045) and *tRNA-Gly* (Scaffold_109:170,885-171,190). MAB-seq shows forward strand. *gapdh* (Scaffold_7:7,145,577-7,145,923) serves as a 5fC-negative control. **C-D.** MAB-seq quantification as percentage C-to-T conversion per CpG at *tRNA-iMet* **(C)** and *tRNA-Gly* **(D)** forward strand in MAB and control (untreated and M.SssI treated) *X. tropicalis* ZGA (st9) samples. The *tRNA* genes and their respective B-boxes (black) are indicated at the bottom of the panel. Untreated, MAB: n=3; M.SssI: n=2 biological replicates. Error bars show SD.

2.3. Tdg manipulations in *Xenopus* embryos

Global genomic 5fC levels can be altered *in vivo* by manipulating the expression of the 5fC eraser, Tdg. In *Xenopus*, *Tdg* mRNA is not detected before ZGA (Owens et al., 2016; Session et al., 2016) and Tdg protein is only detected from st13 onwards (Slenn et al., 2014), suggesting that Tdg is not maternally provided and thus absent before ZGA. Accordingly, *X. tropicalis* embryos microinjected with an established *Tdg* Mo (Schomacher et al., 2016) developed normally through gastrulation and only displayed axial defects at tailbud stages (Figure 2.5A). Moreover, Tdg KD did not affect 5fC levels at ZGA, but only led to 5fC and 5caC accumulation later in development, at tailbud stages (Figure 2.5B). Collectively, the above data suggest that absence of Tdg in early development enables a transient 5fC accumulation during *Xenopus* ZGA (Figures 1.5A and 2.1B). After ZGA, Tdg expression presumably mediates 5fC removal, which in combination with further oxidation of 5fC to 5caC, result in the observed reduction in global 5fC (Figure 2.1B).

To induce early Tdg expression and global 5fC erasure, I overexpressed human TDG in *X. tropicalis* embryos by microinjecting *HsTDG mRNA*. TDG OE delayed embryonic development and induced embryonic lethality and gastrulation defects observed as a cell mass extruding from the blastopore in late gastrula (st12; Figure 2.5C-E). Importantly, overexpression of the catalytically inactive variant TDG^{N140A} (from here on referred to as dTDG; Hardeland et al., 2000) did not impede survival (Figure 2.5C-D); it rather moderately delayed development and reduced pigmentation in tadpoles (Figure 2.5C, E-F). These data support that the developmental defects observed upon early TDG OE are linked to its catalytic activity. We confirmed global 5fC reduction in ZGA at TDG OE embryos by LC-MS/MS. More specifically, while 5mC levels remained unchanged, TDG OE reduced global 5fC levels ~50% at st9 (Figure 2.5G). As expected, catalytically inactive dTDG did not affect global 5fC levels at st9 when overexpressed to similar or higher levels (Figure 2.5H-I).

Overall, Tdg is absent during early *Xenopus* development. Premature Tdg expression results in global 5fC reduction during ZGA, developmental delay, exogastrulation and embryonic lethality.

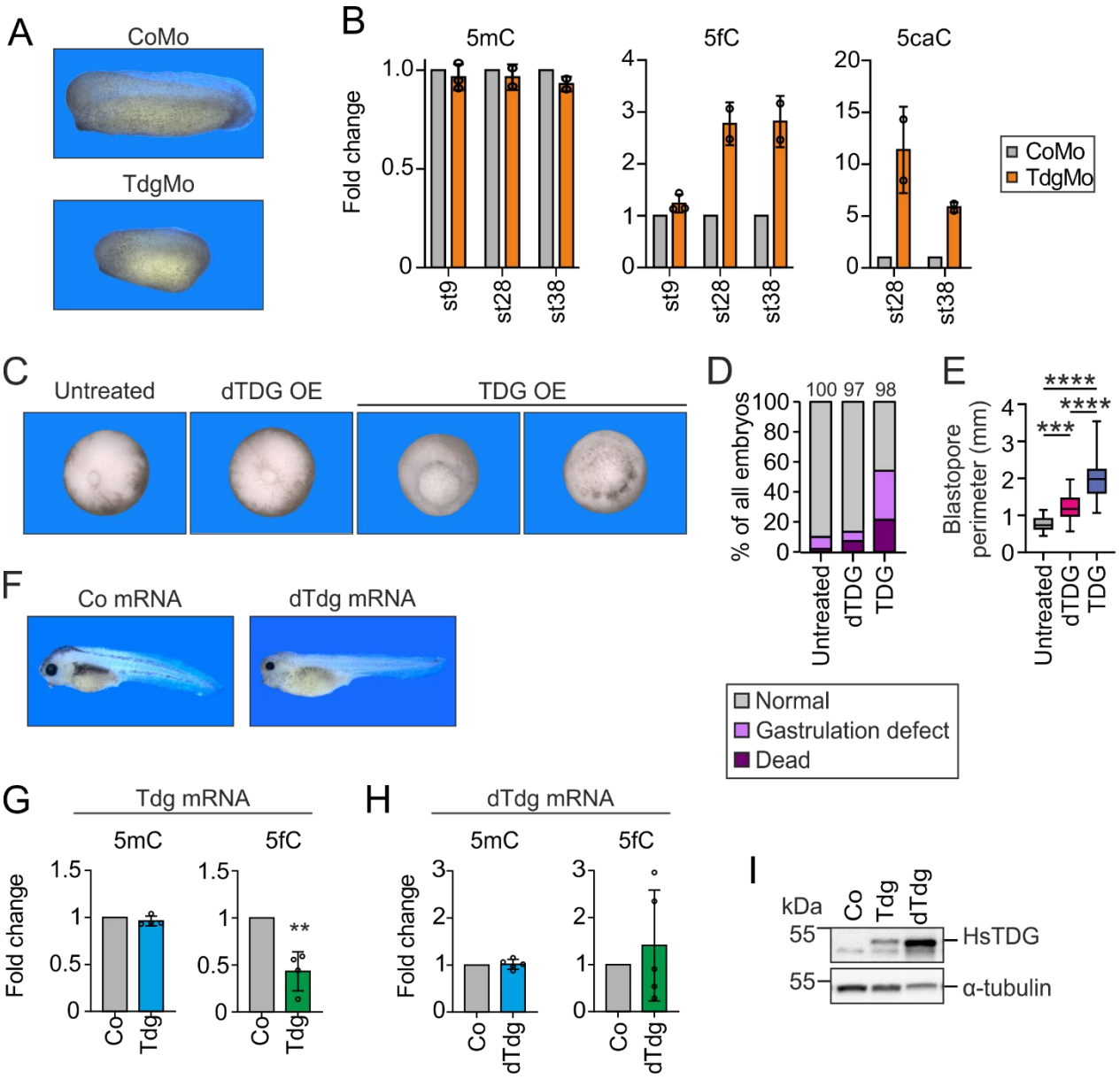


Figure 2.5. TDG OE impedes development progression and embryo survival.

A. Phenotypes of st27 *X. tropicalis* embryos microinjected with control or *Tdg* Mo. **B.** 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. Error bars show SD. **C.** Gastrula stage *X. laevis* embryos in vegetal view injected with *TDG*- or catalytically inactive *dTDG* mRNA. **D.** Phenotype scoring of gastrula stage *X. laevis* embryos injected with *TDG*- or *dTDG*- mRNA or uninjected control. The number of embryos scored is specified on the top. **E.** Blastopore perimeter in mm of gastrula stage *X. laevis* embryos injected with *TDG*- or *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. Error bars show SD. **F.** Phenotypes of *X. tropicalis* tadpoles microinjected with control (*Ppl*) or *dTDG* mRNA. **G.** 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. Error bars show SD. **H.** 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. Error bars show SD. **I.** Immunoblot from ZGA (st9) *X. tropicalis* embryos injected with control (*Ppl*), *TDG* or *dTDG* mRNA. Expected protein size, TDG: 46 kDa, α-tubulin: 50 kDa. α-tubulin, loading control.

2.4. 5fC promotes Pol III binding during ZGA

Since 5fC marks Pol III-bound *tRNA* gene tandem repeats at ZGA (Figure 1.5G-K), I next tested whether the 5fC mark promotes Pol III binding at this stage. Previous work from Dr. Victoria Hatch, showed that Tet2/3 Mo decreased Pol III binding to its target genes at ZGA (st9), suggesting that 5fC might be required for Pol III binding (Figure 1.6C). However, Tet enzymes also have catalytic activity-independent roles in transcription (see section 1.1.3). Furthermore, reduced Pol III binding upon Tet KD could suggest a requirement for DNA demethylation, rather than 5fC, for Pol III binding to its targets during ZGA. To distinguish between 5fC and DNA demethylation promoting Pol III binding at ZGA, I took advantage of the TDG OE system, which *induces* DNA demethylation but *reduces* global 5fC levels. I microinjected 1-cell stage *X. tropicalis* embryos with *TDG* and *dTDG mRNA* and performed Pol III ChIP-qPCR at ZGA (st9). I found that, under control consitions, Pol III selectively binds 5fC-rich targets during ZGA, as Pol III binding to the non-modified *tRNA-Cys* was at background levels, comparable to its binding to Pol II (*Sox2, Tbxt*) and Pol I (*5.8S*) targets (Figure 2.6A, grey bars). TDG OE halved Pol III binding to 5fC-rich Pol III genes (Figure 2.6A, magenta bars). Of note, dTDG OE did not affect Pol III binding (Figure 2.6A, blue bars), showing the effect of TDG OE on Pol III binding is connected to its catalytic activity. This key set of experiment proves that 5fC, and not DNA demethylation, promotes Pol III binding at ZGA. Moreover, the effect of 5fC on RNA polymerase binding is specific to Pol III, as Pol II binding was not affected by TDG OE at st9 embryos (Figure 2.6B). Collectively, the above experiments provide direct evidence that 5fC is required for Pol III binding during *Xenopus* ZGA.

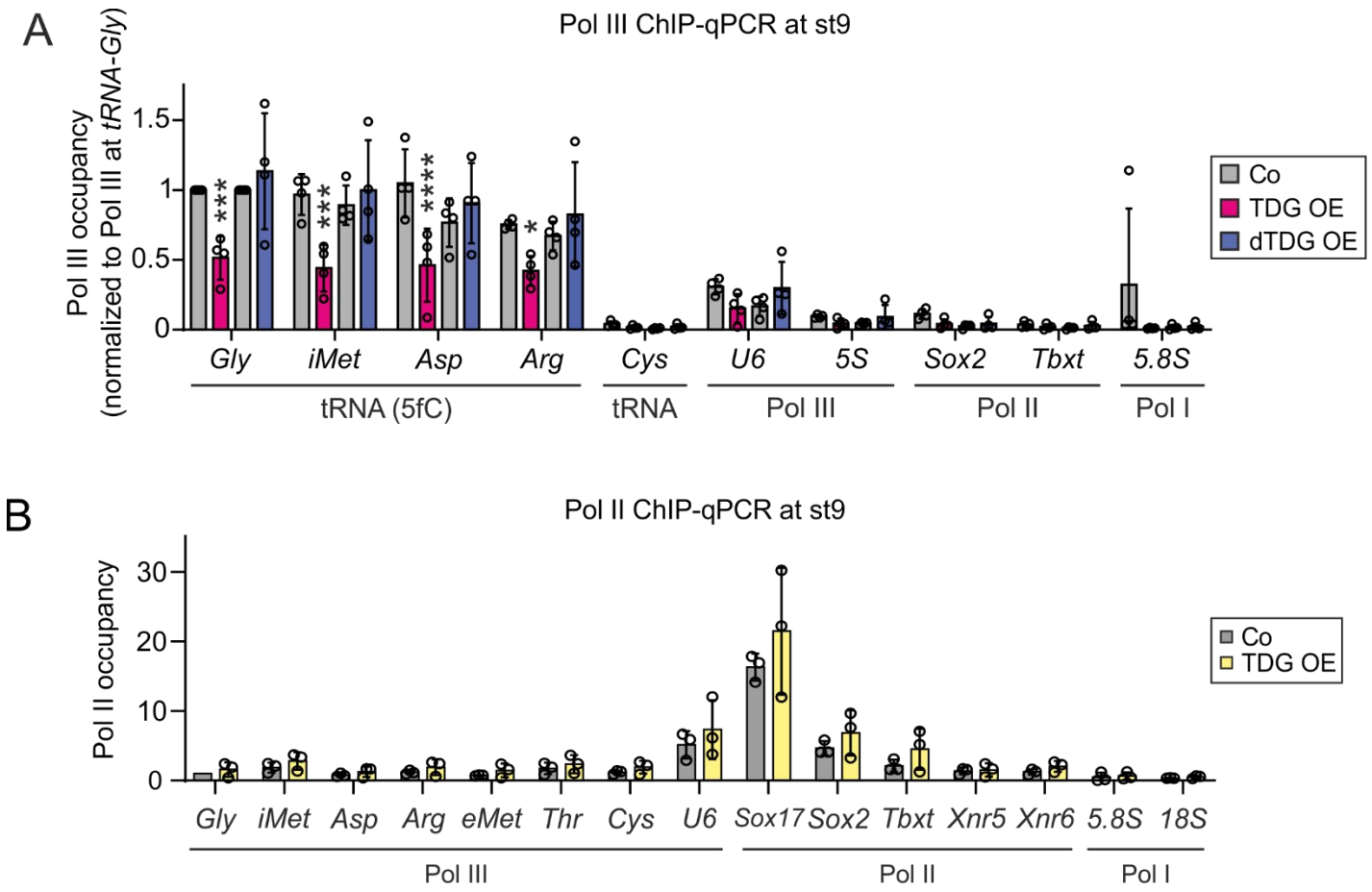


Figure 2.6. 5fC promotes Pol III binding at *Xenopus* ZGA.

A. 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. Error bars show SD. P-values are calculated with 2-way ANOVA and corrected for multiple comparisons with Tukey's test. **B.** Pol II ChIP-qPCR at ZGA (st9) *X. tropicalis* embryos microinjected with control (*Ppl*) or *TDG* mRNA. n=3 biological replicates. Error bars show SD. All comparisons between control and *TDG* mRNA embryos are non-significant, as calculated with 2-way ANOVA.

2.5. Identification of potential 5fC effectors

We next aimed to characterize the mechanism by which 5fC promotes Pol III binding during *Xenopus* ZGA. We hypothesized that 5fC might attract transcription regulators or directly induce Pol III transcriptional machinery binding. To test this hypothesis, I collaborated with Dr. Lars Schomacher, this lab, to perform a 5fC DNA oligo pull down with *X. laevis* ZGA (st9) nuclear protein extracts coupled with proteomics mass spectrometry. We used a biotin labelled, homo-modified dsDNA 30-mer carrying three 5fCs each within CpG context, as well as an unmodified control (Figure 2.7A). Because *X. laevis* genome is poorly annotated, the identified proteins were re-annotated using the PHROG database to match their human homologues (http://kirschner.med.harvard.edu/tools/mz_ref_db.html). We identified 685 proteins present in at least 2 out of 4 technical replicates. Among them, 43 proteins showed at least 1.5-fold differential binding, irrespective of p-value (11 proteins enriched and 32 proteins depleted for the 5fC oligo) (Figure 2.7B, Table 1). Only 12 proteins were significantly enriched and 6 proteins were significantly depleted for 5fC binding (Table 1).

The top 5fC binder was cryptochrome 1 or 2 (Cry1/2; FC=-5.8, p-value=$6.11x10^{-5}$), followed by DNA ligase 3 (Lig3; FC=-4.5, p-value=$2.7x10^{-5}$) (Figure 2.7B). DNA helicase Ascc3 (FC=-1.8, p-value=0.06), DNA polymerase ε catalytic subunit α (Polε; FC=-2.2, p-value=0.15), Pol II transcription factor Tfiih subunit 4 (Gtf2h4; FC=-1.8, p-value= 0.16) and SWI/SNF chromosome remodeling complex ATPase subunit (Smarca4/Smarca2; FC=-1.7, p-value=0.06) were enriched for 5fC binding, although not significantly. Enrichment of the DNA repair factors Atm (FC=-1.7, p-value= 0.04) and Lig3 suggest that the 5fC oligo was targeted for DNA repair. Tdg was not enriched among the 5fC binders, consistent with Tdg being absent before ZGA in *Xenopus* (Slenn et al., 2014). Pdcd11, an Nf-κb binding protein that co-localizes with U3 snRNA at the nucleolus, was uniquely enriched for 5fC oligo binding (FC=-3.3, p-value=0.03), suggesting that the 5fC oligo might promote association with the nucleolus. Notably, we did not find Pol III or any of its TFs among the identified proteins. This is an important limitation of this experiment, as it shows that our oligo did not attract the Pol III machinery.

The proteins depleted from the 5fC oligo might also be of interest, as they are presumably repelled by 5fC. In this scenario, 5fC enables Pol III transcription at ZGA by repelling a transcriptional repressor. Among the identified proteins, the ribosomal protein L7 (Rpl7; FC=4.5, p-value=0.04) showed the highest fold change, followed by cytochrome c oxidase (Cox7a2; FC=2.4, p-value=0.05) (Figure 2.7B). The most significantly enriched factor was Gemin4 (FC=1.8, p-value=6.5×10^{-5}), which is a “survival of motor neurons” (SMN) complex component, localized at

the cytoplasm and nuclear bodies known as Gems (Gemini of Cajal bodies). Another SMN complex component, Gemini2, was enriched for unmodified oligo binding, although not significantly (FC=1.7, p-value=0.1). No known chromatin factors and remodelers were identified.

We selected Cry1/2, the top 5fC binder, and Smarca2/4, known regulators of transcriptional activation in early development (see section 2.6.2), for further validation.

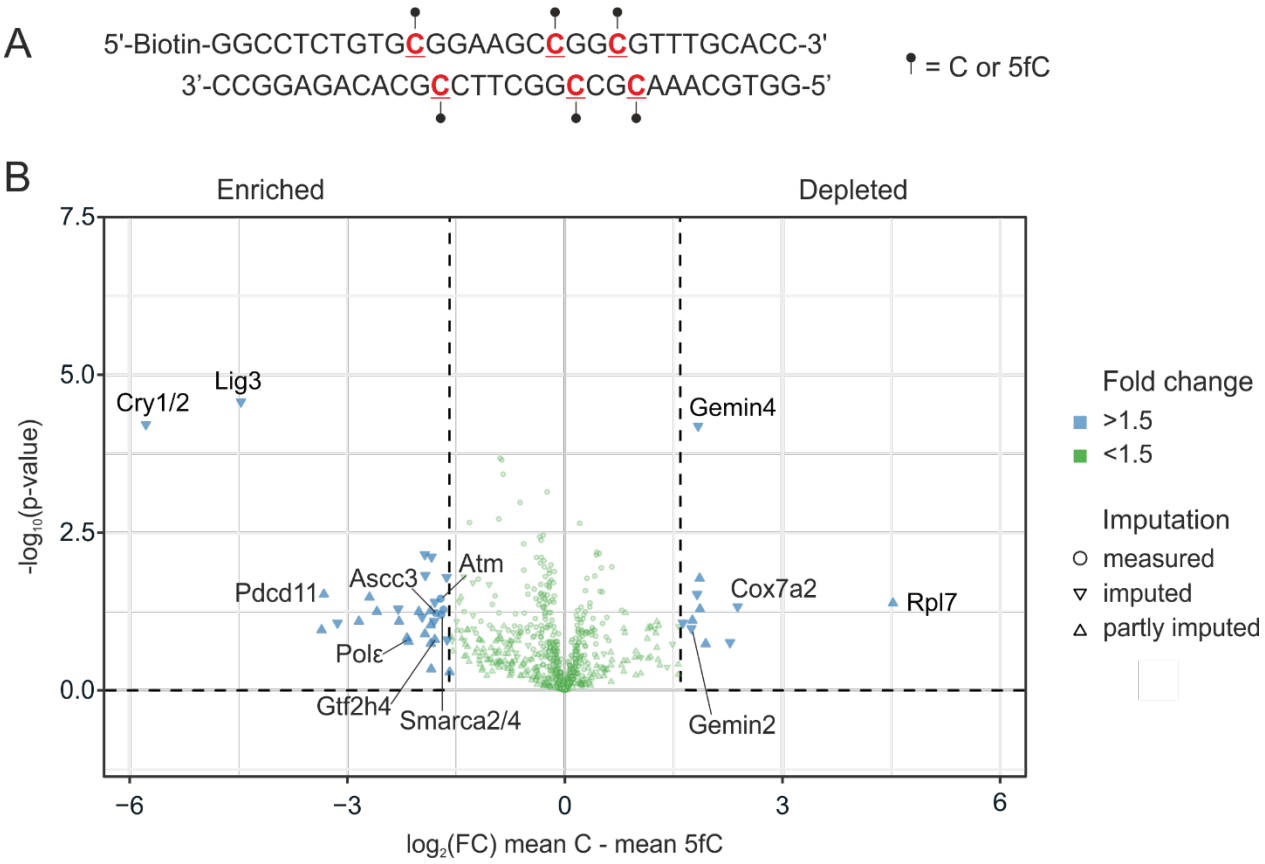


Figure 2.7. Oligo pull down for the identification of potential 5fC effectors at *Xenopus* ZGA.
A. Oligo sequence used in pull down experiment with *X. laevis* st9 (ZGA) nuclear protein extracts. The top strand is 5' biotin labelled. Red cytosines were either unmodified (C) or 5fC in both strands. **B.** Volcano plot of proteins enriched (left) and depleted (right) for the 5fC oligo. n=4 technical replicates. Proteins with fold change >1.5 are marked in blue. Measured: protein was measured in all replicates, in both conditions. Imputed: protein was measured only in one condition, in 2-4 replicates. Partly imputed: protein was measured at least once or twice in both conditions.

Enriched for 5fC oligo		
Gene name	**P-value**	**Fold change**
Lig3	**2.68E-05**	**-4.466263858**
Cry1;Cry2	**6.11E-05**	**-5.776968635**
Smn1	**0.007**	**-1.931361375**
Cul9	**0.008**	**-1.838925144**
Dync1h1	**0.015**	**-1.924382795**
Bckdk	**0.016**	**-1.630430061**
Pdcd11	**0.030**	**-3.322649439**
Ints1	**0.034**	**-2.694488387**
Atm	**0.035**	**-1.714277465**
Hal	**0.040**	**-1.794426641**
Srp68	**0.051**	**-2.298318286**
Nbas	**0.053**	**-1.675949677**
Cenpe	0.055	-1.852310079
Myh9;Flj00279	0.056	-2.591881019
Mycbp2	0.057	-2.013663845
Ascc3	0.060	-1.777553429
Smarca4;Smarca2	0.064	-1.705172761
Urb1	0.069	-1.968048297
Mcoln2	0.080	-1.802168431
Gemin5	0.081	-2.839599274
Dnajc13	0.081	-2.28647039
Uba52;Ubc;Ubb;Ubbp4	0.086	-3.135043834
Rif1	0.092	-1.849729099
Rpl6	0.110	-3.354624124
Ap3d1	0.128	-1.929370551
Pole	0.148	-2.177371978
Gtf2h4	0.157	-1.793114366
Rpa2	0.158	-1.623674219
Opa3	0.169	-2.155053173
Camk2g;Camk2b;Camk2d;Camk2a	0.181	-1.848489212
Rp11-632c17__A.1-001	0.460	-1.843209625
Rpl37	0.515	-1.588870983
Depleted from 5fC oligo		
Gene name	**P-value**	**Fold change**
Gemin4	**6.47E-05**	**1.836555801**
Afg3l2	**0.017**	**1.857765**
Nup107	**0.030**	**1.822732046**
Rpl7	**0.042**	**4.523368168**
Cox7a2	**0.047**	**2.381704094**
Dsp;Dsp variant protein	**0.051**	**1.861829641**

Ogdhl;Ogdh	0.078	1.757472868
Csnk2b	0.085	1.626662676
Gemin2	0.104	1.743590323
Eno1;Eno3	0.174	2.276563896
Nadh1	0.183	1.941013366

Table 1. List of proteins enriched or depleted for 5fC oligo binding.
Proteomics LC-MS/MS analysis of *X. laevis* st9 (ZGA) nuclear protein extract oligo pull down. Fold change >1.5. Significantly changed candidates are highlighted in bold.

2.5.1. Cry1/2 do not preferentially bind 5fC-modified sequences

The cryptochromes are transcriptional repressors that reinforce the negative feedback loop of the circadian clock (Lin & Todo, 2005). As a component of the circadian clock, the cryptochromes regulate the oscillating transcription of more than 40% of all protein-coding genes and many non-coding RNAs (ncRNAs) in mammals, either directly or indirectly (Zhang et al., 2014). To test whether Cry1 and Cry2 preferentially bind 5fC modified *tRNA* genes, I performed fluorescence polarization (FP) assay (Anderson et al., 2008). I used recombinant mouse Cry1 (mCry1) and human CRY2 (hCRY2), which were a kind gift from Prof. Eva Wolf, University of Mainz (Figure 2.8A). As a ligand, I used 28 bp double strand oligos that cover part of the *tRNA-iMet* gene sequence, including the internal promoter B-box (Figure 2.8B). Two cytosines in CpG context, inside and next to the B-box, were unmodified, methylated, hydroxymethylated or formylated, in a homo- or hemi-modified fashion (Figure 2.8B). The forward oligo was 5' fluorescein isothiocyanate (FITC) labelled for fluorescence detection and the reverse oligo was 5' biotin labelled for other applications (Figure 2.8B).

I first determined the optimum oligo concentration for the assay, by measuring the fluorescence of oligo serial dilutions between 2.5 and 80 nM in the absence of protein (Table 2). The lowest oligo concentration that consistently gave stable fluorescence signals for all different oligos was 10 nM (Table 2), so I chose this condition for further experiments. I next sought to calculate the equilibrium dissociation constant (K_D; protein concentration at which half the ligand is bound) for mCry1/hCRY2 and each oligo. I titrated the amount of protein, using a series of 3-fold dilutions from 1 µM to 1 nM, keeping the oligo concentration stable at 10 nM (Figure 2.8C-D). mCry1 did not show increased affinity for any of the tested substrates (Figure 2.8C). However, it is important to note that none of the DNA oligos reached binding saturation under the conditions used (the curves have no plateau; Figure 2.8C), so K_D values were not calculated. Similar to mCry1, hCRY2 bound all substrates with similar affinity (K_D between 112 and 147 nM), with a slight preference

for the hemi-hydroxymethylated oligo, which had the lowest K_D (K_D=112 nM) (Figure 2.8D). Overall, *in vitro* ligand binding experiments did not confirm increased cryptochrome affinity for 5fC-modified dsDNA. mCry1 and hCRY2 binding affinity to single strand (ss) DNA was not tested.

	Fluorescence Polarization [mP]						
Oligos	**no oligo**	**2.5 nM**	**5 nM**	**10 nM**	**20 nM**	**40 nM**	**80 nM**
C	190.0	65.7	57.0	56.3	57.3	60.7	65.7
5mC	170.3	59.7	56.7	57.0	60.0	62.7	66.3
C-5mC	156.3	53.7	52.3	53.0	55.0	61.3	65.7
5hmC	157.0	57.3	52.0	55.3	58.3	64.7	67.3
C-5hmC	166.7	56.7	49.0	51.0	54.7	57.7	61.3
5fC	175.3	51.0	45.0	41.7	46.3	49.3	53.7
C-5fC	169.3	54.7	45.3	44.7	47.0	51.7	51.7

Table 2. Fluorescence polarization values of oligo titration in the absence of protein.
FP signal of homo-modified (C, 5mC, 5hmC, 5fC) and hemi-modified (C-5mC, C-5hmC, C-5fC) dsDNA oligos was measured in the absence of protein to determine the optimum concentration for subsequent experiments. Note that FP values show stable increase for 10 nM oligo, or higher, in all cases. mP: millipolarization, arbitrary units.

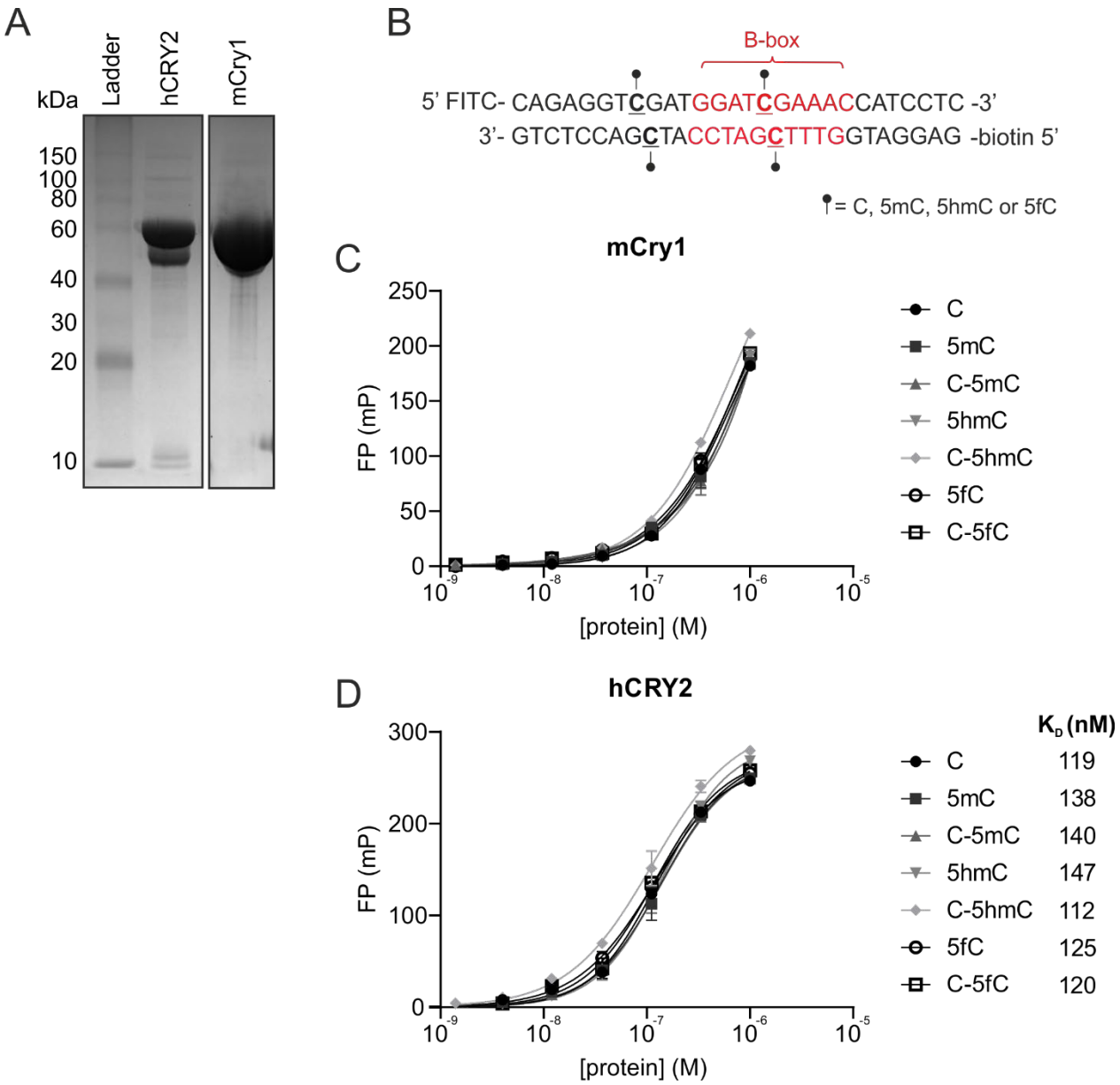


Figure 2.8. mCry1 and hCRY2 do not show increased binding affinity for 5fC modified dsDNA. **A.** SDS-PAGE of recombinant hCry2 and mCry1 with Coomassie staining. **B.** Oligo sequence used in FP experiments. B-box is marked in red. Bold and underlined cytosines were either unmodified (C), 5mC, 5hmC or 5fC. Top strand is 5' FITC labelled. Bottom strand is 5' biotin labelled. **C-D.** Fluorescence polarization of mCry1 **(C)** and hCRY2 **(D)** with homo-modified (C, 5mC, 5hmC, 5fC) and hemi-modified (C-5mC, C-5hmC, C-5fC) dsDNA oligos. Note that the binding curves in **(C)** do not reach a plateau, thus K_D values were not calculated. K_D values in **(D)** were calculated using the Hill equation. n=3 technical replicates.

2.5.2. Smarca2/4 are not required for *tRNA-iMet* transcription in ZGA

Smarca2 and Smarca4 (also known as Brm and Brg1, respectively) are the two mutually exclusive catalytic subunits of the ATP-dependent chromatin-remodeling complex Switch/Sucrose-Nonfermentable (SWI/SNF), also known as BAF (Brg1/Brm associated factor) complex. They are ATPases that use energy obtained from ATP hydrolysis to slide and/or remove nucleosomes and alter chromatin structure (Kwon et al., 1994; Wang et al., 1996). Whole proteome analysis during *X. tropicalis* and *X. laevis* embryogenesis revealed that Smarca4 is maternally provided in both species and gradually increases during development (Figure 2.9A). Smarca2 was mostly expressed upon ZGA in *X. tropicalis*, while we did not detect the protein in *X. laevis* embryos, possibly due to technical reasons (Figure 2.9A).

The SWI/SNF complex is crucial for transcription regulation during embryonic development. Smarca4 (Brg1) is required for mouse ZGA, as embryos from Brg1-depleted oocytes arrest at 2-cell stage and show reduced gene expression of ~30% of Pol II transcribed genes (Bultman et al., 2006). In *Xenopus*, Smarca4 is required for the expression of genes with the highest transcriptional activity burst in ZGA and Smarca4 Mo results in embryonic lethality during gastrulation (Wagner et al., 2017). Furthermore, SWI/SNF was shown to cooperate with Pol II (Brahma & Henikoff, 2024; Wilson et al., 1996) and Pol I (Zhang et al., 2013) to promote gene expression, while evidence for a similar role in Pol III transcription is missing. Since Smarca2/4 binding was enriched for the 5fC oligo in our pull down (Figure 2.7B and Table 1), I hypothesized that they bind 5fC-modified Pol III target genes to induce chromatin remodeling and facilitate Pol III binding during *Xenopus* ZGA. If this is true, Smarca4 (and/or Smarca2) should selectively bind 5fC-modified over non-5fC modified Pol III genes at ZGA, mimicking Pol III occupancy (Figure 1.6C and 2.6A). Moreover, this pattern should be specific to ZGA and absent at later stages.

To test this hypothesis I performed Smarca4 ChIP-qPCR at st9 (ZGA) and st13 (neurula). I first tested Smarca4 antibody specificity by western blot. I tested two independent α-Smarca4 antibodies, a commercial Proteintech antibody and a *Xenopus*-specific antibody (Wagner et al., 2017), kind gift from Prof. Ralf Rupp, Ludwig-Maximilans University (Figure 2.9B-C). Both antibodies gave specific Smarca4 signal in mESCs and *X. tropicalis* embryo protein extracts (Figure 2.9B-C). Western blot showed that Smarca4 expression was relatively stable throughout *X. tropicalis* development, confirming our proteomics findings (Figure 2.9A-C). I selected the Proteintech antibody for subsequent ChIP experiments, as it was more concentrated and thus more compatible with the assay. Smarca4 ChIP-qPCR at ZGA (st9) showed homogeneous Smarca4 occupancy across 5fC-rich (*tRNA-Gly*, *tRNA-iMet*, *tRNA-Asp*, *tRNA-Arg*) and non 5fC-

rich (*tRNA-Thr*, *tRNA-Cys*) *tRNA* genes, as well as other Pol III (*5S*, *U6*), Pol II (*Sox2*, *Tbxt*, *Foxd3*) and Pol I (*5.8S*, *18S*) targets (Figure 2.9D). Additionally, Smarca4 occupancy remained unchanged at neurula (st13; Figure 2.9D). Note that high sample-to-sample variability and lack of a negative control for Smarca4 binding hinder the interpretation of these results. Nevertheless, the above experiments provide no evidence of preferential Smarca4 binding on 5fC-modified Pol III targets during *X. tropicalis* ZGA.

I next tested the effect of Smarca2/4 inhibition on *tRNA* gene expression. I used BRM014 (Chambers et al., 2023; Xiao et al., 2022), a selective dual Smarca2/4 inhibitor, to block Smarca2 and Smarca4 ATPase activity in *X. tropicalis* embryos. More specifically, I microinjected 1-cell stage *X. tropicalis* embryos with increasing dose of BRM014, varying from 5 nM to 5 µM, as well as DMSO control. BRM014 treatment reduced *Irx2*, *Meis3* and *FoxD3* expression at st9, which was previously reported in response to Smarca4 Mo microinjections (Wagner et al., 2017), confirming Smarca2/4 inhibition (Figure 2.9E). Among the different BRM014 doses, 5 nM and 5 µM showed increasing gene expression inhibition, while 500 nM BRM014 only had a mild effect (Figure 2.9E), suggestive of technical complications with these samples. Importantly, northern blot analysis showed no difference in *tRNA-iMet* levels between DMSO, 500 nM and 5 µM BRM014 treated *X. tropicalis* st9 samples (Figure 2.9F), suggesting that Smarca2/4 activity is not required for *tRNA* transcription at ZGA.

The above experiments confirm Smarca2 and Smarca4 expression during *Xenopus* development. However, I report no evidence that (i) Smarca4 selectively binds 5fC-rich Pol III targets, and (ii) Smarca2/4 activity is required for *tRNA* transcription during *Xenopus* ZGA. Based on that, Smarca2 and Smarca4 are likely not involved in 5fC-induced Pol III zygotic reprogramming.

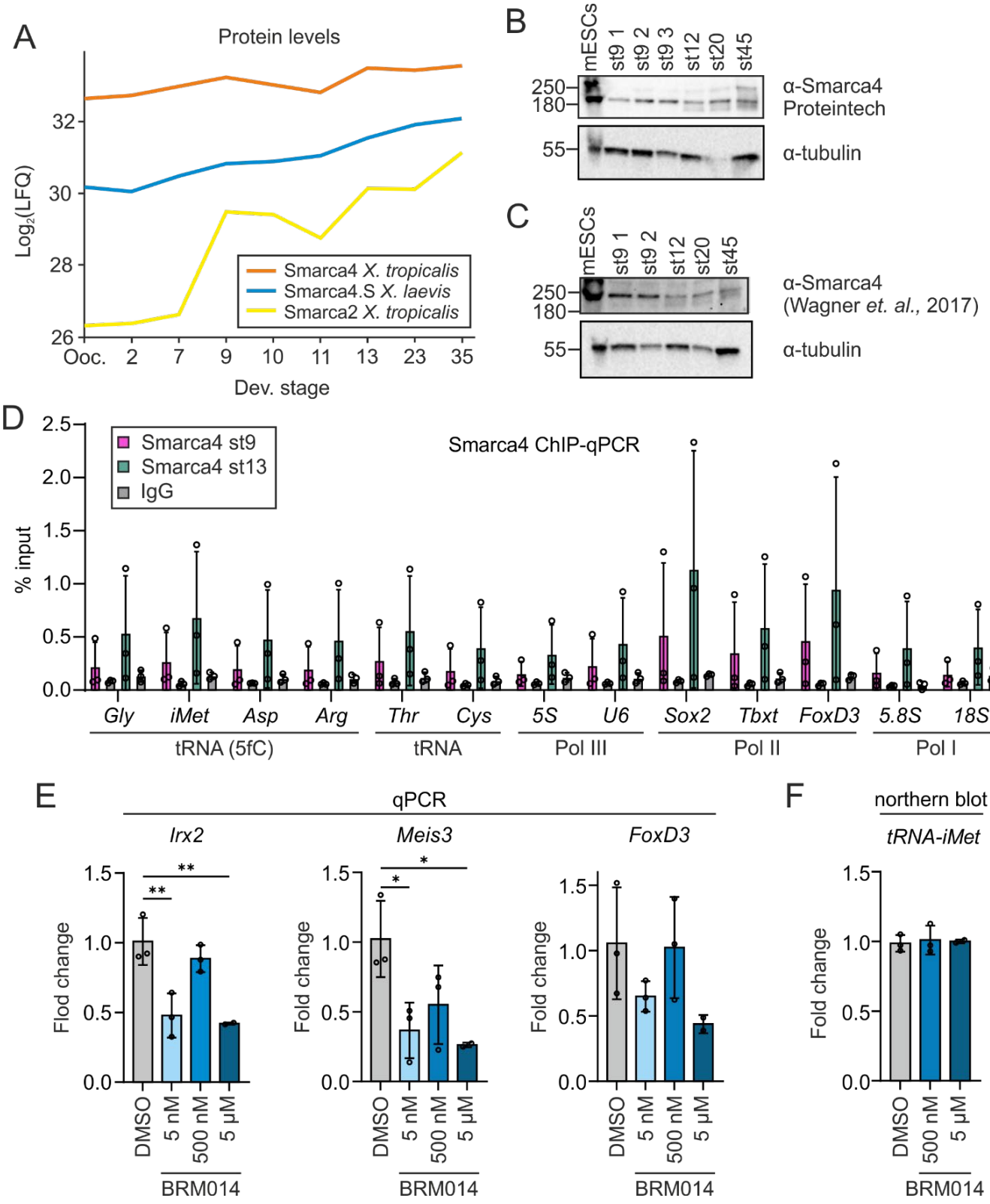

A
Protein levels
Log2(LFQ)
32
30
28
26
Ooc. 2 7 9 10 11 13 23 35
Dev. stage
Smarca4 X. tropicalis
Smarca4.S X. laevis
Smarca2 X. tropicalis
B
mESCs st9 1 st9 2 st9 3 st12 st20 st45
250
180
55
α-Smarca4 Proteintech
α-tubulin
C
mESCs st9 1 st9 2 st12 st20 st45
250
180
55
α-Smarca4 (Wagner et. al., 2017)
α-tubulin
D
Smarca4 ChIP-qPCR
Smarca4 st9
Smarca4 st13
IgG
% input
2.5
2.0
1.5
1.0
0.5
0.0
Gly iMet Asp Arg Thr Cys 5S U6 Sox2 Tbxt FoxD3 5.8S 18S
tRNA (5fC)
tRNA
Pol III
Pol II
Pol I
E
qPCR
Irx2
Meis3
FoxD3
Flod change
Fold change
1.5
1.0
0.5
0.0
**
**
*
*
DMSO 5 nM 500 nM 5 µM
BRM014
F
northern blot
tRNA-iMet
Fold change
DMSO 500 nM 5 µM
BRM014

Figure 2.9. Smarca2/4 are not required for *tRNA* gene expression during *Xenopus* ZGA.
A. Proteomics quantification of *X. tropicalis* and *X. laevis* Smarca4 and *X. tropicalis* Smarca2 during embryonic development. Values are average of n=4 biological replicates. **B-C.** Immunoblot from mESCs and st9 (ZGA), st12 (neurula), st20 (tailbud) and st45 (tadpole) *X. tropicalis* embryos, probed with Proteintech **(B)** or *Xenopus*-specific (Wagner er al. 2017) **(C)** α-Smarca4 antibodies. α-tubulin, loading control. **D.** Smarca4 ChIP – qPCR in ZGA (st9) and neurula (st13) *X. tropicalis* embryos using the Proteintech antibody. Values are shown as % input. Data are mean of n=3 independent ChIP pulldown samples. Error bars show SD. All differences are non-significant as calculated by 2-way ANOVA corrected with Šídák's multiple comparisons test. **E.** *Irx2, Meis3* and *FoxD3* RT-qPCR of ZGA (st9) *X. tropicalis* embryos microinjected with DMSO (control) and 5 nM, 500 nM or 5 μM BRM014. Gene expression is normalized to *gapdh* levels and represented as fold change over control (DMSO). DMSO, 5 nM BRM014 and 500 nM BRM014: n=3; 5 μM BRM014: n=2 biological replicates. Error bars show SD. P-value calculated with one-way ANOVA corrected with Dunnett's test for multiple comparisons. ** = 0.005, * (*Meis3* DMSO-5 nM BRM014) = 0.03, * (*Meis3* DMSO-5 μM BRM014) = 0.02. **F.** *tRNA-iMet* northern blot of *X. tropicalis* ZGA (st9) embryos microinjected with DMSO (control), 500 nM or 5 μM BRM014. *tRNA-iMet* levels are normalized over *5.8S rRNA* and represented as fold change over DMSO (control). n=3 biological replicates. Error bars show SD. All differences are non-significant as calculated by one-way ANOVA corrected with Dunnett's test for multiple comparisons.

2.6. 5fC promotes *tRNA* transcription in ZGA

2.6.1. 5fC promotes endogenous *tRNA* transcription in ZGA

I next sought to investigate whether 5fC promotes *tRNA* transcription during *Xenopus* ZGA. *tRNA* qPCR analysis was not possible because *tRNAs* are heavily chemically modified and the exact position of most of these modifications in *Xenopus* have not been characterized. For this reason, I used northern blot to quantify endogenous *tRNA-iMet* and *tRNA-Gly* transcript levels in ZGA (st9). Maternal *tRNA* transcripts are still present at ZGA, which renders assaying nascent transcript levels challenging. To estimate the fraction of maternally provided *tRNAs* at st9 I used α-amanitin to chemically inhibit Pol III transcription. α-amanitin treatment abolished *U6 snRNA* transcription at st9, a non-maternally provided Pol III target, confirming complete Pol III inhibition (Figure 2.10A). I then used northern blot to assay *tRNA-iMet* and *tRNA-Gly* levels at these conditions. Complete Pol III inhibition reduced *tRNA-iMet* levels by ~25% (Figure 2.10B, grey bars) and *tRNA-Gly* levels by ~35% (Figure 2.10C, grey bars) at st9, suggesting that ~70% of *tRNA* transcripts present at ZGA are maternally provided. I then tested the effect of *Tet2/3* KD and TDG OE on *tRNA* transcript levels around ZGA. *Tet2/3* KD significantly reduced *tRNA-iMet* and *tRNA-Gly* transcripts at st9, to levels comparable to α-amanitin induced Pol III inhibition (17% for *tRNA-iMet* and 14% for *tRNA-Gly*; Figure 2.10B-C). TDG OE also reduced *tRNA* transcripts to a similar extend, but its effect was more pronounced directly after ZGA at st10 (Figure2.10B-C). The delayed effect of TDG OE might be because Tet KD prevents the formation of 5fC, while TDG OE removes 5fC after it already formed. Importantly, overexpression of the catalytically inactive dTDG did not affect *tRNA* transcript levels at st9 and 10 (Figure 2.10D), showing that the observed reduction is due to the catalytic activity of TDG. Overall, 5fC erasure around ZGA reduces endogenous *tRNA* transcripts to a similar extend as complete chemical Pol III inhibition. These data suggest that 5fC promotes endogenous *tRNA* transcription in *Xenopus* ZGA.

To test whether 5fC also promotes the transcription of other Pol III targets apart from *tRNA* genes during *Xenopus* ZGA, I used qPCR to assay *U6 snRNA* (type III Pol III promoter) and *5S rRNA* (type I Pol III promoter) transcript levels upon Tet2/3 KD and TDG OE. There are two distinct types of *5S rRNA* genes in *Xenopus*, high copy oocyte-type *5S rRNA* that is expressed during oogenesis and upon ZGA, and low copy somatic *5S rRNA* that is expressed later in development (Harper et al., 1983; Pardue et al., 1973; Peterson et al., 1980). The sequences of oocyte and somatic 5S genes are only partially homologous, so I was able to design primers specific to the oocyte *5S rRNA*, the gene version expressed during ZGA (from here on referred to as *5S*). Interestingly, neither Tet2/3 KD nor TDG OE significantly affected *U6* and *5S* transcript levels at st9 and 10

(Figure 2.10E-F). This finding supports that while 5fC is required for *tRNA* transcription during ZGA, it is dispensable for transcription of other Pol III targets such as *U6 snRNA* and *5S rRNA*, even though these genes are also enriched for 5fC at st9 based on 5fC DIP data (data not shown). Taken together, the above data support that 5fC specifically promotes type II Pol III promoter transcription during *Xenopus* ZGA.

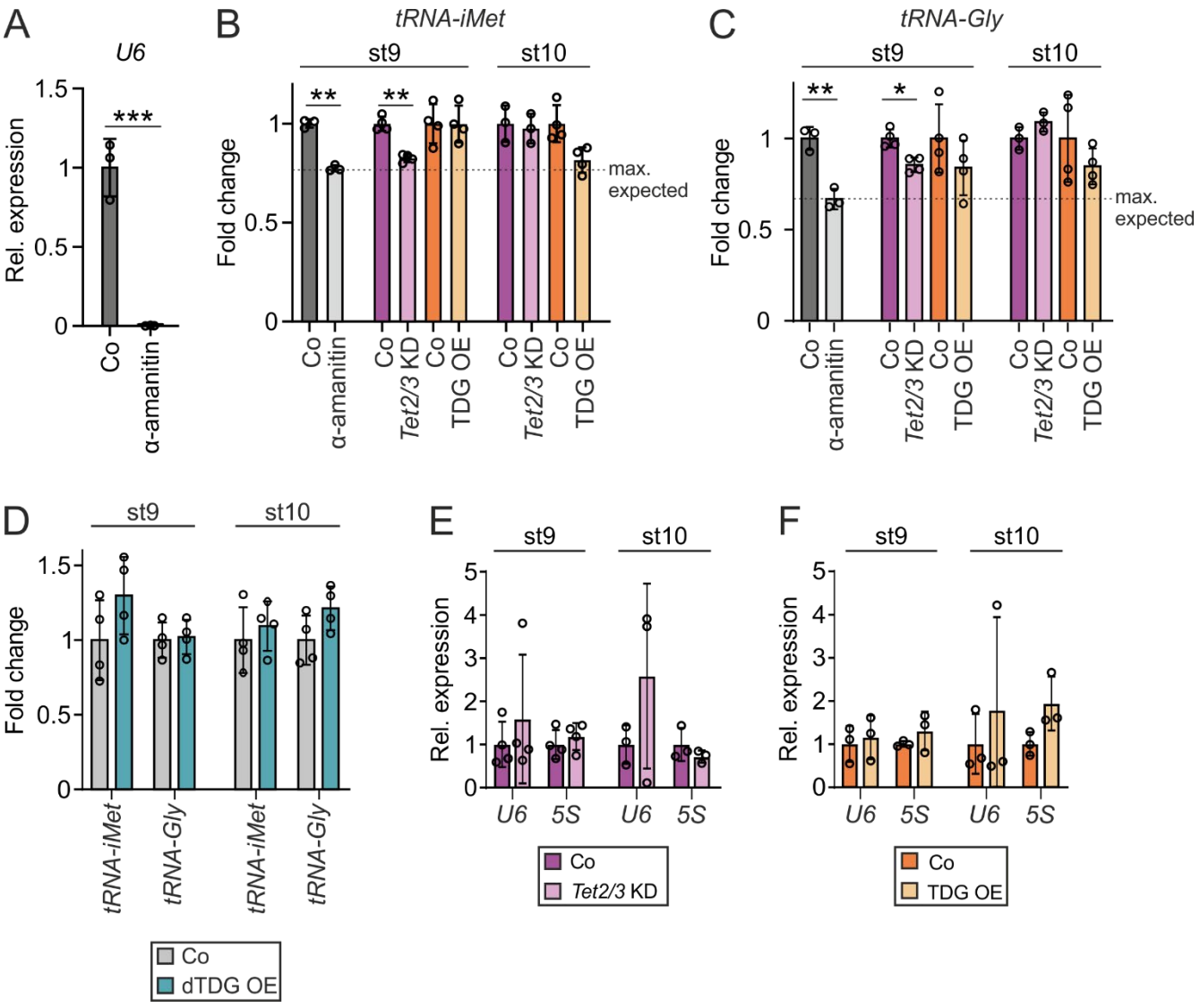


Figure 2.10. 5fC promotes type II Pol III promoter transcription in ZGA.
A. *U6 snRNA* RT-qPCR at ZGA (st9) *X. tropicalis* embryos untreated or microinjected with 2 ng α-amanitin to inhibit Pol III. *U6* expression is normalized over *5.8S rRNA* levels (Pol I target gene) and represented as fold change over the control. n=3 biological replicates. Error bars show SD. P-value calculated with unpaired t-test. **B.** *tRNA-iMet* northern blot of *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. Error bars show SD. P-values calculated with paired t-test. **C.** *tRNA-Gly*

northern blot of *X. tropicalis* embryos microinjected with α-amanitin, Tet2/3 morpholino (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. Error bars show SD. P-values calculated with paired t-test. **D.** *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. Error bars show SD. Paired t-test shows no significance for all differences between Co and dTDG OE. **E-F.** *U6 snRNA* and oocyte-type *5S rRNA* RT-qPCR of ZGA (st9) *X. tropicalis* embryos microinjected with control and Tet2/3 morpholinos (Tet2/3 KD) **(E)** or control (*Ppl*) and *TDG mRNA* (TDG OE) **(F)**. *U6* and *5S* expression is normalized to *5.8S rRNA* levels and represented as fold change over the control. n=3 biological replicates. Error bars show SD. Unpaired t-test shows no significance for all differences between Co and treatment.

2.6.2. 5fC boosts *tRNA-iMet* transgene transcription in ZGA

To provide unequivocal proof that 5fC promotes *tRNA* transcription during *Xenopus* ZGA, I established a modified *tRNA* transgene expression assay. In this assay, unmodified, 5mC-, 5hmC- or 5fC- modified *tRNA* transgenes are injected into *X. laevis* embryos at 1-cell stage, and transcript levels from the different transgenes are quantified by qPCR later in development (Figure 2.11A). I selected *tRNA-iMet* as the *tRNA* transgene, because it showed the highest percentage of 5fC DIP peaks among all *tRNA* genes (Figure 1.5J). Importantly, the transgene had to be cloned into a vector, rather than microinjected as linear DNA, as linear dsDNA microinjected into *X. laevis* oocytes is degraded within 2 h (Maryon & Carroll, 1989), and this also seemed to be the case for early embryos (own observation, data not shown).

As a first step, I sought to identify the minimum transcriptional unit required for *tRNA* gene transcription. Eukaryotic *tRNAs* are transcribed individually by Pol III (Haeusler & Engelke, 2006); however, we could not exclude that additional intergenic sequences might be required for effective *tRNA* transcription. For this reason, I prepared a construct with a 993 bp insert from the *X. tropicalis* genome, which encodes for two consecutive *tRNA-iMet* followed by *tRNA-Gly* (Figure 1.5K and 2.11B, "complete" and "no CMV' plasmids). I also prepared a version of this plasmid with a single *tRNA-iMet* gene (Figure 2.11B, "single *iMet*" and "no CMV, single *iMet'* plasmids). In all cases, the *tRNA-iMet* gene of interest was co-transcribed with 100 bp barcode added directly downstream of the *tRNA-iMet* gene, before the poly-A transcription termination sequence (Figure 2.11B, purple). This barcode would be used in the following qPCR step as a proxy for *tRNA-iMet* transgene quantification in case amplification of the *tRNA* itself was not possible (as we experienced before with endogenous *tRNAs*). Additionally, I used RLuc under a CMV promoter (Pol II target) as a normalization control to compensate for microinjection variations. CMV-RLuc

was either present on the same construct or co-injected as a separate plasmid (Figure 2.11B). This was to exclude that the *tRNA-iMet* transgene was transcribed by Pol II as a run through from the potent CMV promoter, or that ongoing Pol II transcription on the same plasmid interfered with Pol III transcription (Gerber et al., 2020). Finally, to verify that the *tRNA-iMet* transgene is indeed transcribed by Pol III, I prepared a version of each of the above plasmids where the internal promoter element B-box was mutated (Figure 2.11B, bottom and 2.11C).

I injected the different plasmids in *X. laevis* embryos and harvested total RNA at st10.5 for qPCR. I first assayed *tRNA-iMet* transgene transcription using primers targeting the barcode that is co-transcribed with the *tRNA* and I normalized over *RLuc* transcript levels (Figure 2.11D). Barcode transcript levels were overall very low. Importantly, there was no difference between wild type (wt) and mutant B-box, suggesting that the detected transcripts were not Pol III depended. Next, I used the same samples, this time targeting directly *tRNA-iMet* with specific primers and normalized transcript levels over *RLuc* (Figure 2.11E). *tRNA-iMet* levels were 20-100 fold higher than those of the barcode, suggesting that the barcode is either not transcribed or selectively cleaved and degraded. *tRNA-iMet* transcript levels from complete (CMV-RLuc, 2x *tRNA-iMet*, *tRNA-Tyr*), no CMV and single *tRNA-iMet* plasmids were not reduced as a result of B-box mutation (Figure 2.11E). Notably, the plasmid that carried a unique *tRNA-iMet* gene in the absence of CMV promoter showed a ~3 fold reduction of *tRNA-iMet* transcript levels from mutant B-box (Figure 2.11E). These data suggest that the presence of a CMV promoter (Pol II target) or neighboring B-boxes (Pol III targets) on the same plasmid can regulate the transcription of the *tRNA* gene of interest. Based on that, I selected the "no CMV, single *tRNA-iMet*" plasmid for further experiments.

Next, I sought for a method that would allow for addition of DNA modifications on selected nucleotides at the plasmid. The chosen method should allow for production of relatively large amounts of modified plasmids (tens to hundreds of nanograms). I initially tried excising the *tRNA-iMet* gene by restriction digest, gel extraction of the linearized backbone and ligation with a PCR produced *tRNA-iMet* gene synthesized with 50% 5fC dCTPs. The resulting plasmid would carry a *tRNA-iMet* gene with 50% of all cytosines randomly formylated. However, the yield was extremely low because ligation of double strand breaks is very inefficient.

To overcome the low yield problem, I switched from restriction digest/ligation to strand exchange (Lühnsdorf et al., 2012). In this assay, one strand of the target gene is selectively targeted for excision by a nicking endonuclease and replaced by a synthetic strand that carries the DNA modification of interest (Figure 2.11F). This method generally yields higher amounts of modified plasmid compared to restriction digest/ligation, because the remaining strand on the plasmid base

pairs with the synthetic strand, enabling hybridization and successful ligation. I chose to replace *tRNA-iMet* non-coding (template) strand with a synthetic strand that carries 2x 5fC in a CpG context, inside and directly downstream of the B-box (Figure 2.11G). This was based on the following observations: a) 5fC DIP at st9 selective enriched the forward strand, which coincides with the *tRNA-iMet* non-coding strand (Figure 2.3D-G). b) Motif analysis of 5fC DIP targets showed that the 5fC mark correlated with the B-box (data not shown). c) MAB-seq showed that *tRNA-iMet* B-box is indeed formylated (Figure 2.4B-C) d) B-box is the recognition element for TFIIIC, the only type II promoter-specific TF (see section 1.4.1-2).

I was able to control every step of the process based on the principle that nicked circular DNA migrates slower than covalently closed DNA in agarose gels (Figure 2.11H). Accordingly, plasmid incubation with the nicking endonuclease Nb.Bpu20I resulted in a band shift at a higher molecular weight, indicating that the plasmid was successfully nicked (Figure 2.11H, lanes 1 and 2). Upon re-ligation of the endogenous strand with T4 ligase, most of the plasmid successfully ligated, with only a small fraction remaining nicked (Figure 2.11H, lane 3). This is probably because the nicked sequence is fairly long (92 bases) which renders correct hybridization challenging. Furthermore, the excised strand corresponds to a *tRNA* gene, with the potential to form secondary and tertiary structures that would further hinder proper hybridization with the remaining strand. Incubation of the nicked plasmid with 200-fold molar excess of unmodified (C) strand without the addition of polynucleotide kinase (PNK) did not result in covalently closed product (Figure 2.11H, lane 4), because the 5' end of the synthetic strand is not phosphorylated and thus not able to form phosphodiester bond. Finally, unmodified (C), methylated (mC) and formylated (fC) synthetic strands successfully ligated to the vector after PNK incubation, producing covalently closed plasmids carrying the desired modification (Figure 2.11H, lanes 5-7). Note that I did not produce 5hmC plasmids because the production of the synthetic strand failed repeatedly. The covalently closed fraction of the strand exchange plasmids was then gel extracted and purified for microinjections.

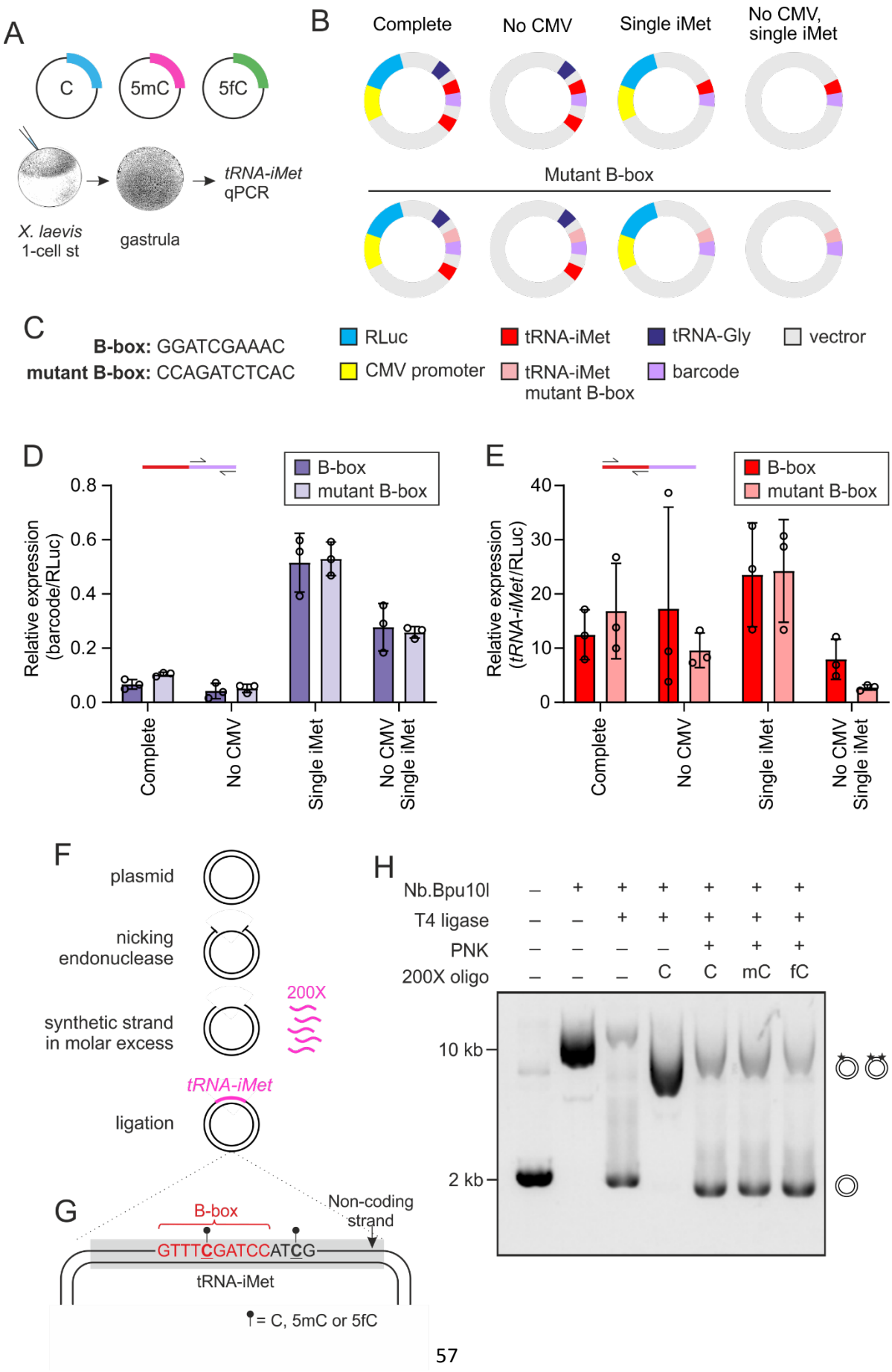

A
C
5mC
5fC
tRNA-iMet
qPCR
X. laevis
1-cell st
gastrula
B
Complete
No CMV
Single iMet
No CMV,
single iMet
Mutant B-box
C
B-box: GGATCGAAAC
mutant B-box: CCAGATCTCAC
RLuc
tRNA-iMet
tRNA-Gly
vector
CMV promoter
tRNA-iMet
mutant B-box
barcode
D
B-box
mutant B-box
Relative expression
(barcode/RLuc)
0.8
0.6
0.4
0.2
0.0
Complete
No CMV
Single iMet
No CMV
Single iMet
E
B-box
mutant B-box
Relative expression
(tRNA-iMet/RLuc)
40
30
20
10
0
Complete
No CMV
Single iMet
No CMV
Single iMet
F
plasmid
nicking
endonuclease
synthetic strand
in molar excess
200X
tRNA-iMet
ligation
G
B-box
Non-coding
strand
GTTTCGATCCATCG
tRNA-iMet
= C, 5mC or 5fC
H
Nb.Bpu10I
T4 ligase
PNK
200X oligo
C
C
mC
fC
10 kb
2 kb

Figure 2.11. *tRNA-iMet* transgene assay optimization.
A. Schematic representation of *tRNA-iMet* transgene assay. Top: Three different variants of the *tRNA-iMet* plasmid, in which selected cytosines are either C, 5mC or 5fC. Bottom: *X. laevis* embryos at 1-cell stage are microinjected with the different variants of *tRNA-iMet* plasmid. Embryos are harvested in later stages and *tRNA-iMet* is quantified by qPCR. **B.** *tRNA-iMet* plasmid variants prepared for transgene assay. Complete (CMV promoter, RLuc, 2x *tRNA-iMet*, *tRNA-Tyr*), No CMV (2x *tRNA-iMet*, *tRNA-Tyr*), Single iMet (CMV promoter, *tRNA-iMet*) and No CMV, single iMet (*tRNA-iMet*) plasmids. *tRNA-iMet* gene of interest is co-expressed with a 100 bp barcode (purple) in all plasmids. Top: *tRNA-iMet* wt B-box. Bottom: *tRNA-iMet* mutant B-box. **C.** *tRNA-iMet* B-box (wt) and mutant B-box sequences. **D.** Barcode relative expression from complete, No CMV, Single iMet and No CMV, single iMet plasmids with wt and mutant B-box at st10.5 *X. tropicalis* embryos. Barcode expression is normalized to *RLuc*. n=3 biological replicates. Error bars show SD. All comparisons between B-box and mutant B-box were non-significant, as calculated by 2-way ANOVA with Šídák's multiple comparisons test. **E.** *tRNA-iMet* relative expression from complete, No CMV, Single iMet and No CMV, single iMet plasmids with wt and mutant B-box at st10.5 *X. tropicalis* embryos. *tRNA-iMet* expression is normalized to *RLuc*. n=3 biological replicates. Error bars show SD. All comparisons between B-box and mutant B-box were non-significant, as calculated by 2-way ANOVA with Šídák's multiple comparisons test. Note, the no CMV, single iMet (*tRNA-iMet*) construct showed a 3-fold decrease in *tRNA-iMet* transcript levels with the mutant B-box. **F.** Scheme of strand exchange assay. The plasmid is nicked by a nicking endonuclease to selectively remove one strand of the target sequence (in this case *tRNA-iMet* non-coding strand). The nicked plasmid is incubated with 200-fold molar excess of the synthetic strand that carries the modification of interest (in this case unmodified C, 5mC, 5fC) and the synthetic strand is subsequently ligated to produce covalently closed, modified plasmids. **G.** Scheme of final plasmid constructs after strand exchange with two modified cytosines in the non-coding (template) strand of the *tRNA-iMet* gene. **H.** Analysis of each step of the strand exchange assay by agarose gel electrophoresis. Lane 1: untreated plasmid. Lane 2: Nb.Bpu10I nicked plasmid. Lane 3: Nicked and self-ligated plasmid. Lane 4: Nicked plasmid incubated with synthetic strand and T4 ligase in the absence of PNK. Lanes 5-7: Nicked plasmid incubated with synthetic strand and T4 ligase after treatment with PNK.

I then microinjected the strand exchange produced plasmids, along with control plasmid that did not undergo strand exchange (No SE), in *X. laevis* embryos and tested *tRNA-iMet* transcription from each plasmid at the onset (st8) and right after (st10-11) ZGA by qPCR. *tRNA-iMet* values were normalized over the amount of plasmid DNA microinjected (see section 4.2.17). *tRNA-iMet* was not produced from any plasmid at st8 (Figure 2.12A). At st10, No SE, unmodified (C) and methylated (5mC) plasmids produced *tRNA-iMet* transgene at similar levels. This is in agreement with previous reports that DNA methylation does not inhibit gene expression in early *Xenopus* development (Bogdanović et al., 2011). Importantly, 5fC induced *tRNA-iMet* transgene transcription 6-fold at st10 (Figure 2.12A). *tRNA-iMet* transcripts were not detected at st11, consistent with transient oocyte-type *tRNA* expression. Of note, 5fC modified plasmid selectively induced transgene expression at the AP, as 5fC plasmid injected at the VP did not show increased *tRNA-iMet* levels compared to unmodified plasmid (Figure 2.12B). To further corroborate the specificity of 5fC-induced *tRNA-iMet* expression, I pre-treated strand exchange plasmids with recombinant TDG to remove 5fC. LC-MS/MS verified 90% 5fC removal upon TDG treatment (Figure 2.12C-D). Accordingly, TDG treatment reduced *tRNA-iMet* expression to control levels at st10, as expected (Figure 2.12E). Collectively, these experiments provide direct proof that 5fC promotes *tRNA* transcription during *Xenopus* ZGA.

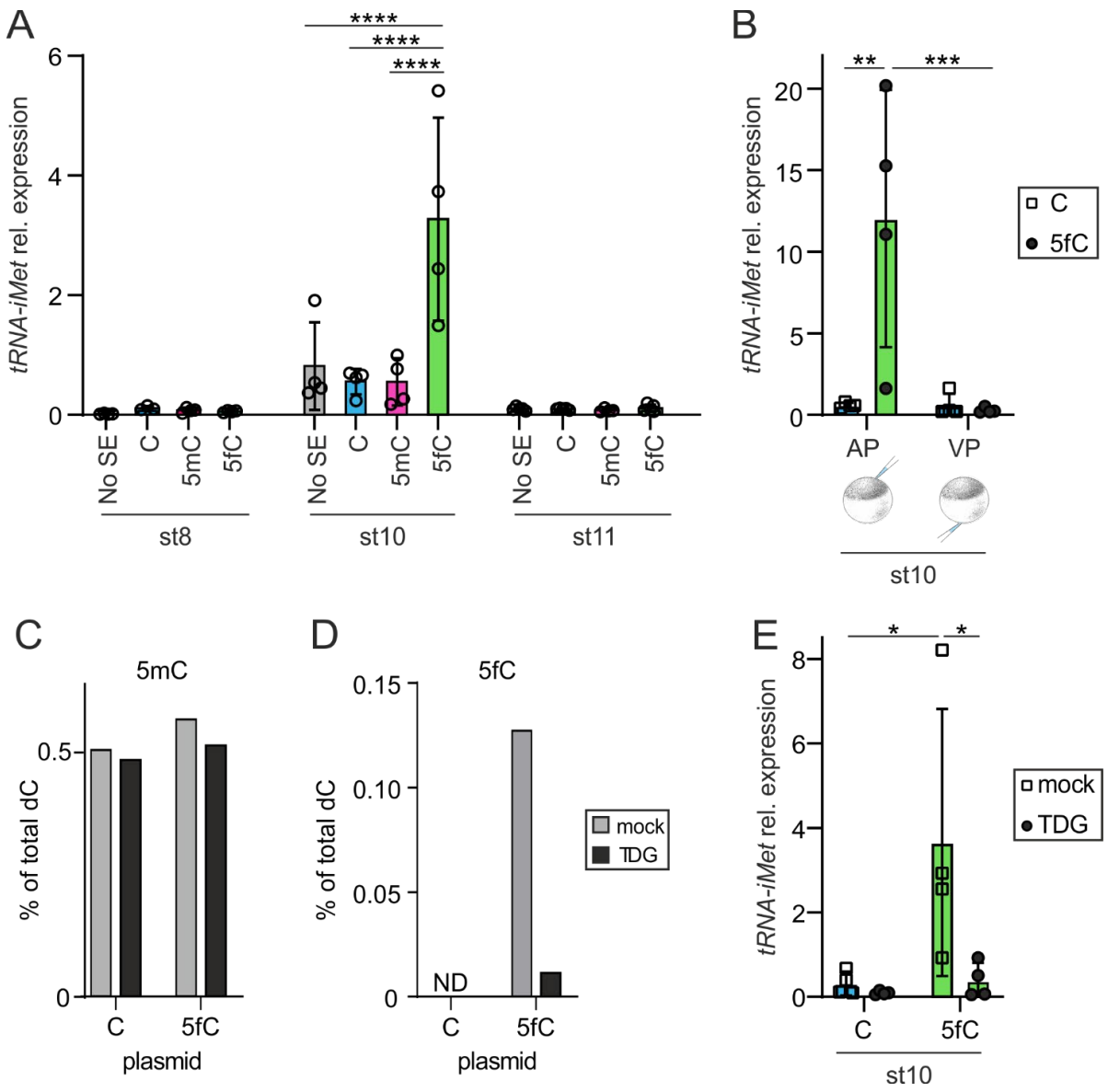


Figure 2.12. 5fC boosts *tRNA-iMet* transgene transcription at ZGA.

A. *tRNA-iMet* transgene expression analysis. *X. laevis* embryos 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'), 5mC, or 5fC modified. Embryos were harvested at st8 (early-ZGA), st10 (early gastrula) and st11 (mid-gastrula) 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 on ΔCp values (normally distributed) and corrected for multiple comparisons with Tukey's test. St8 and 10, n=4; st11, n=5 biological replicates. Each biological replicate contained between 6-12 embryos. Error bars show SD. **B.** *tRNA-iMet* expression analysis as explained in **(A)**. Embryos were microinjected with C or 5fC SE-plasmids into the animal (AP) or the vegetal pole (VP) and harvested at st10 (early gastrula), n=4 biological replicates. **C-D.** LC-MS/MS analysis of plasmids 5mC **(C)** and 5fC **(D)** levels after strand exchange with unmodified and formylated (5fC) oligos, after mock- and *in vitro* TDG treatment. Note, the 5mC levels in both plasmids are due to Dcm methylation during plasmid propagation in *E. coli*. **E.** *tRNA-iMet* expression analysis as explained in **(A)**. *tRNA-iMet* plasmids were mock- or TDG treated prior to microinjection and microinjected embryos were harvested at st10 (early gastrula), n=4 biological replicates. Error bars show SD.

3. Discussion

The main finding of this thesis is that 5fC is an activating epigenetic mark that promotes RNA Pol III binding and transcription during ZGA. My work, described in section 2, together with that of former and current members of this lab, led to the development of the following model of 5fC-mediated *tRNA* gene reprogramming during ZGA (Figure 3.1): *Xenopus* embryos are special in containing large *tRNA* tandem repeats, which are transcribed from ZGA until gastrula stage. During ZGA, Tet enzymes oxidize these hypermethylated *tRNA* loci, leading to 5fC accumulation. The 5fC mark promotes transcription by an unknown mechanism that may involve recruitment of 5fC readers or changes in chromatin conformation. At the end of gastrulation, the *tRNA* tandem repeats are silenced as DNA methylation marks become repressive and 5fC is lost upon expression of Tdg.

This is the first study to identify a physiologically relevant role for 5fC *in vivo*, namely Pol III gene expression activation at ZGA. Our work answers the long-standing question of whether oxCs are functional epigenetic marks with 5mC-independent, instructive roles.

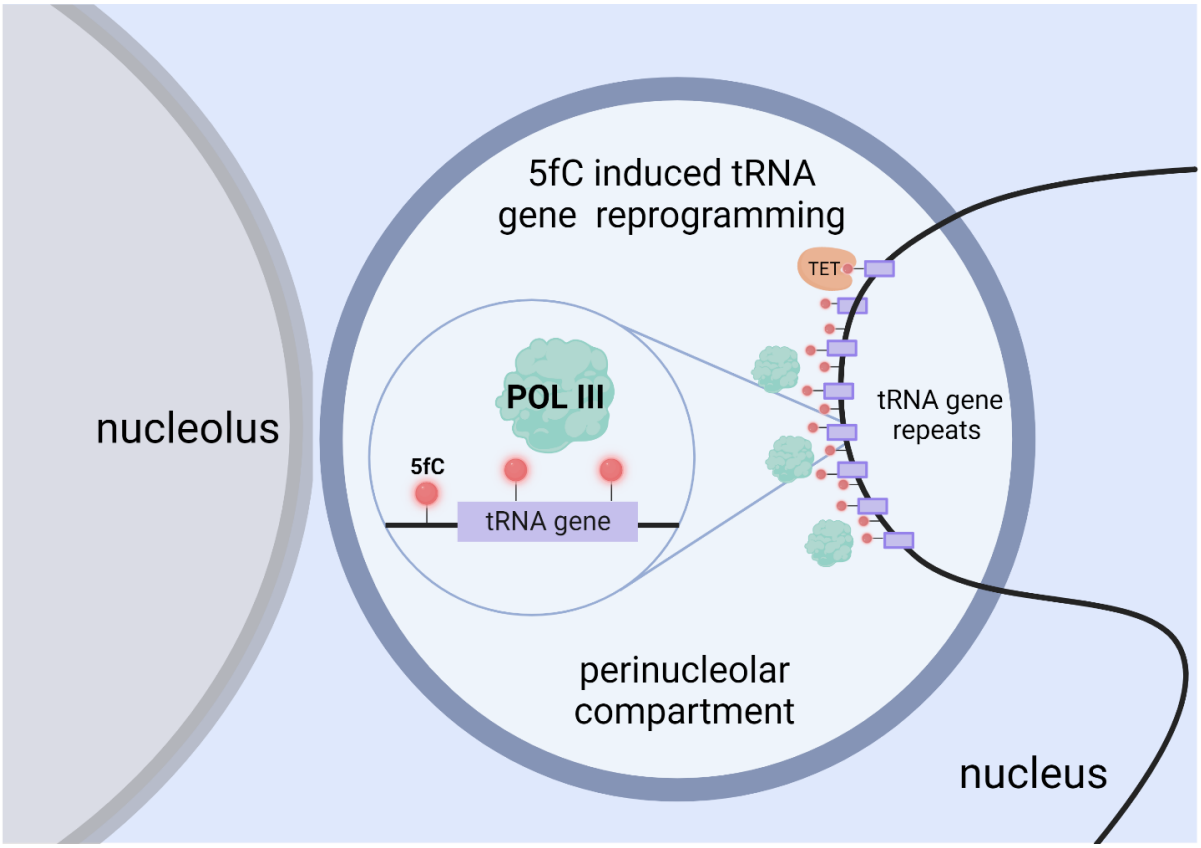


Figure 3.1. Model of 5fC activating *tRNA* transcription at *Xenopus* ZGA.
At the onset of ZGA, 5fC accumulates on *tRNA* gene tandem repeats at the perinucleolar compartment and induces Pol III binding and *tRNA* transcription.

3.1. 5fC is an activating mark

Base methylation is not an inherently epigenetic process. For example, uracil methylation at its 5th carbon produces thymine, so T is in practice 5mU. However, thymidylate synthase mediated deoxyuridine methylation to T takes place in suspension, before the nucleotide is incorporated to the DNA. Subsequently, thymine as one of the four Watson-Crick bases will code for genetic information. On the other hand, 5mC is a product of DNMT induced cytosine methylation, which takes place selectively on CpG dinucleotides in defined sequence context to inhibit the expression of a given gene. In this line, bases could be categorized in two classes: “coding bases”, A, T, C, G, which are produced in suspension before incorporating into the DNA, and “instructive bases”, represented by 5mC, which are produced by epigenetic processing in a sequence-specific manner. In this work, we provide for the first time evidence of a second “instructive base”, 5fC, which is produced by sequential 5mC oxidation by TET dioxygenases. 5fC drives transcriptional activation, specifically *tRNA* expression during ZGA, when gene expression fine-tuning is critical.

5fC accumulation during *Xenopus* ZGA at relatively high levels (Figure 1.5A and 2.1B) offered the unique opportunity to study this otherwise rare DNA modification. While 5hmC transiently accumulated right before ZGA (st6.5-8, Figure 2.1B), its levels dropped up to 3-fold during ZGA. More specifically, during ZGA, 5fC was the second most abundant DNA modification after 5mC, and up to 6-fold more abundant than 5hmC. Notably, this is the first time that 5fC levels are reported to be higher than those of its precursor 5hmC. This 5fC accumulation is due to the concurrent incidence of two conditions. First, Tdg protein is absent before neurula stages (st13) in *Xenopus* (Slenn et al., 2014), suggesting that 5fC is not actively excised during ZGA. These conditions might also be met during mouse ZGA, as *Tdg mRNA* is not detected in 2-cell stage embryos (Huang et al., 2017) and Tdg deficiency does not affect zygotic DNA demethylation (Guo et al., 2014; Santos et al., 2013). Second, 5fC is likely not further oxidized by Tets at ZGA, as 5caC was not detected before neurula (st13; Figure 2.1B). However, how Tet processivity at *tRNA* genes during ZGA is regulated, in order to oxidize 5hmC to 5fC but not 5fC to 5caC remains unclear. A unique feature of 5fC during *Xenopus* ZGA is that it accumulates at long, tandem repeats of *tRNA* genes. Accordingly, in mESCs 5fC/5caC accumulate at major satellite repeats but not at non-repetitive DNA (Shen et al., 2013). It is conceivable that the repetitive nature of these sequences attracts Tet enzymes and/or co-factors that fine-tune Tet processivity, in order for 5fC to accumulate.

The fact that TET/TDG-mediated DNA demethylation can stall at different steps along the pathway, namely at the level of 5hmC (Hu et al., 2015; Ito et al., 2011), 5fC (this study) and

potentially even 5caC, already suggests that there is more to the generation of these marks than simply being intermediates of DNA demethylation. If the only purpose of 5mC oxidation was to remove the mark, then the active DNA demethylation pathway would be a very ineffective one, as it introduces oxCs that are not further processed, despite being prone to mutagenesis and DNA damage response (Jiang et al., 2017; Mahfoudhi et al., 2016). Our work proves that the "active DNA demethylation" pathway is, at least under the given conditions, a "DNA modification turnover" pathway and that its current name reduces its function to just one of few.

While 5mC symmetry is maintained upon DNA replication via DNMT1, there are no reports of a similar mechanism for oxC maintenance. Indeed, our data show that the forward DNA strand of the *tRNA-iMet* and *tRNA-Gly* loci are asymmetrically formylated (Figure 2.3D-G). This could imply that, as opposed to 5mC that mediates stable gene silencing, oxC, and particularly 5fC, are deposited on demand to mediate a transient signal. Accordingly, 5fC transiently accumulates during *Xenopus* ZGA (Figure 2.1B) to induce *tRNA* gene expression. This would offer an alternative explanation for the quick turnover of 5fC, which has so far been viewed as evidence for it being a mere intermediate of DNA demethylation.

Our discovery that 5fC is a transcriptional activator is in agreement with its genomic signature in mESCs, where 5fC is enriched at poised and active enhancers (Iurlaro et al., 2016; Song et al., 2013; Xia et al., 2015) and transcriptionally active gene promoters bound by RNA Pol II (Neri et al., 2015; Raiber et al., 2012), suggesting that 5fC correlates with active gene expression. However, *in vitro* experiments showed 5fC to hinder Pol II transcription rate and fidelity, (Kellinger et al., 2012; Kitsera et al., 2017; Wang et al., 2015; You et al., 2014). Additionally, 5fC was found to interact with transcriptional repressors *in vitro* (Iurlaro et al., 2013; Spruijt et al., 2013), suggesting that it might interfere with gene expression. The absence of a clear, consistent role for 5fC led to the notion that there might not be one at all. Note that 5mC was also initially disregarded as a physiologically irrelevant modification due to its low abundance. Moreover, since the role of 5mC in eukaryotic DNA was initially unclear, it was suggested that 5mC could be a transcriptional activator (Comings, 1972) or redundant for embryonic development (Adams, 1971, 1973). It is conceivable that the role of 5fC in gene expression is context-dependent, given the opposing outcomes between *in vivo* and *in vitro* experiments. Overall, the role of 5fC seems to be more complicated than that of 5mC.

Under what conditions is 5fC an activating mark? *Xenopus* ZGA combines two unique conditions: 5fC accumulation in the absence of Tdg and first time expression of the embryonic genome. It is intriguing to speculate that the 5fC mark selectively dictates first time gene expression of

previously dormant genes. If this is the case, could 5fC be implicated in gene expression initiation of cell-type specific genes during lineage commitment? Indeed, 5fC is enriched at developmental enhancers in a tissue-specific manner at E11.5 mouse embryos (Iurlaro et al., 2016), suggesting that 5fC might play a role in cell-fate specification.

Overall, the discovery of 5fC as an activating mark with an important role in early embryonic development changes the current view on oxCs. Future studies will delineate the full spectrum of 5fC significance in gene expression regulation.

3.2. 5fC promotes *tRNA* expression by Pol III

Comparison of BS-seq and MAB-seq profiles in *X. tropicalis* ZGA (st9) embryos showed that ~85-95% of CpGs at *tRNA* genes are methylated and ~5-15% are formylated (Figure 2.4B). These data suggest that the vast majority of *tRNA* gene copies are methylated rather than formylated at ZGA. However, it is important to note that BS-seq and MAB-seq data represent whole embryos and the 5mC/5fC ratio possibly differs at specific regions, e.g. the chromocenter-positive animal pole (Figure 1.5B), where 5fC levels are expected to be higher than the whole-embryo average. For example, certain Pol II target genes are marked by H3K4me3 and are accordingly transcriptionally active in one cell population, while marked by H3K27me3 and silenced in a different cell population, in *X. tropicalis* embryos (Akkers et al., 2009). This finding reinforces the role of epigenetic modifications in spatial gene expression at the dorso-ventral or animal-vegetal axis of *Xenopus* embryos.

How does 5fC activate embryonic *tRNA* transcription when only a fraction of *tRNA* genes is marked by 5fC at ZGA? Increasing evidence supports that not all *tRNA* gene copies are expressed in cells. In *S. cerevisae*, deletion of one- or all-but-one copies from a family of multiple copy *tRNA* genes did not affect cell fitness, suggesting redundancy between different *tRNA* copies of the same family (Bloom-Ackermann et al., 2014). The same study also showed that a given *tRNA* gene family includes major and minor gene copies, which are more or less essential for cell growth and compensation within the family. In human fibroblasts, 32% of Pol III target genes, including *tRNA* genes, are not bound by Pol III and are devoid of activating histone marks, consistent with permanent transcriptional silencing (Orioli et al., 2016). In line with this, half of the *tRNA* genes are not occupied by Pol III in HeLa cells (Oler et al., 2010). Furthermore, comparison of Pol III occupancy between HeLa and T cells showed that 26% of *tRNA* genes were occupied by Pol III in one cell type but not the other and 16% of *tRNA* genes were not Pol III bound in either cell type (Barski et al., 2010). Importantly, cell type-specific Pol III occupancy might be specific to *tRNA* genes, as all *U6 snRNA* genes were bound by Pol III in both cell types (Barski et al., 2010).

Together, the above show that (a) different copies of the same *tRNA* gene are functionally redundant, (b) not all *tRNA* genes are expressed in all cell types, and (c) the current view of global Pol III transcription regulation through shared TFs at similar core promoters is flawed.

MAB amplicon sequencing for *X. tropicalis tRNA-iMet* and *tRNA-Gly* revealed that all CpGs in- and around these *tRNA* genes are formylated to comparable levels (Figure 2.4B-D). Whether different 5fCpGs co-exist on the same gene copy and if so, what is the average number of modified cytosines per gene remains unclear. This finding suggests that the 5fC sequence context at *tRNA* gene loci may not be important. On the other hand, I showed that 5fC promotes Pol III transcription of a *tRNA* transgene when it marks the intragenic promoter B-box. Notably, only two formylated cytosines in a hemi-modified context suffice to boost *tRNA-iMet* transgene transcription up to 20-fold (Figure 2.12). It is possible that among the different 5fCpGs at the *tRNA* loci, only those at or close to the promoter elements contribute to Pol III transcriptional activation. In mouse embryos, promoter 5fC sites in gametes and parental pronuclei correlate with induced gene expression at 2- and 4- cell stage respectively, while no temporal regulation patterns were observed for intragenic 5fC sites (Zhu et al., 2017). Whether more 5fC marks induce even higher gene expression, how far 5fC can be away from the B-box to exert such an effect, and what is the role of hemi- and homo-modification, will be interesting topics for future studies.

It is important to note that *tRNA* genes are unique among other types of genes for the following reasons: i) *tRNA* genes are considerably short, usually between 70 and 110 nucleotides, ii) within their short length they harbor two intragenic promoters, the A-box and the B-box, and iii) regulatory (promoter) and non-regulatory regions alternate throughout the gene. The unique nature of *tRNA* genes might be relevant for the observed activating role of 5fC. If this is true, SINE elements, the other Pol III target genes with type II promoters, should also be activated by 5fC. In line with this, SINEs are downregulated during ZGA in Tet triple knockdown (TKD) mouse embryos (Arand et al., 2022). Studying the inherent characteristics of 5fC-induced Pol III transcriptional activation of *tRNA* genes could reveal valuable information about the conditions under which 5fC serves as an activating mark.

I showed that 5fC is required for endogenous *tRNA* transcription in *Xenopus* ZGA, as global 5fC downregulation by *Tet2/3* Mo KD or TDG OE reduced endogenous *tRNA-iMet* and *tRNA-Gly* transcripts to levels comparable to those of complete Pol III inhibition by α-amanitin (Figure 2.10A-C). To further support my endogenous *tRNA* transcription data, I designed a transgene assay, which confirmed that *tRNA* transcription from plasmids carrying a 5fC-modified *tRNA*-iMet gene was induced compared to unmodified and methylated controls. More than 40 years ago, Newport

& Kirschner reported that microinjection of plasmid carrying a yeast *tRNA-Leu* gene into *X. laevis* embryos resulted in immediate *tRNA* transcription that was quickly silenced and only reactivated upon ZGA (Newport & Kirschner, 1982b). Accordingly, we observed no transgene expression at st8, but only immediately after ZGA at st10, when the presence of 5fC induced transgene transcript levels up to 20-fold compared to unmodified and methylated transgenes (Figure 2.12A-B, E).

DNA methylation is not transcriptionally repressive in early *Xenopus* embryos (Bogdanović et al., 2011). Accordingly, our methylated *tRNA-iMet* transgene did not impair its transcript levels compared to unmodified control transgenes (Figure 2.12A). Oocyte type *tRNA* genes exist in a binary state during *Xenopus* ZGA, where copies of the same gene are either methylated or formylated (Figure 2.4B). At early stages, a fraction of the methylated loci is presumably processed by Tet enzymes to generate 5hmC, which is then further processed to 5fC at the onset of ZGA (Figure 2.1B). 5fC accumulates in the absence of Tdg and triggers *tRNA* gene transcription by Pol III. At a later time point, during neurula stages, Tdg is present and excises 5fC, which results in oocyte-type *tRNA* gene silencing. The portion of oocyte-type *tRNA* genes that remained methylated during ZGA are presumably not- or lowly expressed during ZGA and they are actively silenced in neurula, when the 5mC mark starts being repressive (Bogdanović et al., 2011). In this case, 5mC functions as a precursor or *intermediate* of 5fC on the *tRNA* genes that are activated at ZGA, and as a repressive mark at later developmental stages.

Additionally, the transgene assay showed that while 5fC induces *tRNA-iMet* transgene transcription when microinjected at the AP, it does not affect transgene transcription when microinjected at the VP (Figure 2.12B). This important finding highlights that while 5fC is necessary to induce *tRNA* transcription, it is not sufficient. Additional factors involved in 5fC-induced *tRNA* transcription are selectively present at the AP during ZGA. Since embryonic transcription initiates at the AP (Chen et al., 2019), these factors could be exclusively produced by the embryonic genome and not maternally provided, hence they are not yet present at the VP. Alternatively, maternal factors required for 5fC mediated Pol III transcription might be restrained at the AP during early stages. Indeed, the distribution of a large fraction of maternal products in the egg is known to be polarized and this polarity is maintained after fertilization and through early development (Kloc, 2014).

3.3. 5fC as a regulator of ZGA across species

The mammalian genome undergoes global DNA demethylation during early development, as reported in human (Guo et al., 2014), mouse (Amouroux et al., 2016; Smith et al., 2012), bovine (Duan et al., 2019) and goat (Deng et al., 2020) embryos. This observation, along with the well-established role of 5mC as a repressive mark, lead to the assumption that 5mC erasure might be implicated in ZGA. However, no global DNA demethylation is observed in early zebrafish (Jiang et al., 2013; Potok et al., 2013) and frog embryos (Bogdanović et al., 2011; G. J. C. Veenstra & Wolffe, 2001 and this study). On the contrary, we showed that global 5mC levels increase at the onset of ZGA in *Xenopus* (Figure 1.5A and 2.1B), although the meaning of 5mC induction around *Xenopus* ZGA remains elusive. A study in goat embryos showed that reduced DNA methylation correlates with increased transcription of selected genes at ZGA (Deng et al., 2020). On the other hand, heavily methylated distal regulatory elements attracted pioneer factors, which facilitated chromatin opening during zebrafish ZGA (Liu et al., 2018). Overall, the role of 5mC in ZGA remains unclear and might differ across species.

Here, we describe for the first time the role of a different DNA modification, 5fC, as an activator of *tRNA* gene expression during *Xenopus* zygotic reprogramming. A few lines of evidence support that the role of 5fC in ZGA might be conserved in mammals. First, Dr. Viviana Vastolo, this lab, performed IF in mouse pre-implantation embryos, which confirmed 5fC staining at the paternal pronucleus (PN) of the zygote, as previously reported (Inoue 2011, Zeng 2019) (Figure 3.2A). However, we noticed that the paternal 5fC staining was not uniform but occurred in puncta in late stage (PN4/5) zygotes, when the minor ZGA takes place (Abe et al., 2018; Aoki et al., 1997; Bouniol et al., 1995). 5fC chromocenters were transient and disappeared by the 2-cell stage (major ZGA). Importantly, the majority of the 5fC chromocenters in PN4/5 zygotes also stained positive for Pol III (Figure 3.2B). These results mirror our finding of transient 5fC chromocenters at *Xenopus* ZGA, which co-localize with Pol III at the PNC (Figure 1.5C-E) and suggest that the role of 5fC in ZGA might be conserved in mouse.

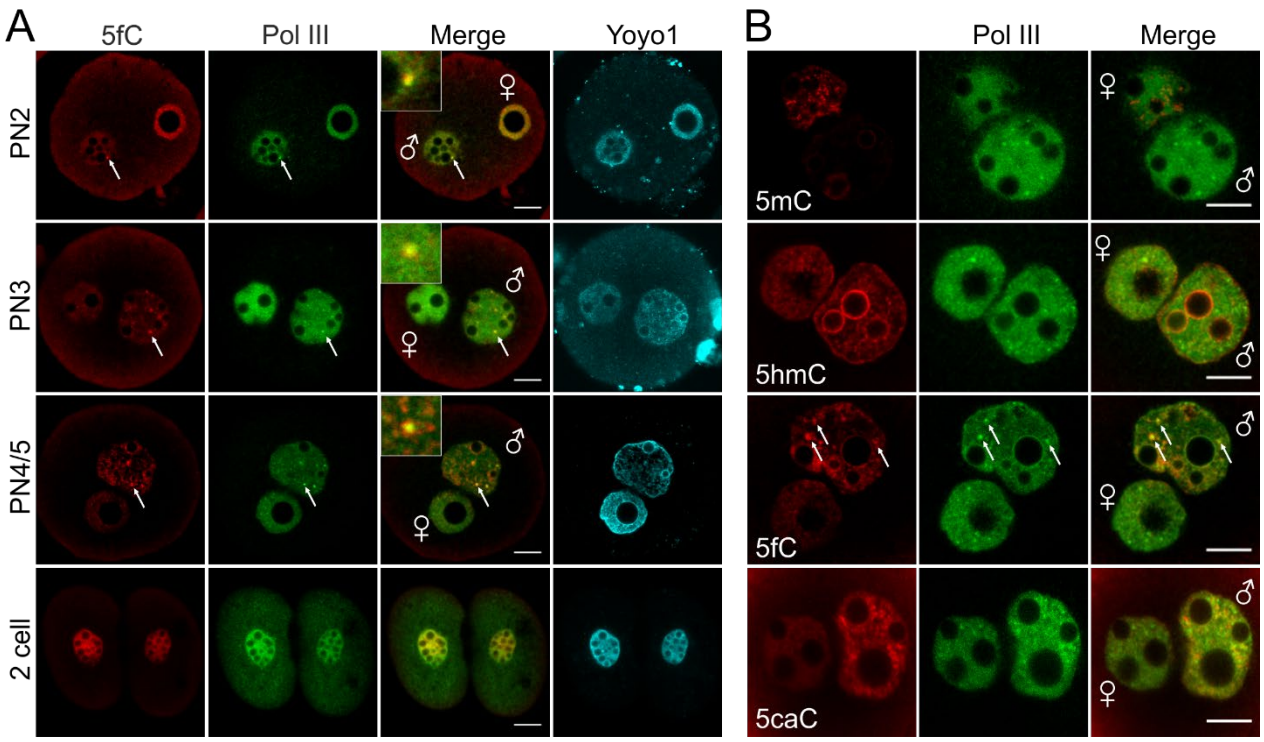


Figure 3.2. 5fC accumulation at Pol III sites is conserved in mouse ZGA.

A, B. Immunofluorescence microscopy (IF) of Pol III (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. Data collected and analyzed by Dr. Viviana Vastolo.

Second, 5fC foci were also reported in metaphase chromosomes of triploid human zygotes (Pendina et al., 2018). While 5caC also formed foci, 5mC and 5hmC chromosome staining was homogeneous. This observation matches our *Xenopus* metaphase chromosome staining at ZGA, where 5fC formed foci on several chromosomes, while 5mC staining was homogeneous (Figure 1.5F). Furthermore, the 5fC foci on human metaphase chromosomes were transient, as they were detected in zygotes but disappeared upon cleavage divisions (Pendina et al., 2018). Importantly, 60-70% of the nucleolar organized regions stained positively for 5fC and 5caC, suggesting that the human 5fC chromocenters associate with the nucleolus. Despite the similarity of human 5fC foci to *Xenopus* 5fC chromocenters, there seems to be a discrepancy at the developmental stage. While human 5fC foci were observed in zygotes, human ZGA takes place at the 8-cell stage (Braude et al., 1988; Tesařék et al., 1988; Vassena et al., 2011). Notably, a recent study reported that human ZGA initiates much earlier than previously thought, namely at the zygote (Asami et al., 2022), which suggests that the human 5fC foci might coincide with ZGA after all. In any case, the

significance of the triploid genome of these zygotes in the formation of 5fC foci is unclear, as 5fC foci were not observed in normal human embryos (Gao et al., 2020), although the resolution of the provided IF images is poor, and metaphase chromosome 5fC staining of diploid human zygotes is not available.

As mentioned above, 5fC chromocenters co-localized with Pol III at the paternal PN of mouse zygotes (Figure 3.2). While maternal PN also stained positively for 5fC, the signal was weaker, consistent with limited active DNA demethylation (Guo et al., 2014; Shen et al., 2014; L. Wang et al., 2014). Importantly, no maternal 5fC chromocenters co-localized with Pol III (Figure 3.2B). This could be due to technical limitations, since 5fC at the maternal genome is rare, rendering visualization by immunostaining challenging. Alternatively, this observation would suggest that, if 5fC-induced *tRNA* transcription activation is indeed conserved in mouse, it takes place exclusively at the paternal PN, which is 5fC-rich. In that case, all zygotic *tRNAs* would be transcribed from the paternal genome, with no maternal contribution. While genomic imprinting of certain genes is well established and refers to epigenetic silencing of the maternal or the paternal gene copy by DNA methylation (Ferguson-Smith & Bourchis, 2018), this would be the first case of asymmetric gene expression through direct marking of a parental copy for expression. This intriguing hypothesis can be readily tested by interrogating publicly available RNA-seq data of zygotes derived from crossings of different strains, about the genetic origin of tRNA transcripts, provided that tRNAs carry strain-specific SNPs.

Several models have been proposed regarding to the timing of ZGA. Among them, the maternal-clock model suggests that accumulation of essential maternal factors precedes embryonic transcription (Schulz & Harrison, 2019). This model traditionally refers to Pol II transcription factors, such as TFIID complex components, which reportedly accumulate before the onset of ZGA in *X. laevis* (Veenstra et al., 1999) and *C. elegans* (Guven-Ozkan et al., 2008). The present study identified the first case of a rate liming DNA modification, namely 5fC. At the onset of ZGA, 5fC accumulates at *tRNA* genes in the absence of Tdg to trigger Pol III transcription. Interfering with 5fC accumulation in *Xenopus* by *Tet2/3* Mo KD (Figure 1.6A-B) or TDG OE (Figure 2.5C-E) results in developmental delay and embryonic lethality. Furthermore, Gene Ontology (GO) enrichment analysis of differentially expressed genes (DEG) between control and *Tet2/3* KD st9 (ZGA) *X. tropicalis* embryos, showed that genes upregulated in *Tet2/3* KD embryos were involved in protein biosynthesis (e.g. ‘mRNA splicing’, ‘translation’ terms), while the top downregulated gene *psmb7* encodes a proteasomal subunit (Figure 3.3A-B). 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. Concordantly, 5fC DIP peaks were enriched on *tRNA* genes corresponding to 'optimal' codons (Figure 3.3C), which promote transcript stability and favor expression, while non-optimal codons promote RNA clearance (Bazzini et al., 2016; Mishima & Tomari, 2016). Together, our findings suggest that 5fC accumulation promotes embryonic *tRNA* transcription, which is required for protein synthesis. 5fC removal before ZGA results in delayed embryo development and lethality, consistent with protein resources not meeting the developmental requirements.

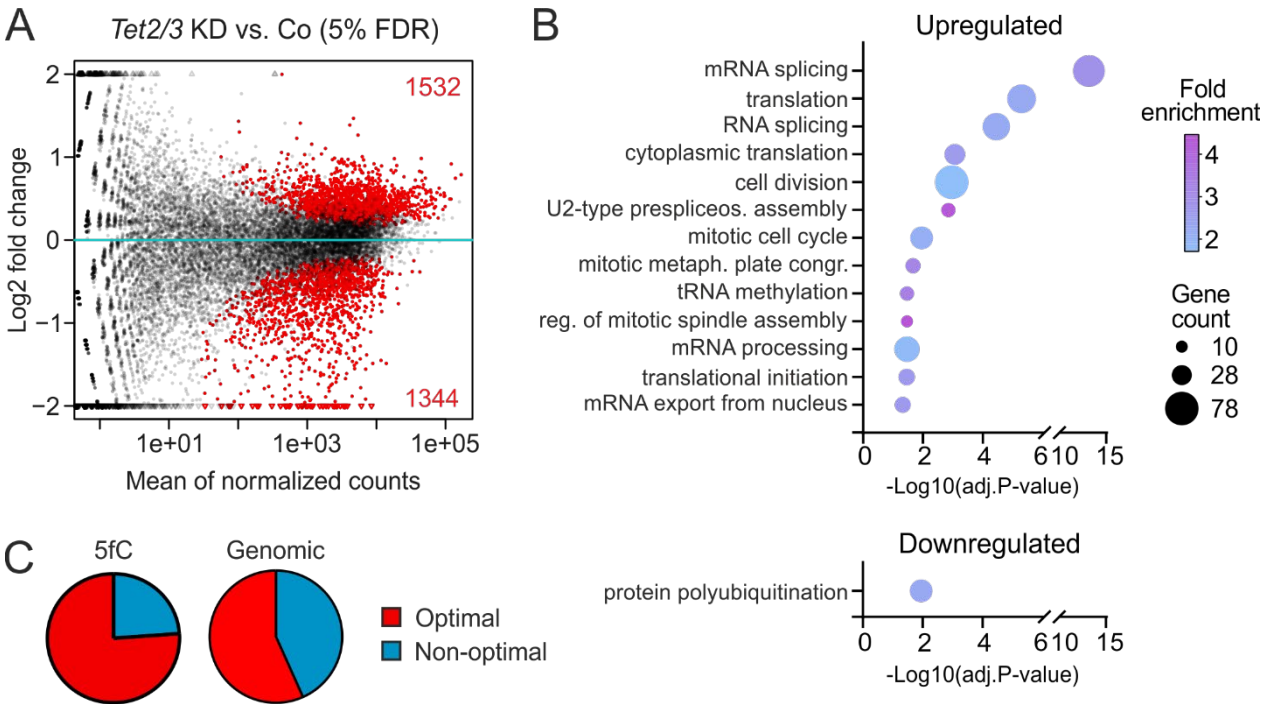


Figure 3.3. 5fC correlates with protein translation at ZGA.
A. Differential expression MA-plot of *Tet2/3* KD in ZGA (st9) *X. tropicalis* embryos. The up- and down-regulated genes (5% FDR) are highlighted in red and their numbers are indicated. **B.** Gene Ontology terms of significantly enriched up- (top) and downregulated (bottom) genes from (A). **C.** Correlation of 5fC DIP enriched *tRNA* genes with optimal and non-optimal codons. 5fC-enriched *tRNA* genes show over-representation of optimal codons (P-value=10^{-5} by χ^2 test). Samples prepared by Dr. Victoria Hatch and data analyzed by Dr. Emil Karaulanov and Dr. Medhavi Mallick.

3.4. The mechanism of 5fC induced epigenetic reprogramming

The mechanism by which 5fC activates *tRNA* transcription at ZGA is currently unknown. Previous i*n vitro* studies identified potential 5fC readers (Iurlaro et al., 2013; Spruijt et al., 2013), but their functional significance *in vivo* remains unclear. Accordingly, 5fC oligo pulldown using *X. laevis* ZGA (st9) nuclear protein extracts followed by proteomics mass spectrometry identified a limited number of potential 5fC interactors at ZGA (Figure 2.7B and Table 1). However, neither Pol III nor any of its transcription factors were listed among them, suggesting our oligo failed to attract the Pol III transcriptional machinery. This is probably due to poor DNA oligo design, as the oligo bait did not include a B-box (or an A-box), but rather another part of a *tRNA* sequence that was enriched in the 5fC DIP. Furthermore, the DNA repair factors Atm and Lig3 were identified among the proteins enriched for the 5fC oligo, suggesting that 5fC can still attract the BER pathway despite the absence of Tdg at ZGA. Since TDG-independent DNA demethylation has not been reported, it is unclear whether this finding is physiologically relevant or a technical artifact.

A promising 5fC-reader candidate would be TFIIIC. 5fC on *tRNA* genes associates with the Tfiiic binding element B-box and Tet2/3 KD reduces Tfiiic binding to its targets at ZGA (Figure 1.6D). Since TFIIIC is required for initial assembly of the transcriptional machinery, it is conceivable that 5fC promotes TFIIIC binding to the B-box to initiate Pol III machinery assembly (Figure 3.4). TFIIIC is also known to share properties with CTCF. Like CTCF, TFIIIC can recruit condensin and cohesin and is enriched at TAD boundaries (Van Bortle & Corces, 2012). Maybe 5fC induces TFIIIC binding to initiate chromatin looping (Figure 3.4). Furthermore, TFIIIC was reported to have histone acetyltransferase activity (Hsieh et al., 1999) and could thereby be attracted by 5fC to reshape chromatin around *tRNA* genes in order to facilitate Pol III transcription (Figure 3.4). Preliminary *in vitro* work by Dr. Lars Schomacher, testing whether Tfiiic exhibits increased affinity for 5fC-modified sequences was inconclusive. However, the finding that 5fC marks all CpGs in- and around *tRNA* genes argues against a specialized TF reader model.

In yeast, high Pol III density was observed at the A-box and B-box of tRNA genes, suggesting that Pol III might stall at these positions (Turowski et al., 2016). Furthermore, TFIIIC interaction with DNA is relatively short-lived compared to that of TFIIIB, which remains attached to DNA to allow for rapid transcription re-initiation (Kassavetis et al., 1990; Soragni & Kassavetis, 2008). It is thus imaginable that 5fC promotes TFIIIC dissociation from the *tRNA* gene body after initial Pol III machinery assembly to facilitate transcription. In this scenario, 5fC promotes *tRNA* transcription by enabling TFIIIC-Pol III turnover instead of TFIIIC binding (Figure 3.4). This hypothesis is also compatible with 5fC being present not only at the TFIIIC binding site but also elsewhere on the

gene body. However, Tfiiic ChIP-qPCR showed reduced Tfiiic binding to 5fC-rich *tRNA* genes in *Tet2/3* KD ZGA (st9) *X. tropicalis* embryos (Figure 1.6D), suggesting that 5fC positively correlates with Tfiiic binding, which argues against this scenario.

Nucleosome positioning would be an attractive mechanism for 5fC-mediated Pol III activation (Figure 3.4). The 5fC mark could repel nucleosomes from the marked sequences allowing for Pol III binding. This scenario would also explain our observation that 5fC also marks the intergenic region around *tRNA* genes, on top of the gene promoters and gene body (Figure 2.4B). In 1995, Almouzni and Wolffe suggested that Pol II and Pol III activation in *Xenopus* ZGA is regulated by two distinct mechanisms (Almouzni & Wolffe, 1995). While Pol II TFs are not readily available before ZGA, rendering TF availability the limiting factor of Pol II gene expression, the Pol III transcriptional machinery is abundantly available and potentially transcriptionally active before ZGA. Instead, Pol III transcription exclusively depends on chromatin organization and nucleosome occupancy, as artificial increase of the embryo DNA amount induced a *tRNA* transgene transcription before ZGA, while premature transcription of a Pol II transcribed CMV promoter also required the addition of recombinant TBP (Almouzni & Wolffe, 1995). Thus, 5fC accumulation on and around *tRNA* genes at the onset of ZGA could cause nucleosome sliding and chromatin reconfiguration towards a more permissive state. Indeed, metagene profiles of 5fC DIP peaks on *tRNA* genes coincide with a local sink in both activating (H3K4me1, H3K4me3) and repressive (H3K9m3) histone modifications (data not shown), suggesting that they might be nucleosome depleted. Analysis of publicly available ATAC-seq data of *X. tropicalis* st9 (ZGA) embryos (Bright et al., 2021) will confirm whether this is the case.

How might 5fC reorganize nucleosome position? The catalytic subunit of the BAF remodeling complex Smarca4 was enriched for 5fC oligo binding in our oligo pulldown, though not significantly and by less than 2-fold (Table 1). Smarca2 and -4 inhibition by BRM014 did not affect endogenous *tRNA-iMet* transcript levels at ZGA, suggesting that the BAF complex is likely not involved in 5fC-induced *tRNA* transcription at ZGA. Alternatively, 5fC might directly interfere with nucleosome-DNA interaction, or signal through different chromatin remodeling complexes, for example the NuRD complex, which is known to control gene expression during embryonic development (Basta & Rauchman, 2015).

In vitro experiments suggest that 5fC promotes double helix unwinding (Dubini et al., 2020; Jaisal et al., 2023; Raiber et al., 2015). Notably, symmetric 5fC destabilizes the dsDNA structure, as measured by a ~2°C decrease of the Tm and up to 10 kJ/mol decrease in ΔG° (Dubini et al., 2020). Accordingly, 5fC on *tRNA* genes could facilitate template opening, which is normally mediated by TFIIIB (Hahn & Roberts, 2000; Kassavetis et al., 1998). Interestingly, our *tRNA-iMet* transgene assay confirmed low transcription levels from unmodified and methylated plasmids at ZGA, which increased dramatically in 5fC-modified plasmids (Figure 2.12A-B, E). These data show that Pol III transcription takes place regardless of 5fC, but the mark renders Pol III transcription faster or more efficient, which could be due to 5fC-induced priming by double helix unwinding (Figure 3.4).

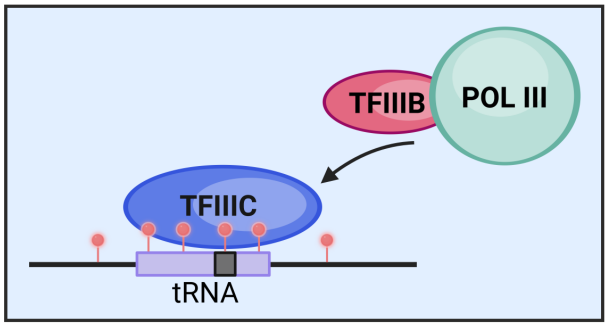
5fC could also indirectly promote Pol III transcription by driving the formation of the PNC (Figure 3.4). Physical proximity of *tRNA* genes by spatial organization at the PNC offers a number of benefits: TF targeting, RNA polymerase re-initiation, downstream transcript processing, such as maturation, RNA modification etc. Indeed, 4C experiments revealed that tRNA insulators make long-range contacts with other tRNA genes (Raab et al., 2012). 5fC could stimulate initial PNC assembly at the onset of ZGA, bringing *tRNA* loci from different chromosomes in contact. In line with these, the nucleolar protein Pdcd11 was selectively enriched for the 5fC oligo in our oligo pull down experiment (Table 1, Figure 2.7B), suggesting that 5fC might promote association with nucleolar components. After the PNC has formed, its structural integrity depends on Pol III transcription, as Pol III inhibition with Tagetin results in PNC resolution in HeLa cells (Wang et al., 2003). Possibly, 5fC-mediated PNC assembly initiates Pol III binding and transcription, which in turn is essential to maintain the PNC structure. Brown and Gurdon showed that *X. laevis* embryos without nucleolus develop slower, show microcephaly phenotypes and lethality (Brown & Gurdon, 1964), reminiscent of our TDG OE phenotype (Figure 2.5C-E). Nucleolar staining of Tet2/3 Mo KD embryos in ZGA will reveal whether 5fC is required for nucleolus formation.

Conversely, *tRNA* gene assembly at the PNC could facilitate 5mC oxidation to 5fC selectively at these genes, by localized Tet targeting to the PNC. Consequently, the PNC would be a Tet oxidation hub, where 5fC is generated and accumulates. Indeed, 5S rRNA genes, which are known to localize to the PNC vicinity, at the nucleolus, are also 5fC modified at ZGA, but 5fC is not required for their transcription (Figure 2.10E-F). IF staining for Tet around ZGA will reveal whether Tet proteins are targeted to the PNC before the onset of ZGA. It would be important to find out what comes first: *tRNA* gene formylation or their assembly to the PNC. Understanding this will help us delineate the mechanism of 5fC-mediated zygotic reprogramming.

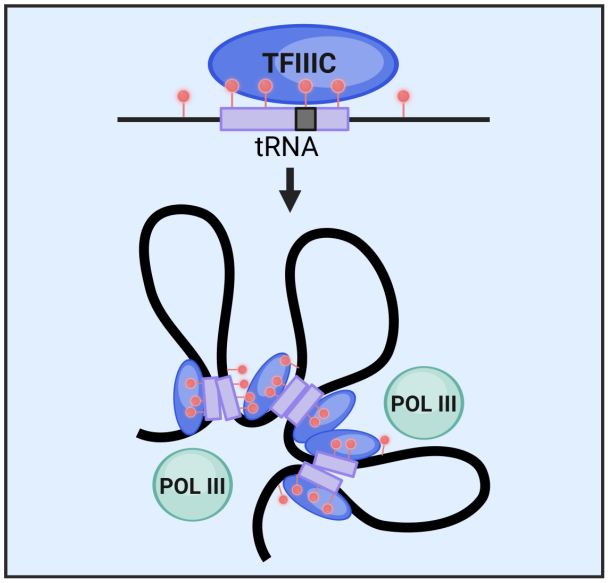
Overall, the mechanism by which 5fC promotes *tRNA* transcription during ZGA remains unresolved. It is currently unclear whether 5fC directly attracts the Pol III machinery to its targets, or whether it creates permissive chromatin around them. Characterizing the protein environment of 5fC at ZGA, by targeting an engineered dCas9-BioID to oocyte *tRNA* repeats *in vivo*, could offer a better perspective on its mode of action. Understanding the mechanism of 5fC-induced Pol III activation at ZGA will ultimately help identify additional roles that 5fC might play in different physiological contexts.

A. Signaling through TFIIIC

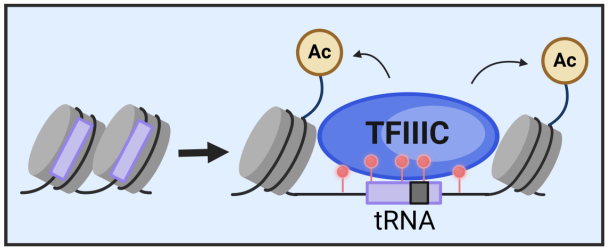
(i) Pol III machinery recruitment



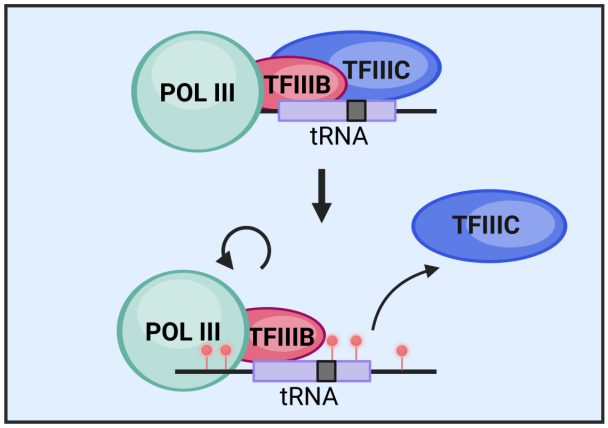
(ii) Chromatin looping



(iii) Histone acetylation

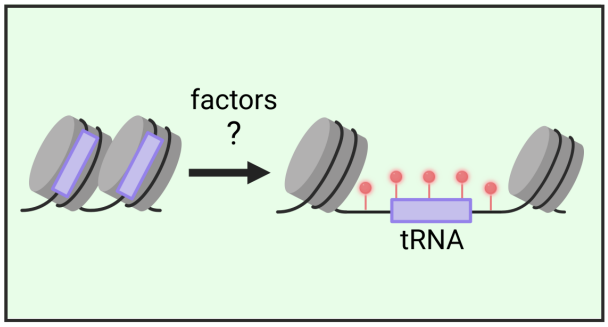


(iv) TFIIIC dissociation, transcription re-initiation

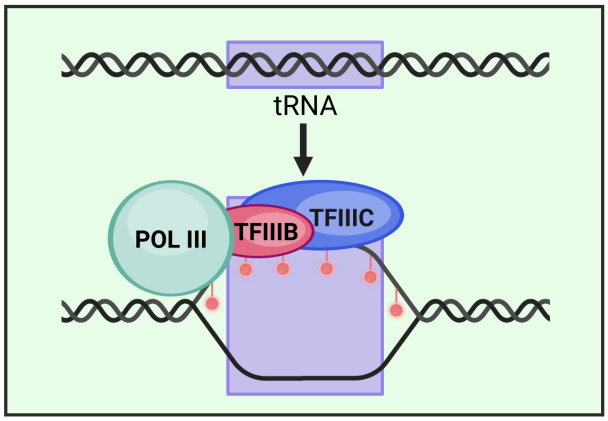


B. DNA & chromatin conformation altering

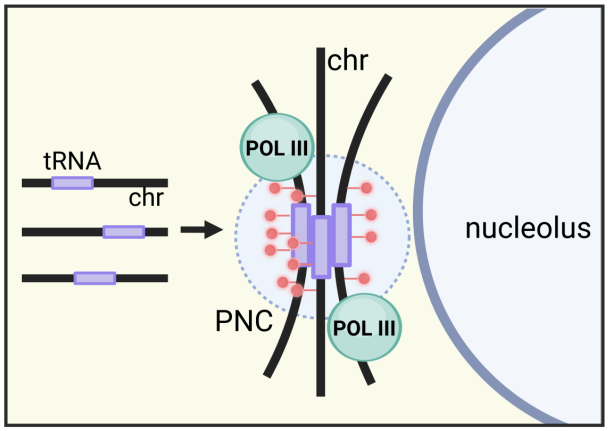
(i) Nucleosome positioning



(ii) DNA unwinding



C. PNC assembly



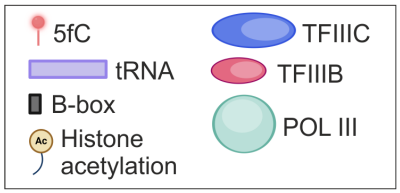


Figure 3.4. Proposed mechanisms for 5fC-mediated *tRNA* transcription activation during ZGA.
A. 5fC signals through TFIIIC, by increasing (i-iii) or decreasing (iv) its affinity for DNA binding. (i) 5fC may promote TFIIIC binding, which in turn recruits TFIIIB and Pol III. (ii) Upon 5fC induction, TFIIIC might mediate chromatin looping to bring in contact distal *tRNA* genes from the same or different arrays, facilitating their coordinated transcription by Pol III. (iii) 5fC could promote TFIIIC binding, which in turn acetylates the histone tails of proximal chromosomes, promoting chromatin opening. (iv) Conversely, after Pol III machinery assembly, 5fC might repel TFIIIC from the B-box to facilitate Pol III transcription re-initiation. **B.** 5fC-mediated chromatin and DNA conformation changes. (i) 5fC might attract unknown chromatin remodelers to promote nucleosome sliding and chromatin opening. (ii) 5fC could mediate DNA double helix unwinding at *tRNA* genes to enable Pol III transcription. **C.** 5fC stimulates PNC assembly. *tRNA* genes on different chromosomes marked by 5fC coalesce at the PNC, potentially via phase separation. Pol III transcribes the spatially associated tRNA genes and maintains the structural integrity of the PNC. PNC: perinucleolar compartment. Chr: chromosome.

3.5. Limitations of this study

To test the role of 5fC in endogenous *tRNA* transcription during ZGA, I used northern blot to assay *tRNA-iMet* and *tRNA-Gly* transcript levels after Tet and Tdg manipulation (Figure 2.10). As mentioned in section 2.6.1., *tRNA* qPCR was not possible due to RNA modifications interfering with primer binding and reverse transcription. However, northern blot has certain limitations. First, the number of different *tRNAs* that can be assessed in parallel is limited by the large amount of total RNA required (10 µg per sample), and the low throughput nature of the method. Second, northern blot is a semi-quantitative assay, not suitable for detection of small differences in *tRNA* abundance.

In the future, it will be useful to apply high throughput methods to quantify transcription across all embryonic *tRNAs* at ZGA. Such an experiment would allow side-by-side comparison of *tRNA* transcription from 5fC modified (*tRNA-iMet*, *tRNA-Gly*, *tRNA-Asp*, *tRNA-Arg*) and unmodified (*tRNA-Cys*, *tRNA-Thr*, *tRNA-eMet*) genes in physiological and Tet KD/TDG OE background, providing further insights into the role of 5fC in *tRNA* expression. To this end, I established four-leaf clover qPCR (Honda et al., 2015), a method designed to selectively amplify *tRNAs* by ligation of a stem-loop adapter that will be subsequently amplified instead of the *tRNA*, in an effort to bypass *tRNA* modifications. Even though I successfully amplified a few human *tRNAs* and reproduced published data, amplification of *Xenopus tRNA* failed (data not shown). This is again due to the lack of information concerning both the position and the type of modifications at *Xenopus tRNAs,* which do not seem to be conserved between frog and human. An alternative method would be enzymatic removal of common reverse transcriptase blocking RNA modifications (Potapov et al., 2018) before cDNA synthesis, followed by qPCR or small RNA sequencing.

A common problem in studying transcription at ZGA is identifying the scarce embryonic transcripts among massive amounts of maternally loaded RNA. To overcome this problem, I used α-amanitin to inhibit embryonic Pol III transcription, in order to estimate the percentage of embryonic versus maternal transcripts in the *tRNA* pool at ZGA (st9). To more precisely study embryonic *tRNA* transcription in the presence and absence of 5fC, one can use metabolic labelling of nascent RNA with ribonucleotide analogues followed by RNA-seq, which has successfully been applied to study embryonic mRNA transcription (Heyn et al., 2014; Kwasnieski et al., 2019; Tani & Akimitsu, 2012).

Concerning conservation of 5fC-mediated zygotic reprogramming across species, confirming 5fC accumulation on *tRNA* genes during mammalian ZGA would provide additional evidence. Published datasets on genomic 5fC signature in early mammalian embryos are currently limited to two studies, one in mouse (Zhu et al., 2017) and another in human embryos (Gao et al., 2020). Both works used malononitrile assisted sequencing to study genome-wide 5fC distribution with base resolution. While the mouse embryo data are publicly available, those on human embryos are not and were not made available to us upon request to the authors. Anyhow, we find these datasets unreliable and thus not useful to our study for the following reasons: First, my own efforts using malononitrile-seq in frog embryos showed that 5fC-M inhibits the DNA polymerases used to acquire the published data (Figure 2.3C). Second, the available datasets lack information critical for our analysis. Single nucleotide polymorphisms (SNPs) and reference genome discrepancies can greatly affect base resolution sequencing data interpretation (e.g. C-to-T transitions unrelated to the chemistry used to call 5fC sites). Hence, untreated and mock treated controls are indispensable for proper data analysis but were not included in the publicly available data.

Lastly, as detailed in section 3.4, this study did not characterize the mechanism of 5fC-mediated *tRNA* transcriptional activation at ZGA. Delineating the 5fC mechanism would be an exciting follow-up project, highly anticipated by the scientific community.

4. Materials and Methods

4.1. Materials

4.1.1. Equipment

2100 Bioanalyzer (Agilent technologies); agarose and polyacrylamide gel molds (Bio-Rad); agarose and polyacrylamide gel well combs (Bio-Rad); agarose gel horizontal electrophoresis system and power supplier (Bio-Rad); AxioCam MRc 5 camera (Zeiss); bacterial incubators (Thermo Scientific); bacterial shaker (Infors-HT); Biorupter Plus (Diagenode); centrifuges (Thermo Scientific); ChemiDoc XRS+ System (Bio-Rad); Concentrator plus (Eppendorf); Cooled Incubator MIR-153 (Sanyo); COVARIS E220; dounce homogenizers (Merck); DynaMag-2 Magnet rack (Invitrogen); EASY-nLC 1200 System (Thermo Scientific); Electronic Multichannel pipette (Sartorius); Electronic Pipette (Sartorius); Fireboy burner (Integra); freezer -20°C (Liebherr); freezer -80°C (Liebherr); fridges (Liebherr); ice maker (Wessamat); Illumina Version 7.0_9/22; LED control unit (Leica); LightCycler 480 (Roche); magnetic stirrer (Heidolph); microcentrifuges (Heraeus); microinjector PLI-100A Plus Pico-Injector (Harvard Apparatus); micromanipulator Mk1 (Singer); microwave oven (Sharp); multidispenser pipette (Eppendorf); Nanodrop 2000 spectrophotometer (Thermo Scientific); orbital shaker (Kisker); Orbitrap Exploris 480 (Thermo Scientific); P-97 micropipet puller (Sutter Instrument); PCR thermocycler (Biometra); Pipet boy (Integra); pipettes (Eppendorf); precision balance (Neolab); Qubit 4.0 Fluorometer (Invitrogen); refrigerated centrifuges (Thermo Scientific); rocking shaker (Kisker); Spark M20 microplate reader (Tecan); SteREO Discovery.V20 (Zeiss); stereomicroscope M80 (Leica); Test-Tube-Rotator (Kisker); ThermoMixer compact (Eppendorf); Thermostatically Controlled Standard Incubator (Aqualytic); Trans-Blot Turbo (Bio-Rad); Typhoon FLA 9500 (GE Healthcare); ultrapure water purification system (Millipore); vortexer (Scientific industries).

4.1.2. Lab supplies

25 cm × 75 µm inner diameter mass spectrometry column (New Objective); Amicon Ultra 30 centrifugal filters (Merk); assay plates, 384 well (3544, Corning); beakers (Schott); capilarries (30-0016, Harvard Apparatus); Criterion TGX Precast Midi Protein Gel, 7.5%, 12% (BioRad); dissecting scissors; DNA low binding 1.5 ml tubes (Eppendorf); Empore C18 StageTips (CDS Analytical); Eppendorf tubes 1.5 ml, 2 ml (Eppendorf); filter-tips, 10, 20, 100, 200, 1000 µl (Starlab); forceps; glass Petri dishes (Schott); Hybond-N+ membrane (RPN203B, GE Healthcare); microloader tips, 20 µl (Eppendorf); micrometer ruler (World Precision Instruments); microtube pestles (Starlab); NextSeq2000 P1 and P3 flow cells (Illumina); pH colorimetric strips (Marcherey-Nagel); Protein low binding 1.5 ml tubes (Eppendorf); ReproSil-Pur 120 C18-AQ (Dr. Maisch

GmbH); Scalpel (Paragon); stripettes, 5, 10, 25, 50 ml (Costar); Trans-Blot Turbo Midi 0.2 µm PVDF Transfer Packs (Bio-Rad); transfer pipets (Sarstedt); tubes, 15, 50 ml (Sarstedt); ultracentrifuge tubes, 30 ml (Beckman Coulter); Wattman paper (1703967, Biorad).

4.1.3. Chemicals

acetonitrile (Sigma-Aldrich); agarose (Sigma-Aldrich); ammonium accetate (Sigma-Aldrich); ammonium bicarbonate (Sigma-Aldrich); Ampicillin (IMB Media Lab); ATP magnesium salt (Sigma-Aldrich); Ca(NO3)2.4H2O, calcium nitrate tetrahydrate (Sigma-Aldrich); CaCl2.2H2O, calcium chloride dihydrate (Carl Roth); cytochalasin D (Sigma-Aldrich); Dimethyl sulfoxide (DMSO; Sigma-Aldrich); DTT, dithiothreitol (Sigma-Aldrich); EDC, Ethyl-3-(3-dimethylaminopropyl)carbodiimide (Sigma-Aldrich); EDTA, ethylenediaminetetraacetic acid (Sigma-Aldrich); EGTA, ethylene glycol-bis(β-aminoethyl ether)-N,N,N′,N′-tetraacetic acid (Sigma-Aldrich); ethanol (Carl Roth); Ficoll PM 400 (Sigma-Aldrich); formaldehyde solution (Sigma-Aldrich); formic acid (Sigma-Aldrich); Freon, 1,1,2-Trichlorotrifluoroethane (Sigma-Aldrich); glycerol (Sigma-Aldrich); glycogen (Sigma-Aldrich); HCl, hydrochloric acid (Sigma-Aldrich); HEPES (Sigma-Aldrich); IGEPAL CA-630 (Sigma-Aldrich); iodoacetamide (Sigma-Aldrich); KCl, potassium chloride (Sigma-Aldrich); L-cysteine hydrocloride monohydrate (Sigma-Aldrich); LiCl, lithium chloride (Sigma-Aldrich); methylimidazole (Sigma-Aldrich); MgSO4.7H2O, magnesium sulfate heptahydrate (Sigma-Aldrich); MOPS (Sigma-Aldrich); MS-222, ethyl-3-aminobenzoate (Sigma-Aldrich); NaCl, sodium chloride (Sigma-Aldrich); NaHCO3, sodium bicarbonate (Sigma-Aldrich); NaOAc, sodium acetate (Sigma-Aldrich); NaOH, sodium hydroxide (Sigma-Aldrich); NEM, N-Ethylmaleimide (Sigma-Aldrich); polyvinyl pyrolidone (Sigma-Aldrich); potassium acetate (Sigma-Aldrich); SDS, Sodium dodecyl sulfate (Sigma-Aldrich); sodium deoxycholate (Sigma-Aldrich); Sucrose (Sigma-Aldrich); trimethylamine-N-oxide (Sigma-Aldrich); Tris, tris(hydroxymethyl)aminomethane (Sigma-Aldrich); Tween-20 (Sigma-Aldrich); urea (Sigma-Aldrich); α-amanitin (Sigma-Aldrich).

4.1.4. Enzymes

AccuPrime DNA polymerase (Thermo Scientific); BglII (NEB); BsrGI HF (NEB); Deep Vent (exo-) DNA Polymerase (NEB); DNAse I Amplification Grade (AMPD1-1KT, Sigma-Aldrich); HindIII HF (NEB); M.SssI (NEB); mass spectrometry grade trypsin (Sigma-Aldrich); Nb.Bpu10I (Thermo Scientific); NotI (NEB); OneTaq (NEB); Phusion U Hot Start DNA Polymerase (Thermo Scientific); Proteinase K, recombinant, PCR grade (Roche); PvuII HF (NEB); Q5 High-Fidelity DNA Polymerase (NEB); recombinant TDG (Dr. Schomacher, this lab); reverse transcriptase (IMB Protein Production core facility; RNase A (Qiagen); RNase-Free Dnase I Set (Qiagen); SacI

(NEB); SalI HF (NEB); SfiI (NEB); T4 DNA ligase (NEB); T4 Polynucleotide kinase (NEB); XbaI (NEB).

4.1.5. Reagents

10% NOVEX gradient SDS gel (Thermo Scientific); 40% polyacrylamide solution, 29:1, 3.3% crosslinker (Biorad); 5-Formyl-2'-deoxycytidine-5'-Triphosphate (N-2064, TriLink); Agarose (Sigma-Aldrich); AMPure XP beads (Beckman Coulter); BRM014 (gift from Dr. Sandra Schick); BSA, Bovine Serum Albumin (Sigma-Aldrich); cOmplete EDTA-free Protease Inhibitor Cocktail (Roche); Coomassie Brilliant Blue G-250 (Biozym); CutSmart buffer (NEB); dNTPs (Thermo Scientific); Dynabeads MyOne Streptavidin C1 (Thermo Scientific); Dynabeads Protein G (10004D, Thermo Scientific); GelRed Nucleic Acid Stain 10000X Water (Millipore); Gibson assembly 2X MasterMix (IMB Protein Production Core Facility); glycine (Sigma-Aldrich); human Chorinonic Gonadotropin (hCG; C6051.1010, Lot No. 1021PHCG24, Genaxxon bioscience); isopropanol (Sigma-Adrich); L-15, Leibovitz's medium (Gibco); LB agar plates (IMB Media Lab); LB media (IMB Media Lab); LightCycler 480 Probes Master (Roche); LightCycler 480 SYBR Green I Master (Roche); MES buffer (Thermo Scientific); minreal oil (M5904, Sigma-Aldrich); Morpholinos (Gene Tools); NP-40, Nonidet P-40 (Sigma-Aldrich); nuclease-free H2O (Qiagen); NuPAGE LDS sample buffer (Thermo Scientific); One Shot TOP10 cells (Invitrogen); Pen Strep (10,000 Units/ml Penicilin, 10,000 µg/ml Streptomycin; Gibco); Pierce Protein G Magnetic Beads (Thermo Scientific); Qiazol (Qiagen); random primers (Thermo Scientific); recombinant hCRY2 (gift from Prof. Eva Wolf); recombinant mCRY1 (gift from Prof. Eva Wolf); ReproSil-Pur 120 C18-AQ (Dr. Maisch GmbH); Restore Western Blot Stripping buffer (Thermo Scientific); RiboLock RNase-Inhibitor (Thermo Scientific); ROTIC/I, chloroform/isoamyl alcohol mixture (Carl Roth); skim milk powder (Sigma-Aldrich); SuperSignal West Pico (Thermo Scientific); UltraPure Salmon Sperm DNA Solution (Thermo Scientific); UPL probe library (Roche); XL1-Blue Competent Cells (Agilent); Zero Blunt PCR Cloning vector (Invitrogen); γ^{32}P ATP 3000Ci/mmol 10mCi/ml (Perkin Elmer).

4.1.6. Kits

DNA Clean & Concentrator-5 (Zymo Research); DNeasy Blood & Tissue kit (Qiagen); dsDNA HS Assay Kit for Qbit (Invitrogen); EpiTect Bisulfite Kit (Qiagen); High Sensitivity DNA kit for Bioanalyzer(Agilent technologies); MEGAscript SP6 Transcription kit (Thermo Scientific); MEGAscript T3 Transcription kit (Thermo Scientific); MEGAscript T7 Transcription kit (Thermo Scientific); Microspin G-25 Columns (27-5325-01, GE Healthcare); NEBNext Ultra II DNA Library Prep Kit (NEB); QIAfilter Plasmid Midi kit (Qiagen); QIAprep Spin Miniprep kit (Qiagen); QIAquick

Gel Extraction kit (Qiagen); QIAquick PCR Purification kit (Qiagen); Qubit dsDNA HS Assay Kit (Thermo Scientific); RNeasy Mini kit (Qiagen); xGen ssDNA & Low-Input DNA Prep Kit (IDT)

4.1.7. Antibodies

Primary antibodies	WB dilutions
BRG1 (Smarca4; 21634-1-AP, Proteintech)	1:800
Brg1 (Smarca4; Wagner et. Al 2017)	1:5
IgG (ab171870, Abcam)	N/A
Pol II (C15200004, Diagenode)	N/A
POLR3F (ab151495, Abcam)	N/A
TDG (61437, Active Motif)	1:1000
α-tubulin (T5168, Sigma)	1:2000
Secondary antibodies	
goat anti-mouse IgG-HRP (115-035-146, Dianova)	1:10,000
goat anti-rabbit IgG-HRP (111-035-144, Dianova)	1:10,000
goat anti-rat IgG-HRP (AP136P, Millipore)	1:10,000

Table 3. List of primary and secondary antibodies.
Dilutions for western blot (WB) working solutions are shown on the right hand column. N/A refers to antibodies that were not used in WB. 5 µg antibody were used for DIP and ChIP in all cases where relevant.

4.1.8. Buffers and Solutions

Annealing buffer (1X): 10 mM Tris pH 7.5, 50 mM NaCl, 1 mM EDTA

B+ buffer (1X): 10 mM HEPES-KOH pH 7.7, 10% glycerol, 125 mM potassium acetate, 1.5 mM MgCl2, 1 mM DTT, 0.5% NP-40, 1 M NaCl

B+W, binding and wash buffer (2X): 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl

Barth (10X): 889 mM NaCl, 10 mM KCl, 24 mM $NaHCO_3$, 8 mM $MgSO_4.7H_2O$, 3.3 mM $Ca(NO_3)_2.4H_2O$, 4.1 mM CaCL2.2H2O, 100 mM HEPES, pH 7.6

Coomassie solution: 0.25% Coomassie Brilliant Blue G-250, 10% acetic acid, 43% ethanol

Crosslinking buffer (1X): 1% 1-methylimidazole, 12.5 mM HCl, 3.1% EDC

Cushion: 10 mM HEPES-KOH pH 7.7, 900 mM sucrose, 50 mM KCl, 2.5 mM MgCl2

ELB, embryo lysis buffer (1X): 10 mM HEPES-KOH pH7.7, 250 mM Sucrose, 50 mM KCl, 2.5 mM $MgCl_2$, 1 mM DTT, cOmplete EDTA-free Protease Inhibitor Cocktail

Elution buffer (1X): 100 mM $NaHCO_3$ pH 8.8, 1% SDS

FP, fluorescence polarization buffer (1X): 20 mM HEPES pH 7.0, 100 mM NaCl, 5% glycerol

Glycosylase buffer (1X): 10 mM HEPES-KOH pH 7.4, 1 M trimethylamine-N-oxide, 100 mM KCl, 10 mM EDTA, 0.5 mM DTT, and 100 µg/ml BSA

Hybridization buffer (1X): 5X SSC buffer, 0.1% Ficoll, 0.1% polyvinyl pyrolidone, 0.1% BSA

Incubation buffer (1X): 50 mM Tris–HCl pH 8, 100 mM NaCl, 2 mM EDTA, 1 mM DTT, 1% Nonidet P-40 and cOmplete EDTA-free Protease Inhibitor Cocktail

IP, immunoprecipitation buffer (1X): 10 mM Na-Phosphate pH 7, 140 mM NaCl, 0.05% Triton X-100

LDS (4%): 1X NuPAGE, 300 mM DTT, 2% SDS

Lysis buffer (1X): 250 mM sucrose, 1% NP-40, 10 mM EDTA, 25 mM HEPES pH 7.2, 10 µM cytochalasin D

MEMFA (1X): 100 mM MOPS pH 7.4, 2 mM EGTA, 1 mM $MgSO_4.7H_2O$, 4% formaldehyde

MS Buffer A: 0.1% formic acid

MS Buffer B: 0.1% formic acid, 80% acetomitrile

MS-222 (0.15%): 1.5 g/l MS-222, 1.4 g/l NaHCO3

NP-40 lysis buffer (1X): 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 PhosSTOP phosphatase inhibitor

Protein binding buffer (1X): 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM DTT, 0.25% IGEPAL CA-630, cOmplete EDTA-free Protease Inhibitor Cocktail

Protein destaining solution: 50% ethanol, 50 mM ammonium bicarbonate pH 8

Proteinase K digestion buffer: 50 mM Tris pH 8, 10 mM EDTA, 0.5% SDS

Ringer (1X): 115 mM NaCl, 2.9 mM KCl, 1.8 mM CaCl2.2H2O, 5 mM HEPES, pH 7.4

Sonication buffer (1X): 20 mM Tris–HCl pH 8, 70 mM KCl, 1 mM EDTA, 10% glycerol, 5 mM DTT, 0.125% Nonidet P-40 and cOmplete EDTA-free Protease Inhibitor Cocktail

SSC, saline-sodium citrate buffer (20X): 3 M NaCl, 300 mM trisodium citrate, pH 7

TBE, Tris-Borate-EDTA (5X): 5.4 g/l Tris, 27.5 g/l boric acid, 10 mM EDTA

TBS, Tris-buffered saline (20X): 132 g/l Tris-HCl, 19.5g/l Trizma base, 175 g/l NaCl, 4 g/l KCl

TBST, Tris-buffered saline 0.1% Tween-20 (1X): 1X TBS, 1 ml Tween-20

TE, Tris-EDTA (1X): 10 mM Tris, 1 mM EDTA, pH 8

Wash buffer 1 (1X): 50 mM Tris–HCl pH 8, 100 mM NaCl, 2 mM EDTA, 1 mM DTT, 1% Nonidet P-40, 0.1% sodium deoxycholate, and cOmplete EDTA-free Protease Inhibitor Cocktail

Wash buffer 2 (1X): wash buffer 1 supplemented with 400 mM NaCl

Wash buffer 3 (1X): wash buffer 1 supplemented with 250 mM LiCl

4.1.9. Morpholinos

Target	Morpholino sequence (5' -> 3')
Standard Control Mo	CCTCTTACCTCAGTTACAATTTATA
Tet2 (*X. tropicalis*) Mo	GGGTCATTCAGTTATATCCGAGAAA
Tet3 Mo	CTGAGTCTCCATCATGTTTCAGCCA
Tdg Mo	GCCTCCATTCTGCTGTTCGATATTA

Table 4. List of Morpholinos.
The listed Morpholinos are complementary to both *X. laevis* and *X. tropicalis mRNA*, unless otherwise stated.

4.1.10. Plasmids

Plasmid	Source	Niehrs ID
pBpu10I	this work	3920
pBpu10I-mut-B-box	this work	3921
pCS2+	Niehrs group	
pCS2+iMet -no-CMV	this work	3924
pCS2+iMet-iMet-Tyr-no-CMV	this work	3929
pCS2+iMet-iMet-Tyr-no-CMV-mut-B-box	this work	3930
pCS2+iMet-no-CMV-mut-B-box	this work	3925
pCS2+RLuc	this work	3917
pCS2+RLuc-iMet-iMet-Tyr	this work	3941
pCS2+RLuc-iMet-iMet-Tyr-SacI	this work	3942
pCS2+single-iMet	this work	3922
pCS2+single-iMet-mut-B-box	this work	3923
pCS2-flag-hTDG	Niehrs group	513
pHH25-HA-Tdg-N140A	Hardeland et al., 2000	470
pRL-TK	Niehrs group	376
pRN-PpI	Niehrs group	663
ZeroBlunt-iMet-iMet-Tyr	this work	3932

Table 5. List of plasmids.
Plasmid prepared by me during the present study are stated as "Source: this work". Niehrs ID column indicates the identification number of each plasmid at the Niehrs Lab database (https://niehrs.imb.uni-mainz.de).

4.1.11. Oligonucleotides

MAB Amplicon sequencing PCR primers			
Target sequence	**1st PCR primer sequence (5' -> 3')**	**2nd PCR primers**	
		Forward primer sequence (5' -> 3')	**Reverse primer sequence (5' -> 3')**
tRNA-iMet forward strand	CTCCTCTGACTGTAACCACG GCCTGCCACCATTCATCCTC	GAGTCCTCTCACC AGCATCTGAC	CTCCTCTGACTGT AACCACG
tRNA-Gly forward strand	CTCCTCTGACTGTAACCACG GCAGGCATGGCTGCAAAGT	TAAAGGCCACAGG TGGGTATGGAG	
Gapdh forward strand	CTCCTCTGACTGTAACCACG CATCATCCCTGCCTCAACTG	GCTGCAGTCTTAA TGGCAGC	
tRNA-iMet BS forward strand	N/A	TATTTTAGATAGT GGATTGAAAGTTA G	CTCTCACCAACAT CTAACCAAAAAC
tRNA-Gly BS forward strand	N/A	GTTATAGGTGGGT ATGGAG	TACCCTACATAAA CAAACATAACTAC AAAAT
Gapdh BS forward strand	CTCCTCTGACTGTAACCACG CATCATCCCTACCTCAACTA	TGTTGTAGTTTTAA TGGTAGTTTTG	CTCCTCTGACTGT AACCACG
untreated 5fC Spike-in	CTCCTCTGACTGTAACCACG CACTCATAACTTCCTAACCA TCTTC	ATTTGCAGCATAT CTTGAACCATTC	
BS treated 5fC spike-in forward strand	N/A	ATTTGTAGTATATT TTGAATTATTTAAA GAG	CACTCATAACTTC CTAACCATCTTC

Table 6. List of primers used for 5fC MAB-seq amplicon preparation.
N/A 1st PCR primer: these amplicons that were prepared by a single PCR (instead of a two-step PCR).

MAB-seq and 5fC DIP-seq spike-ins	
5fC DIP 5fC spike-in	
Component	**Sequence (5'->3')**
Forward strand	GTGCTTGTTTGGCATTTCATTATAGCTATGAGCATCAAGATAAGATCAAAGCAATAG TTCACGCTGAAAGTGTAGTAGATGTGATTGAATCATGGGATGAATGGCCTGATATT GAAGAAGATATTGCG
Reverse strand	TATTCTCCAAAACCATTTTTTCTCCTTCTTCAGATTTGATCAACGCAATATCTTCTTC AATATCAGGCCATTCATCCCATGATTCAATCACATCTACTACACTTTCAGCGTGAAC TATTGCTTTGATCTTATCTTGATGCTCATAGCTATAATGAAATGCCAAACAAGCAC
Product	GTGCTTGTTTGGCATTTCATTATAGCTATGAGCATCAAGATAAGATCAAAGCAATAG TTCACGCTGAAAGTGTAGTAGATGTGATTGAATCATGGGATGAATGGCCTGATATT GAAGAAGATATTGCGTTGAT5FCAAAT5FCTGAAGAAGGAGAAAAAATGGTTTTGGA G AATA
MAB-seq 5fC spike-in	
Component	**Sequence (5'->3')**
Forward strand	ATTTGCAGCATATCTTGAACCATTCAAAGAGAAAGGTGAAGTTCGTCGTCCGACAT TATCATGGCCTCGTGAAATCCCGTTAGTAAAAGGTGG
Reverse strand	CACTCATAACTTCCTAACCATCTTCGTACAACGTCAGGTTTACCACCTTTTACTAAC GGGATTTCACGAGGCCATGATAATGTCGGACGACGAACTTCACCTTTCTCTTTGAA TGGTTCAAGATATGCTGCAAAT
Product	ATTTGCAGCATATCTTGAACCATTCAAAGAGAAAGGTGAAGTTCGTCGTCCGACAT TATCATGGCCTCGTGAAATCCCGTTAGTAAAAGGTGGTAAA5FC5FCTGA5FCGTTG TA5FCGAAGATGGTTAGGAAGTTATGAGTG
5fC DIP unmodified spike-in	
Component	**Sequence (5'->3')**
Forward primer	AGTATTCAACATTTCCGTGTCGCC
Reverse primer	CTTACCGCTGTTGAGATCCAGTT
Product	AGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCT GTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGG TGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAG

Table 7. MAB-seq and 5fC DIP-seq spike-ins.
The synthetic strands or primer sequences used in preparation of each spike-in, as well as the forward strand of the final product are listed.

ChIP qPCR primers			
Target gene	**Forward primer sequence (5' -> 3')**	**Reverse primer sequence (5' -> 3')**	**Roche UPL Probe #**
iMet	CCCTCCCAATAAGCAGAGTG	GTTTCGATCCATCGACCTCT	107
Gly	CTTCCGCTGTCACAGAACG	AAGGCAGCTATGCTCACCA	119
Cys	GGTGCCCCTGTTAGGAGAA	GTCGGCTTTTCACTCTACTGC	69
Thr	CCCCGTTACCAGCAGAGTAG	CACACAAAAGATGAGGATCTCG	82
eMet	CGCTAACGAGGCTGCAAG	GTAAGGGTGTATGCAGTAGAGCAC	163
Asp	GGCATAGCTAATTGCTGTTCACT	GACTTTGAATCCAGCGATCC	31
Arg	AGCCTCCATAGCAAACATGC	TGAAACGTATTAAAAGTGGACAAAA	31
5S rRNA	GCACGTCAGGACCGCTAC	CCGAAAGATGGTGAACTATGC	6
U6 snRNA	TTTAAGGGTCGGACGATTTG	CCATGTCAATAGCCCTACCAA	111
Sox2	AGAAAGTGAGGCAGCAACG	CACAGTTTCCCAAACTCTCTTCA	70
Sox17a	GCCACAGGCTGAGCAAAT	CTCTGGCATCCCATCATACC	26
Tbxt	GGCTCTACGCCTCATTACTCA	GAGATCCTGTGGCTGGTGA	150
Xnr5	TTCATCTGTCCTTCATGGAAAA	GCACCATGTAGACTGGATATTTCA	60
Xnr6	GATCCATCTGTTGCACAAGG	TCTGCATCATTCTGGAGATATTG	79
5.8S rRNA	AGTCAAGTTTGATCGTCTTCTCG	TTAGTGAGGTCCTCGGATCG	81
18S rRNA	CGAAAGTTGATAGGGCAGACA	GCCGCTTTGGTGACTCTAGATA	141
FoxD3	AGCCCCTGGCCCTTAGTAG	TGAGGTTTGTGGGGACACTT	46

Table 8. List of ChIP-qPCR primers and Roche UPL probe numbers.

Northern blot probes	
Target gene	**Sequence (5' -> 3')**
tRNA-iMet	TCCGCTGCGCCACTCTGCT
tRNA-Gly	GAATCTTGCATGATA
5.8S rRNA	AGCTAGCTGCGAGAATTAGTGT

Table 9. List of northern blot probes.

Cloning primers	Forward primer (5' -> 3')	Reverse primer (5' -> 3')
RLuc-a	TGCCTGGGGACGTCGGAGCAA TGACTTCGAAAGTTTATGATC	GAGGCCATGATAATGTCGGACGACG AAC
RLuc-b	TCCGACATTATCATGGCCTCGT GAAATC	CTATAGTGTCACCTAAATCATTATTAT TGTTCATTTTTGAGAAC TCG
Xt_iMet-iMet-Tyr	TGGCGGCACCCTATCAAAAGAA GCG	GGCGTGTGTGTATAGGCAAACTAGG GTGAC
Xt_iMet-iMet	CCAGTCGGGAAACCTGTCGTG CCAGGGCGTGTGTGTATAGGC AAAC	CGAGAATAGATCTTGGTAGCAGAGGA TGGTTTCGATC
Xt_Tyr	CAAGAGATTGATCTCCCTTTTT GGCAGGCAG	CGCGTTGGCCGATTCATTAATGCAGT GGCGGCACCCTATCAA AAG
Barcode	CCTCTGCTACCAAGATCTATTC TCGGCAATC	GCCAAAAAGGGAGATCAATCTCTTGG AGGC
iMiMT-SacI-1	CCAGTCGGGAAACCTGTCGTG CCAGGGCGTGTGTGTATAGGC AAAC	GCCACTCTGCTTGAGCTCAGATGCTG GTGAG
iMiMT-SacI-2	AGCATCTGAGCTCAAGCAGAGT GGCGCAGCG	CGCGTTGGCCGATTCATTAATGCAGT GGCGGCACCCTATCAA AAG
iMet_Nb.Bpu10I	CTCTCACCAGCACCTGAGCTCA AGCAGAGTGGC	AAGTCGCTCCGTAGATTGGCTCAGGT AGATCTTGGTAGCAGAGGATGGTTTC
B-box_mut_RV	N/A	TAGATTGGCTCAGGTAGATCTTGGTA GCAGAGGATGTGAGATCTGGATCGA CCTC
iMet-100bp2	CGGGAAACCTGTCGTGCCAGT GTTGAGATTATGATTGAACTAAT G	TGGCCGATTCATTAATGCAGAGATTC TAGGCTGGTTTG

Table 10. List of primers used for cloning.

Oligos used in cloning	Oligo sequence (5' -> 3')
Barcode sequence	AGATCTATTCTCGGCAATCTACGGAGCGACTTGATTATCAACAGCTGTC TAGCAGTTCTAATCTTTTGCCAACATCGTAATAGCCTCCAAGAGATTGAT C
mut-B-box-top-strand	GAG TCC TCT CAC CAG CAC CTG AGC TCA AGC AGA GTG GCG CAG CGG AAG CGT GCT GGG CCC ATA ACC CAG AGG TCG ATC CAG ATC TCA CAT CCT CTG CTA CCA AGA TCT ACC TGA GCC AAT CTA CGG A
mut-B-box-bottom-strand	TCC GTA GAT TGG CTC AGG TAG ATC TTG GTA GCA GAG GAT GTG AGA TCT GGA TCG ACC TCT GGG TTA TGG GCC CAG CAC GCT TCC GCT GCG CCA CTC TGC TTG AGC TCA GGT GCT GGT GAG AGG ACT C

Table 11. List of synthetic oligos used for cloning.

***tRNA-iMet* strand for strand exchange**	
DNA modification	**Sequence (5' -> 3')**
Unmodified C	TCAGGTAGATCTTGGTAGCAGAGGATGGTTTCGATCCATCGACCTCTGGGTTATG GGCCCAGCACGCTTCCGCTGCGCCACTCTGCTTGAGC
5mC	TCAGGTAGATCTTGGTAGCAGAGGATGGTTT**5mC**GATCCAT**5mC**GACCTCTGGGT TATGGGCCCAGCACGCTTCCGCTGCGCCACTCTGCTTGAGCC
5fC	TCAGGTAGATCTTGGTAGCAGAGGATGGTTT**5fC**GATCCAT **5fC**GACCTCTGGGTTATGGGCCCAGCACGCTTCCGCTGCGCCACTCTGCTTGAGC

Table 12. Sequence of *tRNA-iMet* strand used in strand exchange assay.
Unmodified, 5mC and 5fC versions of the gene's forward, non-coding strand sequence are listed.

RT-qPCR primers			
Target gene	**Forward primer sequence (5' -> 3')**	**Reverse primer sequence (5' -> 3')**	**Source**
U6 snRNA	TCGCTGGAGTTTCGTG CTT	ATTTGCGTGTCATCCTT GCG	This work
Oocyte *5S rRNA*	GGCTTGGTTAGTACCT GGA	AAAGCCTACAACAACTGGT	This work
5.8S rRNA	CATCGATGAAGAACGCAGCT	AAGGTGCGTTCGAGGTGT	This work
tRNA-iMet transgene	GCAGAGTGGCGCAGCG	AGCAGAGGATGGTTTCGATCCA	This work
Barcode	TCTATTCTCGGCAATCTACGG	CAATCTCTTGGAGGCTATTACG	This work
RLuc	CCAGGATTCTTTTCCAA TGC	CTTGCGAAAAATGAAGACCTTT	This work
Irx2	AATCACCAACAGGACGGAAG	GATCCGAGGTGGCTAT TTCA	Wagner et al., 2017
Meis3	TTATGGGCACCCTCTGT TTC	AATGAGTCCGGCTGTT GTCT	Wagner et al., 2017
FoxD3	CCCAGGCAAGGGAAACTACT	AATGAGTCCGGCTGTT GTCT	Wagner et al., 2017
Gapdh	ACCCAGAAGACTGTGGATGG	GTTGAGGCAGGGATG ATGTT	Wagner et al., 2017

Table 13. List of RT-qPCR primers.

Oligo pulldown	
Oligo	**Sequence (5' -> 3')**
Unmodified	Biotin-GGC CTC TGT GCG GAA GCC GGC GTT TGC ACC
Unmodified Reverese Complement	GGT GCA AAC GCC GGC TTC CGC ACA GAG GCC
5fC	Biotin-GGC CTC TGT G5fCG GAA GC5fC GG5fC GTT TGC ACC
5fC Reverse Complement	GGT GCA AA5fC GC5fC GGC TTC 5fCGC ACA GAG GCC

Table 14. List of oligos for pulldown with *X. laevis* ZGA (st9) nuclear protein extracts.

Fluorescence polarization	
Oligo	**Sequence (5' -> 3')**
Unmodified	6-FAM-CAGAGGTCGATGGATCGAAACCATCCTC
Unmodified Reverse Complement	Biotin-GAGGATGGTTTCGATCCATCGACCTCTG
5mC	6-FAM-CAGAGGT**5mC**GATGGAT**5mC**GAAACCATCCTC
5mC Reverse Complement	Biotin-GAGGATGGTTT**5mC**GATCCAT**5mC**GACCTCTG
5hmC	6-FAM-CAGAGGT**5hmC**GATGGAT**5hmC**GAAACC ATCCTC
5hmC Reverse Complement	Biotin-GAGGATGGTTT**5hmC**GATCCAT**5hmC**GA CCTC TG
5fC	6-FAM-CAGAGGT**5fC**GATGGAT**5fC**GAAACCATCCTC
5fC Reverse Complement	Biotin-GAGGATGGTTT**5fC**GATCCAT**5fC**GACCTCTG

Table 15. List of oligos used in mCry1/hCRY2 fluorescence polarization assay.
Forward oligos are 5' FAM labelled and reverse oligos are 5' biotin labelled.

4.1.12. Software

GraphPad Prism v.9; IGV v2.16.0; Image Lab v. 6.1.0 (Bio-Rad); ImageJ 64 v5; LightCycler 480 v. 1.5.1.62 (Roche); MaxQuant v2.4.2.0 ; NEBuilder Assembly Tool v.2.10.1 (NEB, www.nebuilder.neb.com); SparkControl v.3.2 (Tecan); Typhoon FLA 9500 Control Software v. 1.0 (GE Healthcare)

4.2. Methods

4.2.1. *Xenopus* animals

X. laevis and *X. tropicalis* adult animals were obtained from EXRC, NXR and Nasco. *X. laevis* aquarium conditions: <0.05 ppm NH_3/NH_4, 0.05 ppm NO_2, 5 ppm NO_3, 990 – 1000 conductivity, 8.5 pH, 18ºC (water and room temperature), 19ºdH total hardness, 10ºdH carbonate hardness. *X. tropicalis* aquarium conditions: 0 ppm NH_3/NH_4, 0 ppm NO_2, 25 ppm NO_3, 990 – 1000 conductivity, 7.5 pH, 24ºC (water and room temperature), 19ºdH total hardness, 10ºdH carbonate hardness. Light/dark cycle is 12 h/12 h for both species.

Xenopus experiments were approved by the state review board of Rheinland Pfalz, Germany (Landesuntersuchungsamt, reference number 23177–07/A17-5-002 HP) and performed according to federal and institutional guidelines.

4.2.2. *Xenopus* priming

X. laevis females were injected with 150 U human chorionic gonadotropin (hCG) in 100 µl volume at the dorsal lymph sacs area, at least 16 h before *in vitro* fertilization (IVF). The animals were kept O/N at 18ºC.

X. tropicalis females were injected with 10 U hCG in 100 µl volume at the dorsal lymph sack area 24 h before IVF and injected again with 100 U hCG in 100 µl volume 4 h before IVF. The animals where kept at 21ºC.

4.2.3. *Xenopus* testis isolation

X. laevis and *X. tropicalis* males were anesthetized in 0.15% MS-222 for 10 min and sacrificed by transection between the brainstem and the spinal cord. Testis were dissected and kept in 66% L-15 in 1X Ringer buffer supplemented with 1:1000 Pen Strep until used for IVF.

4.2.4. *In vitro* fertilization

Eggs were collected in glass Petri dishes by gently squeezing the females. Sperm solution was prepared by homogenizing the testis in 1X Ringer buffer using a microtube pestle to release the spermatozoa. The eggs were then mixed with ~300 µl sperm solution and incubated for 3 min in RT in order to attach sperm onto the eggs. The eggs were then flooded with 0.1X Barth solution to activate the sperm and fertilized for 20 min in RT. To remove the jelly coat, embryos were transferred into 200 ml beakers containing ~100 ml 2% L-Cysteine solution pH 7.8-8 for *X. laevis* or 3% L-cysteine solution pH 7.6-7.8 for *X. tropicalis* and incubated for 6-10 min with gentle shaking at 80 rpm on a tabletop orbital shaker. The embryos were subsequently washed 3x with cool tap water and transferred into 10 mm Petri dishes, containing 0.1X Barth. Unfertilized/dead embryos were picked out with a transfer pipet. Embryos were kept at 16ºC until microinjections.

4.2.5. Embryo microinjections

Glass capillaries were pulled to produce microinjection needles using a P-97 micropipette puller with the following conditions: heat 575, pull 80, velocity 180, delay 90. Needles were loaded to a micromanipulator connected to a microinjector and calibrated under the stereomicroscope to release 5 or 10 nl droplets in 0.3 – 0.8 sec injection time upon 10 PSI pressure, using a micrometer ruler and mineral oil.

The embryos were transferred into 10 mm glass Petri dishes with 16ºC 0.3X Barth and microinjected centrally at the AP (unless otherwise stated) while in 1- or 2-cell stage. After microinjection, the embryos were kept in 0.3X Barth in RT for 2 h. Subsequently, the media were

changed to 0.1X Barth and the embryos were transferred to an incubator to develop. Incubation was at lower temperature for slower development or at higher temperature for faster development. Incubation temperature: 15-20ºC for *X. laevis* and 18-24ºC for *X. tropicalis*.

Total amounts of mRNA injected per embryo was 1 ng *TDG* mRNA for TDG overexpression and 1 ng TDG^{N140A} mRNA for dTDG overexpression. Control mRNA used for injections was bovine preprolactin (Ppl). ATG- blocking Morpholino (Mo) antisense oligonucleotides (Gene Tools) were designed to target the respective genes. Total amounts of Mo injected per embryo was 30 ng Tet2 and 30 ng Tet3 Mo for Tet2/3 KD, 30 ng *Tdg* Mo for Tdg KD or 60 ng of standard control Mo. For Pol III inhibition, *X. tropicalis* embryos were injected with 2 ng α-amanitin 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.

4.2.6. *Xenopus* embryos imaging

X. laevis and *X. tropicalis* embryos were fixed in MEMFA for 2 h in RT with rocking shaking and subsequently washed 3x with 0.1X Barth. Embryo images were obtained with a ZEISS SteREO Discovery.V20 stereomicroscope and an AxioCam MRc 5 camera.

4.2.7. Genomic DNA extraction 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) for 10 min in RT. DNA was subsequently ethanol precipitated with ammonium acetate. Dr. Michael Musheev carried out quantification of 5mC and oxC as previously described (Schomacher et al., 2016).

4.2.8. 5fC DNA immunoprecipitation (5fC DIP)

5fC DIP 5fC spike-in was prepared using two complementary synthetic oligonucleotides, one containing a 5’ overhang with two cytosines (see section 4.1.11, Table 7). 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 mixed in 1:1 ratio to produce 50% hemi-modified 5fC spike-in which was added to the 5fC DIP samples. 5fC-DIP unmodified spike-in, which corresponds to part of the AmpR sequence, was prepared by PCR (see section 4.2.26). All spike-ins were purified with DNA Clean & Concentrator-5 (Zymo research) according to the instructions of the supplier.

Genomic DNA from *X. tropicalis* st9 embryos was extracted using the DNeasy Blood and Tissue kit (Qiagen). 28 µg DNA was sheared to ~200 bp using a Covaris E220 focused ultrasonicator. 5 pg unmodified spike-in and 20 pg 5fC spike-in were added 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 5 µg 5fC (61227, Active Motif) or IgG (ab46540, Abcam) antibody O/N at 4°C with overhead shaking. 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 h 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 h at 50°C and DNA was purified with DNA Clean & Concentrator-5 (Zymo research) according to the supplier's instructions.

4.2.9. Malononitrile treatment

5fC DIP samples were mixed with 150 mM malononitrile in 300 mM NH_4Ac buffer in final volume 50 µl. The samples were mixed by vortexing and incubated at 37°C, 850 rpm for 20 h in a thermoblock with heated lid to avoid evaporation. Subsequently DNA underwent ethanol precipitation and the DNA pellet was resuspended in 10 µl nuclease-free H2O.

For WGS samples, 2 µg *X. tropicalis* gDNA were mixed with 0.2 ng 5fC DIP unmodified spike-in and 0.2 ng 5fC DIP 5fC spike-in. The DNA was mixed with 150 mM malononitrile in 300 mM NH_4Ac buffer in final volume 100 µl and the mix was processed as described above.

For DNA polymerase amplification efficiency test (Figure 2.3C), 5 µg unmodified or 5fC modified *tRNA-iMet* strand used in strand exchange (Table 12) were mixed with 150 mM malononitrile in 300 mM NH_4Ac buffer in final volume 100 µl and the mix was processed as described above.

4.2.10. Malononitrile assisted 5fC DIP and WGS sequencing

Library preparation and DNA sequencing were performed by IMB's Genomics Core Facility, as follows: DNA library preparation was performed with xGen ssDNA & Low-Input DNA Prep Kit (June 2022) from IDT according to the supplier's instructions. Reagent W2, buffer W3 and enzyme W4 were replaced by Deep Vent (exo-) DNA Polymerase kit (NEB). 5fC DIP libraries were prepared with 0.4 ng DNA and amplified in 14 PCR cycles and WGS libraries were prepared with 14 ng DNA and amplified in 7 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 4.0 Fluorometer (Thermo Scientific). All samples were pooled in equimolar ratio and sequenced on a NextSeq 2000 P3 flow cell, in paired-mode for 2x 150 cycles plus 8 + 8 cycles for the dual index read.

Bioinformatics analysis of malononitrile 5fC DIP-seq and WGS was performed by Dr. Alexandr Gopanenko, as follows: Raw sequencing data were quality-checked with fastqc (v. 0.11.9). According to the sequencing kit manual, R2 reads were trimmed using cutadapt (v. 4.0) with (–u 15). Trimmed sequencing reads were mapped using bowtie2 (v. 2.4.5) (--local) with scaffold_58 or scaffold_109 sequences of *X. tropicalis* v.7.1 genome assembly (www.xenbase.org) as reference. The strand-specific bigwig tracks were generated with deepTools (v.3.5.1) command bamCoverage --binSize 10 --filterRNAstrand reverse.

4.2.11. Methylation assisted bisulfite sequencing (MAB-seq)

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 the forward strand, and several cytosines in CpN (N=A, T, C) context. Spike-in DNA controlled 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 (Table 7). 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 (Table 6). 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.

Library preparation and DNA sequencing were performed by IMB's Genomics Core Facility, as follows: 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.

4.2.12. 5mC and 5fC base resolution profiling

Bioinformatics analysis described in this section was performed by Dr. Alexandr Gopanenko.

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. tropicalis* 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.

4.2.13. Chromatin immunoprecipitation (ChIP)

50-200 *X. tropicalis* embryos were cross-linked in 1% methanol-free formaldehyde for 30 min in RT. Cross-linking was quenched with 125 mM glycine for 30 min in RT and the embryos were washed 3x with 0.3X Barth. Embryos were subsequently homogenized with trituration in 200 µl sonication buffer and chromatin was sonicated to ~500 bp fragments using a Biorupter Plus (Diagenode) and centrifuged at 15,000 g for 5 min to remove yolk and pigment. 10% chromatin was kept aside as input and 100 µl chromatin was added to 100 µl incubation buffer. 5 µg Pol II (C15200004, Diagenode), Pol III (POLR3F; ab151495, Abcam), Smarca4 (21634-1-AP, Proteintech) or IgG (ab171870, Abcam) antibody were added to chromatin and incubated O/N at 4°C with overhead shaking. 50 µl Pierce™ Protein G Magnetic Beads (Thermo Scientific) per reaction were blocked with incubation buffer + 0.1% BSA for 10 min at 4°C with overhead shaking. The beads were washed 2x in incubation buffer for 5 min at 4°C, the chromatin was added and the samples were incubated O/N at 4°C with overhead shaking. Beads were washed sequentially with wash buffer 1, wash buffer 2, wash buffer 3, wash buffer 1 and TE buffer. 200 µl elution buffer was added to the beads and incubated for 15 min in RT with overhead shaking. Elution was repeated with an extra 200 µl elution buffer. 16 µl 5 M NaCl was added to elution and input samples. Cross-linking was reversed by incubating at 65°C O/N with shaking. DNA underwent phenol/chloroform extraction and ethanol precipitation with ammonium acetate and glycogen. To use for qPCR, chromatin was diluted 10x with nuclease-free H2O.

4.2.14. RNA extraction

X. laevis or *X. tropicalis* embryos were homogenized in 700 µl Qiazol. 200 µl chloroform were added and the samples were incubated 3 min in RT. The samples were centrifuged at 11,000 g for 15 min at 4°C and the top, aqueous phase was transferred into fresh Eppendorf tubes and mixed with 400 µl isopropanol. The samples were incubated 10 min in RT and RNA was precipitated at 11,000 g for 15 min at 4°C. The pellet was washed twice with 70% ethanol and resuspended in 30 µl nuclease-free H_2O.

4.2.15. *In vitro* transcription

For Tdg, *TDG*^{N140A} (dTDG) and *Ppl* (control) mRNA preparation, pCS2-flag-hTDG (Niehrs group), pHH25-HA-Tdg-N140A (Hardeland 2000) and pRN-Ppl (Niehrs group) plasmids were linearized with NotI (NEB), XbaI (NEB) and SfiI (NEB) restriction enzymes, respectively, according to the instructions of the supplier. The linearized plasmids were purified using the QIAquick PCR Purification kit (Qiagen) and used as template for *in vitro* transcription. *TDG*, *TDG*^{N140A} and *Ppl* mRNA was produced using the MEGAscript SP6, T7 and T3 transcription kits (Thermo Scientific),

respectively, according to the instructions of the supplier. DNA was digested with RNase-free DNase I (Qiagen) and RNA was purified using the RNeasy Mini kit (Qiagen) according to the supplier's instructions. RNA amount and purity were estimated with a Nanodrop 2000 spectrophotometer.

4.2.16. Reverse transcription (RT)

For relative quantification of transcript levels, 5 μg total RNA were treated with 2 μl DNase I Amplification Grade 1 unit/ml (Sigma-Aldrich) in the supplied Reaction buffer in final volume 20 μl. The reaction was allowed to proceed for 30 min in RT, before it was inhibited by the addition of the supplied Stop buffer and heat inactivation at 70ºC for 10 min. After DNase I digestion, 10 μl RNA (~2.5 μg) were mixed with 2 μl 100 μM random primers and 1 μl 10 mM dNTPs and the mix was heated at 65ºC for 5 min and quickly cooled down to 4ºC to destroy the secondary structures of the nucleic acids. Subsequently, reverse transcriptase buffer (IMB Protein Production core facility) to 1X, 2 μl 100 mM DTT, 1 μl RiboLock RNase inhibitor and 100 U reverse transcriptase (IMB Protein Production core facility) per reaction were added. The reaction was incubated at 25ºC for 10 min, followed by reverse transcription at 42ºC for 95 min, heat inactivation at 70ºC for 15 min and cooling at 4ºC. To use for qPCR, the cDNA was diluted 5x with nuclease-free H2O.

4.2.17. qPCR

qPCR was conducted in 384-well format using the LightCycler480 System from Roche. All reactions were performed in technical duplicates.

For ChIP-qPCR, the Universal Probe Library (UPL) System was used. Each reaction contained 1X LightCycler® 480 Probes Master (Roche), 550 nM forward primer, 550 nM reverse primer, 100 nM UPL probe and 5 μl chromatin, in final volume 11 μl. The samples underwent an initial denaturation at 95ºC for 10 min (ramp rate 4.8ºC/sec), followed by 40 cycles of denaturation at 95ºC for 10 sec (ramp rate 4.8ºC/sec), annealing at 60ºC for 15 sec (ramp rate 2.5ºC/sec) and elongation at 72ºC for 10 sec (ramp rate 4.8ºC/sec), with single signal acquisition in each cycle. Finally, the samples were cooled at 40ºC for 30 sec (ramp rate 2.5ºC/sec). ChIP-qPCR primers and UPL probes are listed at Table 8.

For RT-qPCR, the SYBR Green system was used. Each reaction contained 1X LightCycler 480 SYBR Green I Master (Roche), 600 nM forward primer, 600 nM reverse primer and 5 μl cDNA, in final volume 11 μl. The samples underwent an initial denaturation at 95ºC for 10 min (ramp rate 4.8ºC/sec), followed by 40 cycles of denaturation at 95ºC for 10 sec (ramp rate 4.8ºC/sec), annealing at 58ºC for 15 sec (ramp rate 2.5ºC/sec) and elongation at 72ºC for 10 sec (ramp rate

4.8°C/sec), with single signal acquisition in each cycle. For the melting curve, the samples were denatured at 95°C for 10 sec (ramp rate 4.8°C/sec), followed by melting at 58°C for 1 min (ramp rate 2.5°C/sec), melting at 97°C with ramp rate 0.11°C/sec and continues signal acquisition per 5°C, and finally cooling at 40°C for 30 sec (ramp rate 2.5°C/sec). RT-qPCR primers are listed at Table 13.

Cp values were calculated using the Abs Quant/2nd Derivative Max analysis mode of the LightCycler 480 software v. 1.5.1.62. ChIP-qPCR values were normalized over IgG and *tRNA-Gly* levels. *U6* and *5S* RT-qPCR values were normalized over *5.8S* levels. *Irx2*, *Meis3* and *FoxD3* RT-qPCR values were normalized over *Gapdh* levels. For transgene plasmid optimization (Figure 2.11D-E), barcode and *tRNA-iMet* transcription levels were normalized over *RLuc*. For *tRNA-iMet* transgene plasmid assay (Figure 2.12A-B, E), 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.

4.2.18. Northern blot

10 µg *X. tropicalis* RNA per sample were boiled at 70°C 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 TBE buffer. Transfer was performed onto a Hybond-N+ membrane with a Trans-Blot Turbo system at 25 V for 10 min. For chemical crosslinking, the membrane was placed on a Wattman paper soaked in crosslinking buffer RNA facing up and incubated at 55°C for 2 h. The blot was blocked in 100 µg/ml heat denatured salmon sperm DNA in hybridization buffer at 37°C for 2 h. Probes (Table 9) were 5' end-labelled with γ32P ATP and T4 Polynucleotide kinase according to manufacturer's instructions at 37°C for 1 h and purified with Microspin G-25 Columns. Blots were probed with 250 ng radioactive probe in hybridization buffer at 37°C O/N and washed 3x in hybridization buffer at 42°C for 30 min before autoradiography. *tRNA-iMet* and *tRNA-Gly* levels were normalized over *5.8S rRNA* levels.

4.2.19. Western blot

X. tropicalis embryos were dissolved in 10 µl NP-40 lysis buffer per embryo. Protein was extracted by the addition of 2 volumes Freon and centrifugation at 20,000 g at 4°C for 15 min. Protein was mixed with LDS buffer, boiled at 95°C for 5 min and quickly transferred to ice. Protein was loaded at a 7.5% (for Brg1 detection) or 12% (for TDG detection) PAGE and run in SDS running buffer at 100 V for up to 2 h. Proteins were then transferred to a PVDF membrane using the Trans-Blot Turbo (BioRad) transfer apparatus for 10 min at the "High MW bands" program for Brg1 or 7 min at the "Mixed bands" program for TDG detection. The blots were blocked with 5% milk – TBST for

1 h and probed with primary antibody O/N at 4°C, washed 3x in TBST and probed with secondary antibody for 1 h in RT. Antibodies were diluted in 5% milk – TBST. For antibody working concentrations, see section 4.1.7. In order to probe a single blot for TDG (46 kDa) and α-tubulin (50 kDa), the blot was first probed for TDG and then stripped with Restore Western Blot Stripping buffer (Thermo Scientific) for 10 min in RT, washed 3x in TBST, blocked and re-probed for α-tubulin. 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.

4.2.20. Restriction digest

Plasmids were linearized for cloning with restriction digest. 20-50 U restriction enzyme were combined with plasmid DNA (up to 5 µg) and CutSmart buffer (NEB) in final volume 50 µl. The digests were incubated at 37°C for 1-2 h, followed by enzyme deactivation at 80°C for 20 min. Linearized vectors were gel extracted and purified with QIAquick® Gel Extraction kit (Qiagen), according to the instructions of the supplier.

4.2.21. Gibson assembly

Overlapping fragments were designed using the NEBuilder Assembly Tool v.2.10.1 and generated by restriction digests or PCR. 0.05 pmol backbone were combined with 5-7 fold molar ratio insert and 1X Gibson assembly MasterMix (IMB Protein Production Core Facility) and mixed by trituration. The mix was incubated at 50 °C for 2 h.

4.2.22. Bacterial transformation

XL1-blue bacteria (Agilent) were thawed on ice for 15 min. 3 µl Gibson reaction was added to the bacteria, mixed by trituration and incubated for 30 min on ice. The bacteria were then heat-shocked at 42°C for 45 sec and immediatelly transferred on ice for 5 min. Subsequently, 500 µl LB media without antibiotics were added to the cells and gently mixed by trituration. Bacteria were incubated at 37°C for 1 h with shaking at 500-700 rpm. 100 µl bacterial suspension were plated on an LB agar – Ampicilin plate and incubated O/N at 37°C for colony formation.

4.2.23. Plasmid maintenance

Bacterial stocks were prepared by mixing liquid bacterial culture with 80% glycerol in 1:1 ratio and maintained at -80°C. To recover the bacteria, a small amount of frozen stock was scraped off using a pipet tip and used to inoculate 5 ml liquid cultures of LB media supplemented with the appropriate antibiotic. The bacteria were grown O/N at 37°C with orbital shaking.

4.2.24. Plasmid DNA isolation

Plasmid DNA was isolated with the QIAprep Spin Miniprep kit and QIAfilter Plasmid Midi kit (Qiagen) according to the instructions of the supplier. DNA amount and purity were estimated with a Nanodrop 2000 spectrophotometer. Plasmid sequences were confirmed by Sanger sequencing by StarSeq GmbH.

4.2.25. Cloning strategy

To prepare pCS2+RLuc, RLuc gene was PCR amplified in two fragments using AccuPrime DNA polymerase (Thermo Scientific), RLuc-a and RLuc-b primer sets and pRL-TK as template DNA. The two fragments were necessary to remove the XcmI restriction sequence inside the gene. pCS2+ backbone was linearized with HindIII HF (NEB). Linearized backbone and the two RLuc PCR products were combined in molar ratio 1:3:3 and underwent Gibson assembly.

To prepare ZeroBlunt-iMet-iMet-Tyr, a 993 bp sequence containing two consecutive *tRNA-iMet* genes followed by a *tRNA-Gly* gene was PCR amplified from *X. tropicalis* genome with Q5 DNA polymerase (NEB) and the Xt_iMet-iMet-Tyr primer set. The insert was cloned into Zero Blunt® PCR Cloning vector (Invitrogen) and propagated in One Shot® TOP10 cells, according to the instructions of the supplier.

To prepare pCS2+RLuc-iMet-iMet-Tyr, the Zero Blunt iMet-iMet-Tyr construct was used as template DNA for two PCR products: the first included the two *tRNA-iMet* genes, and the second the *tRNA-Tyr* gene. PCR was performed with AccuPrime DNA polymerase and Xt_iMet-iMet and Xt_Tyr primer sets. 5' and 3' overhangs overlapping with the iMet-iMet and Tyr PCR products, respectively, were added to a synthetic 100 bp barcode sequence by PCR with AccuPrime DNA polymerase and the Barcode primer set. pCS2+RLuc was linearized with PvuII HF (NEB). The three PCR products and the linearized backbone were combined in molar ratio 7:7:7:1 and underwent Gibson assembly.

To be compatible with strand exchange, pCS2+RLuc-iMet-iMet-Tyr had to be altered so that the non-coding strand of the second *tRNA-iMet* gene would be flanked by the Nb.Bpu10I recognition sequence. First, a SacI recognition sequence was inserted upstream of the second *tRNA-iMet* gene. For this, the iMet-iMet-barcode-Tyr sequence was amplified in two fragments, introducing the SacI sequence between them. PCR was performed with AccuPrime DNA polymerase, using pCS2+RLuc-iMet-iMet-Tyr as template DNA and iMiMT-SacI-1 and iMiMT-SacI-2 primer sets. The two PCR fragments were combined with PvuII HF linearized pCS2+RLuc in molar ratio 3:3:1 and underwent Gibson assembly.

Next, pCS2+RLuc-iMet-iMet-Tyr-SacI was linearized with SacI (NEB) and BglII (NEB) to excise the second *tRNA-iMet* gene. *tRNA-iMet* gene sequence flanked by Nb.Bpu10I recognition sites was prepared by PCR, using AccuPrime DNA polymerase, pCS2+RLuc-iMet-iMet-Tyr-SacI as template DNA and iMet_Nb.Bpu10I primers. Backbone and insert were combined in molar ratio 1:5 for Gibson assembly. The resulting plasmid is named pBpu10I and referred to as "Complete" in the Results section (Figure 2.11B-E).

A version of pBpu10I with a mutated B-box in the second *tRNA-iMet* gene (pBpu10I-mut-B-box) was prepared using SacI/BglII linearized pCS2+RLuc-iMet-iMet-Tyr-SacI vector as a backbone and a synthetic mutated B-box *tRNA-iMet* insert. The top and bottom synthetic strands were mixed at 1:1 ratio in annealing buffer and incubated at 95°C for 5 min, followed by slow cool down to RT in order to anneal. Gibson assembly was performed using 1:5 vector:insert molar ratio.

To prepare pCS2+iMet-iMet-Tyr-no_CMV (referred to as "no CMV' plasmid in the Results section) and pCS2+iMet-iMet-Tyr-no_CMV-mut-B-box, the CMV promoter and RLuc sequence were excised from pBpu10I and pBpu10I-mut-B-box by double restrictive digestion with BsrGI HF (NEB) and SalI HF (NEB). A small piece of the RLuc sequence was used as an insert to circularize the vector. The RLuc piece was PCR amplified using the RLuc_piece primer set and AccuPrime DNA polymerase. The linearized vector and the insert were combined in molar ratio 1:6 and underwent Gibson assembly.

To prepare pCS2+iMet-RLuc (referred to as "single *tRNA-iMet*" plasmid in the Result section) and pCS2+iMet-RLuc-mut-B-box, pCS2+RLuc linearized with PvuII HF was used as a backbone. *tRNA-iMet* and *tRNA-iMet* mutant B-box followed by the barcode sequence were PCR amplified using AccuPrime DNA polymerase, pBpu10I and pBpu10I-mut-B-box as template DNA and the iMet-100bp2 primer set. Linearized vector and insert were combined in molar ratio 1:5 and underwent Gibson assembly.

Finally, to prepare pCS2+iMet (referred to as "no CMV, single *tRNA-iMet*" plasmid in the Result section) and pCS2+iMet-mut-B-box, the CMV promoter and RLuc sequence were extracted from pCS2+iMet-RLuc and pCS2+iMet-RLuc-mut-B-box by double restrictive digestion with BsrGI HF and SalI HF. The linearized vector was re-circularized with the RLuc piece used in the production of pCS2+iMet-iMet-Tyr-no_CMV. Vector and insert were combined in molar ratio 1:6 and underwent Gibson assembly.

Cloning primers and oligo sequences are listed at Tables 10-11.

4.2.26. PCR

Cloning inserts were amplified by PCR using either AccuPrime DNA polymerase (Thermo Scientific) or Q5 DNA polymerase (NEB). For PCR with AccuPrime, 1X AccuPrime PCR buffer I, 200 nM of each primer, 50 – 100 ng template DNA and 1 µl AccuPrime DNA polymerase were combined in final volume 50 µl. Initial denaturation was at 94ºC/5 min, followed by 15-25 cycles of denaturation at 94ºC/30 sec, primer annealing at 55-65ºC/30 sec, and primer extension at 68ºC/45 sec. Final extension was at 68ºC for 2 min. For PCR with Q5, 1X Q5 DNA polymerase reaction buffer, 200 µM dNTPs, 500 nM of each primer, 100 ng gDNA template and 1 U Q5 DNA polymerase were combined in final volume 50 µl. Initial denaturation was at 98ºC/45 sec, followed by 30 cycles of denaturation at 98ºC/10 sec, primer annealing and extension at 72ºC/1 min. All PCR products were purified with QIAquick® PCR Purification kit (Qiagen), according to the instructions of the supplier.

5fC DIP unmodified spike-in, which corresponds to part of the AmpR sequence, was prepared using 100 ng pCS2+RLuc as DNA template, 0.2 µM primers, 200 µM dNTPs and 1.25 U OneTaq (NEB) in final volume 50 µl. Initial denaturation was at 94ºC for 2 min, followed by 15 cycles melting at 94ºC/30 sec, primer annealing at 53ºC/30 sec and primer extension at 68ºC/30 sec. The spike-in was purified with DNA Clean & Concentrator-5 (Zymo Research) according to the instructions of the supplier.

To test the amplification efficiency of different commercial DNA polymerases on 5fC-M template DNA (Figure 2.3C), unmodified and 5fC-modified *tRNA-iMet* strand used for strand exchange (see section 4.2.27) were treated with malononitrile (see section 4.2.9) and used as PCR template. For OneTaq (NEB) amplification, 10 ng template DNA was mixed with 1X OneTaq Standard Reaction Buffer, 0.2 µM primers, 200 µM dNTPs and 1.25 U OneTaq enzyme in final volume 50 µl. Initial denaturation was at 94ºC for 2 min, followed by 5-15 cycles melting at 94ºC/30 sec, primer annealing at 55ºC/30 sec and primer extension at 68ºC/30 sec. For PhusionU (Thermo Scientific) amplification, 10 ng template DNA was mixed with 1X PhusioU HF Buffer, 0.2 µM primers, 200 µM dNTPs and 0.01 U PhusionU enzyme in final volume 50 µl. Initial denaturation was at 98ºC for 30 sec, followed by 5-15 cycles melting at 98ºC/10 sec, primer annealing at 66.6ºC/30 sec and primer extension at 72ºC/20 sec. For Q5 (NEB) amplification, 10 ng template DNA was mixed with 1X Q5 reaction Buffer, 0.5 µM primers, 200 µM dNTPs and 0.01 U Q5 enzyme in final volume 50 µl. Initial denaturation was at 98ºC for 30 sec, followed by 5-15 cycles melting at 98ºC/10 sec, primer annealing at 69ºC/30 sec and primer extension at 72ºC/20 sec. For KapaHiFi HotStart U+ (Takara) amplification, 10 ng template DNA was mixed with 1X KAPA HiFi HotStart U+ Mix and

0.3 μM primers in final volume 25 μl. Initial denaturation was at 95ºC for 3 min, followed by 5-15 cycles melting at 98ºC/20 sec, primer annealing at 63ºC/15 sec and primer extension at 72ºC/15 sec. Finally, for Deep Vent (exo-) (NEB) amplification, 10 ng template DNA was mixed with 1X Thermopol Reaction Buffer, 0.25 μM primers, 200 μM dNTPs and 1.5 U Deep Vent (exo-) enzyme, in final volume 50 μl. Initial denaturation was at 95ºC for 2 min, followed by 5-15 cycles melting at 95ºC/30 sec, primer annealing at 57ºC/30 sec and primer extension at 72ºC/30 sec.

Cloning primers and oligo sequences are listed at Tables 10-11.

4.2.27. Strand exchange assay

Protocol adopted from Lühnsdorf et al., 2012. 50 μg pCS2+iMet were digested with 65 U Nb.Bpu10I in final volume 250 μl for 1 h at 37ºC, then another 50 U enzyme were added and the plasmid was digested for an additional hour. The enzyme was inactivated at 80ºC for 20 min. Subsequently, 10 μg digested *tRNA-iMet* plasmid were mixed with T4 DNA ligase buffer, 200X molar excess synthetic unmodified, 5mC-, or 5fC-modified strand (Table 12), 1 mM ATP and 125 U polynucleotide kinase in final volume 350 μl. The mix was split to 20 μl reactions and incubated at 37ºC for 30 min to phosphorylate the synthetic strand, followed by melting at 90ºC for 1 sec and cooling at 65ºC at a cooling rate 0.02ºC/sec and then quickly to 4ºC. After strand exchange, all aliquots with the same synthetic strand were pooled. 10 μl from each sample were loaded on a 0.8% agarose gel stained with GelRed, along with undigested plasmid as positive control and Nb.Bpu10I digested plasmid, no PNK mock and no synthetic strand mock as negative controls, to validate successful completion of all individual steps. After successful strand exchange was verified, each sample was concentrated to final volume 40 μl using the Amicon Ultra 30 centrifugal filters (Merk) according to the instructions of the supplier, and the concentrated plasmids were loaded to a 0.8% agarose gel and run to separation. The covalently closed plasmid was isolated from the gel with a blade and extracted using the QIAquick Gel Extraction kit (Qiagen) according to the instructions of the supplier. Plasmid DNA concentration and purity were estimated with Nanodrop 2000 spectrophotometer.

4.2.28. *In vitro* TDG glycosylase assay

The protocol was adapted from Schomacher et al., 2016. 75 nM *tRNA-iMet* plasmid were incubated with 750 nM recombinant TDG (1:10 molar ratio) in glycosylase buffer at 37ºC for 1 h. Mock samples were incubated with the buffer alone, without enzyme. Plasmid DNA was subsequently phenol/chloroform purified and ethanol precipitated.

4.2.29. Agarose gel electrophoresis

Agarose powder was dissolved in 1X TBE buffer and boiled for 3 min or until the agarose was completely dissolved. Agarose solution was mixed with 1:10,000 GelRed (Millipore), poured into agarose gel molds and let solidify for approximately 30 min. The gels were transferred into horizontal electrophoresis chambers filled with 1X TBE buffer and run in variable voltage for variable amount of time, based on gel thickness and target band size. Gel pictures were taken with ChemiDoc XRS+ System (Bio-Rad) using the Image Lab v. 6.1.0 software.

4.2.30. Oligo pulldown

To isolate nuclear protein, 5000 *X. laevis* st9 embryos were homogenized in 10 µl ELB buffer per embryo by trituration. 17.5 ml cushion was added into 30 ml ultracentrifuge tubes (70% of total volume) and 7.5 ml cell lysate was carefully placed on top of the cushion, by tilting the tube and slowly releasing cell lysate. The samples were spinned at 11,000 g, 4°C for 6 min, the supernatant was removed and the pellet was resuspended in 10 µl ELB buffer per embryo. The cushion/spin step was repeated once more, the supernatant was removed and the pellet was resuspended in 5 µl B+ buffer per embryo using a dounce homogenizer. Samples were incubated at 4°C for 1 h with overhead rotation, followed by spinning at 11,000 g, 4°C for 15 min to remove nuclear debris. The supernatant containing the nuclear proteins was transferred into fresh tubes.

The following pulldown protocol was performed by Dr. Lars Schomacher and Carola Scholz. Double strand (ds) DNA oligos with 5' end biotin label on the forward strand were used for the pulldown. A 5fC oligo carried three formylated cytosines in CpG context per strand, while an unmodified oligo with the same sequence was used as control (Table 14). For hybridization, equimolar ratio of forward and reverse oligos were mixed with SSC buffer in final volume 50 µl and incubated at 90°C for 15 sec, followed by 80°C for 3 min, 50°C for 15 min and 20°C for another 15 min.

For the pulldown, 20 µl Dynabeads MyOne Streptavidin C1 (Thermo Scientific) per sample were washed 3x with 1X B+W buffer using a magnetic rack and resuspended in 20 µl 2X B+W buffer. 100 pmol double strand oligo in final volume 20 µl per reaction were added and the samples were incubated at 25°C for 30 min with gentle shaking. Subsequently, the beads were washed three times with 1X B+W buffer and transferred into low binding Eppendorf tubes. The buffer was removed and the beads were resuspended in 20 µl protein binding buffer. Each sample was prepared in 4 technical replicates.

The samples were subsequently processed for proteomics analysis as previously described (Kappei et al. 2017).

4.2.31. Developmental proteome mass spectrometry

Protein extraction protocol was adapted from Gupta et al., 2018. *X. laevis* and *X. tropicalis* oocytes and st2, 7, 9, 10, 11, 13, 23 and 35 embryos were harvested for protein extraction (4 biological replicates, 6 *X. laevis* embryos and 12 *X. tropicalis* embryos per sample). The embryos were lysed in 1 ml lysis buffer by trituration and incubated 10 min on ice to allow the cytoskeleton protein disassemble. The samples were vortexed for 10 sec and centrifuged at 2,500 g at 4ºC for 4 min. The supernatant was transferred into fresh tubes and HEPES pH 7.2 was added to final concentration 100 mM. The proteins were denatured with 2% SDS. For cysteine protection, the samples were reduced with 5 mM DTT and incubated at 60ºC for 20 min with gentle shaking. After the samples cooled down, NEM was added to 20 mM and the samples were incubated in RT for 20 min. 10 mM DTT were added to the samples and incubated in RT for 10 min, to consume unreacted NEM. The samples then underwent methanol/chloroform precipitation and the pellets were resuspended in 2.5 µl 4% LDS per *X. tropicalis* embryo or 5 µl 4% LDS per *X. laevis* embryo.

The following steps of mass spectrometry sample preparation, data acquisition and data analysis were performed by Dr. Albert Fradera Sola, Institute of Molecular Biology. 30 µl *X. laevis* and *X. tropicalis* proteome extracts were heat denatured at 70°C for 10 min. Proteins were separated on a 10% NOVEX gradient SDS gel (Thermo Scientific) in 1X MES buffer (Thermo Scientific) at 180 V for 8 min. For visualization, proteins were stained with Coomassie solution. Each gel lane was cut in slices, minced and destained with protein destaining solution. Proteins were reduced in 10 mM DTT at 56°C for 1 h and then alkylated with 50 mM iodoacetamide in RT for 45 min, in the dark. Proteins were digested with mass spectrometry grade trypsin (Sigma-Aldrich) at 37°C O/N. Peptides were gel extracted with two rounds of 30% acetonitrile, 50 mM ammonium bicarbonate pH 8 mix, followed by three rounds of 100% acetonitrile, which was subsequently evaporated using a centrifuge concentrator (Eppendorf). Samples were then processed with Empore C18 StageTips (CDS Analytical) for desalting and fractionation, as previously described (Rappsilber, Mann, and Ishihama 2007).

Quantitative peptide analysis was performed on EASY-nLC 1200 System (Thermo Scientific) coupled to Orbitrap Exploris 480 Plus mass spectrometer (Thermo Scientific), using a self-packed column (25 cm × 75 µm inner diameter, New Objective) filled with ReproSil-Pur 120 C18-AQ (Dr. Maisch GmbH). LC running buffers were 0.1% formic acid (Buffer A) and 0.1% formic acid, 80% acetonitrile (Buffer B). Separations were performed at a flow rate of 250 nl/min, using the following

gradient: 0% of buffer B followed by 2-40% linear increase of buffer B in 88 minutes. The mass spectrometer was operated in positive ion mode with a data-dependent acquisition of one MS full scan (scan range 300 - 1,650 m/z; 60,000 resolution; normalized AGC target 300%; max IT 28 ms) and up to twenty MS/MS scans (15,000 resolution; AGC target 100%, max IT 40 ms; isolation window 1.4 m/z) with higher-energy collisional dissociation (HCD) fragmentation. Proteins were identified with Andromeda search engine (Cox et al., 2011) of MaxQuant suite v2.4.2.0 (Cox & Mann, 2008) using the Xenbase (Fisher et al., 2023) *X. laevis* (XENLA_10.1_Xenbase.pep.fa) and *X. tropicalis* (XENTR_10.0_Xenbase.pep.fa) reference proteomes. For the analysis, carbamidomethylation at cysteine was set as a fixed modification and methionine oxidation and protein N-acetylation were set as variable modifications. The 'match between runs' option was activated.

For protein quantification, contaminants, reverse database hits, protein groups identified only by site and protein groups with less than 2 peptides (with at least one classified as unique) were filtered out. Missing values were imputed by shifting a beta distribution derived from label-free quantification (LFQ) intensity values to quantitation limit. Further analysis and graphical representation were conducted within an R framework (R Core Team 2021) incorporating ggplot2 (Wickham, 2016) for visualization, among other packages. For the protein dynamicity analysis, protein abundance stability across embryonic developmental stages was evaluated using the Gini coefficient. For the pairwise comparison of stages, the protein enrichment threshold was set to a p-value < 0.05 (Welch's t-test) and fold change > 2, c = 0.05. The corresponding Welch's t-test was performed with 4 replicates per condition.

4.2.32. Fluorescence Polarization assay

5' FAM labelled and 3' biotin labelled oligos carrying two modified cytosines in CpG contex per strand (Table 15) were hybridized to prepare dsDNA. For oligo hybridization, equimolar ratio of forward and reverse oligo were mixed with FP buffer in final volume 20 µl and annealed by heating to 95°C for 30 sec and then cool down to 4°C with cooling rate 0.1°C/sec. 10 nM dsDNA oligo were mixed with 5 nM - 1 µM mCRY1 or hCRY2 protein in FP buffer in final volume 10 µl. Salt concentration was at near physiological conditions (100 mM NaCl). Each sample was prepared in technical triplicates and plated in 384-well clear bottom plates (Corning). Fluorescence intensity was measured using a Tecan Spark M20 microplate reader and analyzed with SparkControl software v.3.2. Dissociation constants (K_D) were calculated with the Hill equation in GraphPad Prism v.9.

4.2.33. ImageJ analysis

Image analysis was performed with ImageJ 64 v5. The blastopore perimeter (Figure 2.5E) was calculated using the measure tool in ImageJ. Northern blot quantification (Figure 2.10B-D) was performed with the Analyze Gels tool in ImageJ.

4.2.34. Statistical analysis

Data throughout this dissertation 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 v.9 and displayed as *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and ****$p \leq 0.0001$.

5. References

Abe, K.-I., Funaya, S., Tsukioka, D., Kawamura, M., Suzuki, Y., Suzuki, M. G., Schultz, R. M., & 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*(29), E6780–E6788. https://doi.org/10.1073/pnas.1804309115

Adamson, E. D., & Woodland, H. R. (1974). Histone synthesis in early amphibian development: histone and DNA syntheses are not co-ordinated. *Journal of Molecular Biology*, *88*(2), 263–285. https://doi.org/10.1016/0022-2836(74)90481-1

Adams, R. L. P. (1971). The relationship between synthesis and methylation of DNA in mouse fibroblasts. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, *254*(2), 205–212. https://doi.org/10.1016/0005-2787(71)90829-X

Adams, R. L. P. (1973). Delayed Methylation of DNA in Developing Sea Urchin Embryos. *Nature New Biology*, *244*(131), 27–29. https://doi.org/10.1038/newbio244027a0

Akahori, H., Guindon, S., Yoshizaki, S., & Muto, Y. (2015). Molecular Evolution of the TET Gene Family in Mammals. *International Journal of Molecular Sciences*, *16*(12), 28472–28485. https://doi.org/10.3390/ijms161226110

Akkers, R. C., van Heeringen, S. J., Jacobi, U. G., Janssen-Megens, E. M., Françoijs, K.-J., Stunnenberg, H. G., & Veenstra, G. J. C. (2009). A Hierarchy of H3K4me3 and H3K27me3 Acquisition in Spatial Gene Regulation in Xenopus Embryos. *Developmental Cell*, *17*(3), 425–434. https://doi.org/10.1016/j.devcel.2009.08.005

Almeida, R. D., Loose, M., Sottile, V., Matsa, E., Denning, C., Young, L., Johnson, A. D., Gering, M., & Ruzov, A. (2012). 5-hydroxymethyl-cytosine enrichment of non-committed cells is not a universal feature of vertebrate development. *Epigenetics*, *7*(4), 383–389. https://doi.org/10.4161/epi.19375

Almouzni, G., & Wolffe, A. P. (1995). Constraints on transcriptional activator function contribute to transcriptional quiescence during early Xenopus embryogenesis. *The EMBO Journal*, *14*(8), 1752–1765. https://doi.org/10.1002/j.1460-2075.1995.tb07164.x

Amaldi, F., & M Buongiorno-Nardelli, M. (1971). Molecular hybridization of Chinese hamster 5S, 4S and “pulse-labelled” RNA in cytological preparations. *Experimental Cell Research*, *65*(2), 329–334. https://doi.org/10.1016/0014-4827(71)90009-7

Amaya, E., Offield, M. F., & Grainger, R. M. (1998). Frog genetics: Xenopus tropicalis jumps into the future. *Trends in Genetics*, *14*(7), 253–255. https://doi.org/10.1016/S0168-9525(98)01506-6

Ambrosi, C., Manzo, M., & Baubec, T. (2017). Dynamics and Context-Dependent Roles of DNA Methylation. *Journal of Molecular Biology*, *429*(10), 1459–1475. https://doi.org/10.1016/j.jmb.2017.02.008

Amodeo, A. A., Jukam, D., Straight, A. F., & Skotheim, J. M. (2015). Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition. *Proceedings of the National Academy of Sciences*, *112*(10). https://doi.org/10.1073/pnas.1413990112

Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P. W. S., D'Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E., Encheva, V., Kudo, N. R., Koseki, H., Sasaki, H., & Hajkova, P. (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nature Cell Biology*, *18*(2), 225–233. https://doi.org/10.1038/ncb3296

Ananiev, E. V., Barsky, V. E., Ilyin, Y. V., & Churikov, N. A. (1981). Localization of nucleoli in Drosophila melanogaster polytene chromosomes. *Chromosoma*, *81*(4), 619–628. https://doi.org/10.1007/BF00285853

Andersen, I. S., Reiner, A. H., Aanes, H., Aleström, P., & Collas, P. (2012). Developmental features of DNA methylation during activation of the embryonic zebrafish genome. *Genome Biology*, *13*(7), R65. https://doi.org/10.1186/gb-2012-13-7-r65

Anderson, B. J., Larkin, C., Guja, K., & Schildbach, J. F. (2008). *Using Fluorophore-Labeled Oligonucleotides to Measure Affinities of Protein–DNA Interactions* (pp. 253–272). https://doi.org/10.1016/S0076-6879(08)03412-5

Anderson, K. V., & Lengyel, J. A. (1980). Changing rates of histone mRNA synthesis and turnover in Drosophila embryos. *Cell*, *21*(3), 717–727. https://doi.org/10.1016/0092-8674(80)90435-3

Andrews, M. T., Loo, S., & Wilson, L. R. (1991). Coordinate inactivation of class III genes during the gastrula-neurula transition in Xenopus. *Developmental Biology*, *146*(1), 250–254. https://doi.org/10.1016/0012-1606(91)90466-G

Aoki, A., Suetake, I., Miyagawa, J., Fujio, T., Chijiwa, T., Sasaki, H., & Tajima, S. (2001). Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases. *Nucleic Acids Research*, *29*(17), 3506–3512. https://doi.org/10.1093/nar/29.17.3506

Aoki, F., Worrad, D. M., & Schultz, R. M. (1997). Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Developmental Biology*, *181*(2), 296–307. https://doi.org/10.1006/dbio.1996.8466

Arand, J., Chiang, H. R., Martin, D., Snyder, M. P., Sage, J., Reijo Pera, R. A., & Wossidlo, M. (2022). Tet enzymes are essential for early embryogenesis and completion of embryonic genome activation. *EMBO Reports*, *23*(2). https://doi.org/10.15252/embr.202153968

Aran, D., Toperoff, G., Rosenberg, M., & Hellman, A. (2011). Replication timing-related and gene body-specific methylation of active human genes. *Human Molecular Genetics*, *20*(4), 670–680. https://doi.org/10.1093/hmg/ddq513

Arber, W. (1965). Host-Controlled Modification of Bacteriophage. *Annual Review of Microbiology*, *19*(1), 365–378. https://doi.org/10.1146/annurev.mi.19.100165.002053

Arimbasseri, A. G., & Maraia, R. J. (2015). *Biochemical Analysis of Transcription Termination by RNA Polymerase III from Yeast Saccharomyces cerevisiae* (pp. 185–198). https://doi.org/10.1007/978-1-4939-2392-2_10

Arimbasseri, A. G., Rijal, K., & Maraia, R. J. (2013). Transcription termination by the eukaryotic RNA polymerase III. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1829*(3–4), 318–330. https://doi.org/10.1016/j.bbagrm.2012.10.006

Artee Luchman, H., Paterno, G. D., Kao, K. R., & Gillespie, L. L. (1999). Differential nuclear localization of ER1 protein during embryonic development in Xenopus laevis. *Mechanisms of Development*, *80*(1), 111–114. https://doi.org/10.1016/S0925-4773(98)00206-8

Asami, M., Lam, B. Y. H., Ma, M. K., Rainbow, K., Braun, S., VerMilyea, M. D., Yeo, G. S. H., & Perry, A. C. F. (2022). Human embryonic genome activation initiates at the one-cell stage. *Cell Stem Cell*, *29*(2), 209-216.e4. https://doi.org/10.1016/j.stem.2021.11.012

Bachman, M., Uribe-Lewis, S., Yang, X., Burgess, H. E., Iurlaro, M., Reik, W., Murrell, A., & Balasubramanian, S. (2015). 5-Formylcytosine can be a stable DNA modification in mammals. *Nature Chemical Biology*, *11*(8), 555–557. https://doi.org/10.1038/nchembio.1848

Bachvabova, R., & Davidson, E. H. (1966). Nuclear activation at the onset of amphibian gastrulation. *Journal of Experimental Zoology*, *163*(3), 285–295. https://doi.org/10.1002/jez.1401630307

Bai, L., Yang, G., Qin, Z., Lyu, J., Wang, Y., Feng, J., Liu, M., Gong, T., Li, X., Li, Z., Li, J., Qin, J., Yang, W., & Ding, C. (2021). Proteome-Wide Profiling of Readers for DNA Modification. *Advanced Science* , *8*(19), e2101426. https://doi.org/10.1002/advs.202101426

Baker, R. E., Camier, S., Sentenac, A., & Hall, B. D. (1987). Gene size differentially affects the binding of yeast transcription factor tau to two intragenic regions. *Proceedings of the National Academy of Sciences*, *84*(24), 8768–8772. https://doi.org/10.1073/pnas.84.24.8768

Ball, M. P., Li, J. B., Gao, Y., Lee, J.-H., LeProust, E. M., Park, I.-H., Xie, B., Daley, G. Q., & Church, G. M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nature Biotechnology*, *27*(4), 361–368. https://doi.org/10.1038/nbt.1533

Barski, A., Chepelev, I., Liko, D., Cuddapah, S., Fleming, A. B., Birch, J., Cui, K., White, R. J., & Zhao, K. (2010). Pol II and its associated epigenetic marks are present at Pol III–transcribed noncoding RNA genes. *Nature Structural & Molecular Biology*, *17*(5), 629–634. https://doi.org/10.1038/nsmb.1806

Basta, J., & Rauchman, M. (2015). The nucleosome remodeling and deacetylase complex in development and disease. *Translational Research*, *165*(1), 36–47. https://doi.org/10.1016/j.trsl.2014.05.003

Bazzini, A. A., Viso, F., Moreno-Mateos, M. A., Johnstone, T. G., Vejnar, C. E., Qin, Y., Yao, J., Khokha, M. K., & Giraldez, A. J. (2016). Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *The EMBO Journal*, *35*(19), 2087–2103. https://doi.org/10.15252/embj.201694699

Bestor, T., Laudano, A., Mattaliano, R., & Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *Journal of Molecular Biology*, *203*(4), 971–983. https://doi.org/10.1016/0022-2836(88)90122-2

Bird, A., Taggart, M., Frommer, M., Miller, O. J., & Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, *40*(1), 91–99. https://doi.org/10.1016/0092-8674(85)90312-5

Bisbee, C. A., Baker, M. A., Wilson, A. C., Haji-Azimi, I., & Fischberg, M. (1977). Albumin Phylogeny for Clawed Frogs (*Xenopus*). *Science*, *195*(4280), 785–787. https://doi.org/10.1126/science.65013

Blitz, I. L., & Cho, K. W. Y. (2021). Control of zygotic genome activation in Xenopus. In Sergei Y. Sokol (Ed.), *Amphibian Models of Development and Disease* (pp. 167–204). https://doi.org/10.1016/bs.ctdb.2021.03.003

Bloom-Ackermann, Z., Navon, S., Gingold, H., Towers, R., Pilpel, Y., & Dahan, O. (2014). A Comprehensive tRNA Deletion Library Unravels the Genetic Architecture of the tRNA Pool. *PLoS Genetics*, *10*(1), e1004084. https://doi.org/10.1371/journal.pgen.1004084

Blythe, M. J., Kocer, A., Rubio-Roldan, A., Giles, T., Abakir, A., Ialy-Radio, C., Wheldon, L. M., Bereshchenko, O., Bruscoli, S., Kondrashov, A., Drevet, J. R., Emes, R. D., Johnson, A. D., McCarrey, J. R., Gackowski, D., Olinski, R., Cocquet, J., Garcia-Perez, J. L., & Ruzov, A. (2021). LINE-1 transcription in round spermatids is associated with accretion of 5-carboxylcytosine in their open reading frames. *Communications Biology*, *4*(1), 691. https://doi.org/10.1038/s42003-021-02217-8

Boettiger, A. N., & Levine, M. (2009). Synchronous and Stochastic Patterns of Gene Activation in the Drosophila Embryo. *Science*, *325*(5939), 471–473. https://doi.org/10.1126/science.1173976

Bogdanović, O., Long, S. W., van Heeringen, S. J., Brinkman, A. B., Gómez-Skarmeta, J. L., Stunnenberg, H. G., Jones, P. L., & Veenstra, G. J. C. (2011). Temporal uncoupling of the DNA methylome and transcriptional repression during embryogenesis. *Genome Research*, *21*(8), 1313–1327. https://doi.org/10.1101/gr.114843.110

Bogdanović, O., Smits, A. H., de la Calle Mustienes, E., Tena, J. J., Ford, E., Williams, R., Senanayake, U., Schultz, M. D., Hontelez, S., van Kruijsbergen, I., Rayon, T., Gnerlich, F., Carell, T., Veenstra, G. J. C., Manzanares, M., Sauka-Spengler, T., Ecker, J. R., Vermeulen, M., Gómez-Skarmeta, J. L., & Lister, R. (2016). Active DNA demethylation at enhancers during the vertebrate phylotypic period. *Nature Genetics*, *48*(4), 417–426. https://doi.org/10.1038/ng.3522

Borst, P., & Sabatini, R. (2008). Base J: discovery, biosynthesis, and possible functions. *Annual Review of Microbiology*, *62*, 235–251. https://doi.org/10.1146/annurev.micro.62.081307.162750

Bošković, A., Eid, A., Pontabry, J., Ishiuchi, T., Spiegelhalter, C., Raghu Ram, E. V. S., Meshorer, E., & Torres-Padilla, M.-E. (2014). Higher chromatin mobility supports totipotency and precedes pluripotency in vivo. *Genes & Development*, *28*(10), 1042–1047. https://doi.org/10.1101/gad.238881.114

Bostick, M., Kim, J. K., Estève, P.-O., Clark, A., Pradhan, S., & Jacobsen, S. E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science*, *317*(5845), 1760–1764. https://doi.org/10.1126/science.1147939

Bouniol, C., Nguyen, E., & Debey, P. (1995). Endogenous transcription occurs at the 1-cell stage in the mouse embryo. *Experimental Cell Research*, *218*(1), 57–62. https://doi.org/10.1006/excr.1995.1130

Bourc'his, D., & Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, *431*(7004), 96–99. https://doi.org/10.1038/nature02886

Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., & Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science*, *294*(5551), 2536–2539. https://doi.org/10.1126/science.1065848

Boyes, J., & Bird, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell*, *64*(6), 1123–1134. https://doi.org/10.1016/0092-8674(91)90267-3

Braglia, P., Percudani, R., & Dieci, G. (2005). Sequence Context Effects on Oligo (dT) Termination Signal Recognition by Saccharomyces cerevisiae RNA Polymerase III. *Journal of Biological Chemistry*, *280*(20), 19551–19562. https://doi.org/10.1074/jbc.M412238200

Brahma, S., & Henikoff, S. (2024). The BAF chromatin remodeler synergizes with RNA polymerase II and transcription factors to evict nucleosomes. *Nature Genetics*, *56*(1), 100–111. https://doi.org/10.1038/s41588-023-01603-8

Braude, P., Bolton, V., & Moore, S. (1988). Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature*, *332*(6163), 459–461. https://doi.org/10.1038/332459a0

Brewer, A. C., Guille, M. J., Fear, D. J., Partington, G. A., & Patient, R. K. (1995). Nuclear translocation of a maternal CCAAT factor at the start of gastrulation activates Xenopus GATA-2 transcription. *The EMBO Journal*, *14*(4), 757–766. https://doi.org/10.1002/j.1460-2075.1995.tb07054.x

Bright, A. R., van Genesen, S., Li, Q., Grasso, A., Frölich, S., van der Sande, M., van Heeringen, S. J., & Veenstra, G. J. C. (2021). Combinatorial transcription factor activities on open chromatin induce embryonic heterogeneity in vertebrates. *The EMBO Journal*, *40*(9). https://doi.org/10.15252/embj.2020104913

Brown, D. D., & Gurdon, J. B. (1964). ABSENCE OF RIBOSOMAL RNA SYNTHESIS IN THE ANUCLEOLATE MUTANT OF XENOPUS LAEVIS. *Proceedings of the National Academy of Sciences*, *51*(1), 139–146. https://doi.org/10.1073/pnas.51.1.139

Bultman, S. J., Gebuhr, T. C., Pan, H., Svoboda, P., Schultz, R. M., & Magnuson, T. (2006). Maternal BRG1 regulates zygotic genome activation in the mouse. *Genes & Development*, *20*(13), 1744–1754. https://doi.org/10.1101/gad.1435106

Camier, S., Dechampesme, A. M., & Sentenac, A. (1995). The only essential function of TFIIIA in yeast is the transcription of 5S rRNA genes. *Proceedings of the National Academy of Sciences*, *92*(20), 9338–9342. https://doi.org/10.1073/pnas.92.20.9338

Cannatella, D. C., & de Sa, R. O. (1993). Xenopus Laevis as a Model Organism. *Systematic Biology*, *42*(4), 476–507. https://doi.org/10.1093/sysbio/42.4.476

Chambers, C., Cermakova, K., Chan, Y. S., Kurtz, K., Wohlan, K., Lewis, A. H., Wang, C., Pham, A., Dejmek, M., Sala, M., Loeza Cabrera, M., Aguilar, R., Nencka, R., Lacorazza, H. D., Rau, R. E., & Hodges, H. C. (2023). SWI/SNF Blockade Disrupts PU.1-Directed Enhancer Programs in Normal Hematopoietic Cells and Acute Myeloid Leukemia. *Cancer Research*, *83*(7), 983–996. https://doi.org/10.1158/0008-5472.CAN-22-2129

Chang, Y., Sun, L., Kokura, K., Horton, J. R., Fukuda, M., Espejo, A., Izumi, V., Koomen, J. M., Bedford, M. T., Zhang, X., Shinkai, Y., Fang, J., & Cheng, X. (2011). MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferase GLP/G9a. *Nature Communications*, *2*(1), 533. https://doi.org/10.1038/ncomms1549

Chen, H., Einstein, L. C., Little, S. C., & Good, M. C. (2019). Spatiotemporal Patterning of Zygotic Genome Activation in a Model Vertebrate Embryo. *Developmental Cell*, *49*(6), 852-866.e7. https://doi.org/10.1016/j.devcel.2019.05.036

Ciesla, M., Skowronek, E., & Boguta, M. (2018). Function of TFIIIC, RNA polymerase III initiation factor, in activation and repression of tRNA gene transcription. *Nucleic Acids Research*, *46*(18), 9444–9455. https://doi.org/10.1093/nar/gky656

Clarkson, S. G., Birnstiel, M. L., & Purdom, I. F. (1973). Clustering of transfer RNA genes of Xenopus laevis. *Journal of Molecular Biology*, *79*(2), 411–429. https://doi.org/10.1016/0022-2836(73)90014-4

Clarkson, S. G., & Kurer, V. (1976). Isolation and some properties of DNA coding for tRNA1met from Xenopus laevis. *Cell*, *8*(2), 183–195. https://doi.org/10.1016/0092-8674(76)90002-7

Cliffe, L. J., Kieft, R., Southern, T., Birkeland, S. R., Marshall, M., Sweeney, K., & Sabatini, R. (2009). JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Research*, *37*(5), 1452–1462. https://doi.org/10.1093/nar/gkn1067

Cloix, C., Tutois, S., Mathieu, O., Cuvillier, C., Espagnol, M. C., Picard, G., & Tourmente, S. (2000). Analysis of 5S rDNA Arrays in Arabidopsis thaliana: Physical Mapping and Chromosome-Specific Polymorphisms. *Genome Research*, *10*(5), 679–690. https://doi.org/10.1101/gr.10.5.679

Collart, C., Owens, N. D. L., Bhaw-Rosun, L., Cooper, B., De Domenico, E., Patrushev, I., Sesay, A. K., Smith, J. N., Smith, J. C., & Gilchrist, M. J. (2014). High-resolution analysis of gene activity during the Xenopus mid-blastula transition. *Development* , *141*(9), 1927–1939. https://doi.org/10.1242/dev.102012

Comings, D. E. (1972). Methylation of euchromatic and heterochromatic DNA. *Experimental Cell Research*, *74*(2), 383–390. https://doi.org/10.1016/0014-4827(72)90391-6

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., Abramowitz, L. K., Bartolomei, M. S., Rambow, F., Bassi, M. R., Bruno, T., Fanciulli, M., Renner, C., Klein-Szanto, A. J., Matsumoto, Y., … Bellacosa, A. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*, *146*(1), 67–79. https://doi.org/10.1016/j.cell.2011.06.020

Costa, Y., Ding, J., Theunissen, T. W., Faiola, F., Hore, T. A., Shliaha, P. V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., Das, S., Levasseur, D. N., Li, Z., Xu, M., Reik, W., Silva, J. C. R., & Wang, J. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature*, *495*(7441), 370–374. https://doi.org/10.1038/nature11925

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, *26*(12), 1367–1372. https://doi.org/10.1038/nbt.1511

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., & Mann, M. (2011). Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *Journal of Proteome Research*, *10*(4), 1794–1805. https://doi.org/10.1021/pr101065j

Cozzarelli, N. R., Gerrard, S. P., Schlissel, M., Brown, D. D., & Bogenhagen, D. F. (1983). Purified RNA polymerase III accurately and efficiently terminates transcription of 5S RNA genes. *Cell*, *34*(3), 829–835. https://doi.org/10.1016/0092-8674(83)90540-8

Crawford, D. J., Liu, M. Y., Nabel, C. S., Cao, X.-J., Garcia, B. A., & Kohli, R. M. (2016). Tet2 Catalyzes Stepwise 5-Methylcytosine Oxidation by an Iterative and de novo Mechanism. *Journal of the American Chemical Society*, *138*(3), 730–733. https://doi.org/10.1021/jacs.5b10554

Dai, H. Q., Wang, B. A., Yang, L., Chen, J. J., Zhu, G. C., Sun, M. L., Ge, H., Wang, R., Chapman, D. L., Tang, F., Sun, X., & Xu, G. L. (2016). TET-mediated DNA demethylation controls gastrulation by regulating Lefty-Nodal signalling. *Nature*, *538*(7626), 528–532. https://doi.org/10.1038/nature20095

Damal Villivalam, S., You, D., Kim, J., Lim, H. W., Xiao, H., Zushin, P.-J. H., Oguri, Y., Amin, P., & Kang, S. (2020). TET1 is a beige adipocyte-selective epigenetic suppressor of thermogenesis. *Nature Communications*, *11*(1), 4313. https://doi.org/10.1038/s41467-020-18054-y

Daniels, L. M., & Delany, M. E. (2003). Molecular and cytogenetic organization of the 5S ribosomal DNA array in chicken (Gallus gallus). *Chromosome Research*, *11*(4), 305–317. https://doi.org/10.1023/A:1024008522122

Davies, M. N., Volta, M., Pidsley, R., Lunnon, K., Dixit, A., Lovestone, S., Coarfa, C., Harris, R. A., Milosavljevic, A., Troakes, C., Al-Sarraj, S., Dobson, R., Schalkwyk, L. C., & Mill, J. (2012). Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biology*, *13*(6), R43. https://doi.org/10.1186/gb-2012-13-6-r43

Dawlaty, M. M., Breiling, A., Le, T., Barrasa, M. I., Raddatz, G., Gao, Q., Powell, B. E., Cheng, A. W., Faull, K. F., Lyko, F., & Jaenisch, R. (2014). Loss of Tet Enzymes Compromises Proper Differentiation of Embryonic Stem Cells. *Developmental Cell*, *29*(1), 102–111. https://doi.org/10.1016/j.devcel.2014.03.003

Delatte, B., Wang, F., Ngoc, L. V., Collignon, E., Bonvin, E., Deplus, R., Calonne, E., Hassabi, B., Putmans, P., Awe, S., Wetzel, C., Kreher, J., Soin, R., Creppe, C., Limbach, P. A., Gueydan, C., Kruys, V., Brehm, A., Minakhina, S., … Fuks, F. (2016). RNA biochemistry.

Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science*, *351*(6270), 282–285. https://doi.org/10.1126/science.aac5253

Deng, M., Zhang, G., Cai, Y., Liu, Z., Zhang, Y., Meng, F., Wang, F., & Wan, Y. (2020). DNA methylation dynamics during zygotic genome activation in goat. *Theriogenology*, *156*, 144–154. https://doi.org/10.1016/j.theriogenology.2020.07.008

Deniz, Ö., Frost, J. M., & Branco, M. R. (2019). Regulation of transposable elements by DNA modifications. *Nature Reviews Genetics*, *20*(7), 417–431. https://doi.org/10.1038/s41576-019-0106-6

Deplus, R., Delatte, B., Schwinn, M. K., Defrance, M., Méndez, J., Murphy, N., Dawson, M. A., Volkmar, M., Putmans, P., Calonne, E., Shih, A. H., Levine, R. L., Bernard, O., Mercher, T., Solary, E., Urh, M., Daniels, D. L., & Fuks, F. (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *The EMBO Journal*, *32*(5), 645–655. https://doi.org/10.1038/emboj.2012.357

Deprez, E., Arrebola, R., Conesa, C., & Sentenac, A. (1999). A Subunit of Yeast TFIIIC Participates in the Recruitment of TATA-Binding Protein. *Molecular and Cellular Biology*, *19*(12), 8042–8051. https://doi.org/10.1128/MCB.19.12.8042

De Renzis, S., Elemento, O., Tavazoie, S., & Wieschaus, E. F. (2007). Unmasking Activation of the Zygotic Genome Using Chromosomal Deletions in the Drosophila Embryo. *PLoS Biology*, *5*(5), e117. https://doi.org/10.1371/journal.pbio.0050117

Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M., & Pagano, A. (2007). The expanding RNA polymerase III transcriptome. *Trends in Genetics*, *23*(12), 614–622. https://doi.org/10.1016/j.tig.2007.09.001

Dimitrov, S., Almouzni, G., Dasso, M., & Wolffe, A. P. (1993). Chromatin transitions during early Xenopus embryogenesis: changes in histone H4 acetylation and in linker histone type. *Developmental Biology*, *160*(1), 214–227. https://doi.org/10.1006/dbio.1993.1299

Ding, J., Huang, X., Shao, N., Zhou, H., Lee, D.-F., Faiola, F., Fidalgo, M., Guallar, D., Saunders, A., Shliaha, P. V., Wang, H., Waghray, A., Papatsenko, D., Sánchez-Priego, C., Li, D., Yuan, Y., Lemischka, I. R., Shen, L., Kelley, K., … Wang, J. (2015). Tex10 Coordinates Epigenetic Control of Super-Enhancer Activity in Pluripotency and Reprogramming. *Cell Stem Cell*, *16*(6), 653–668. https://doi.org/10.1016/j.stem.2015.04.001

Dobson, A. T., Raja, R., Abeyta, M. J., Taylor, T., Shen, S., Haqq, C., & Pera, R. A. R. (2004). The unique transcriptome through day 3 of human preimplantation development. *Human Molecular Genetics*, *13*(14), 1461–1470. https://doi.org/10.1093/hmg/ddh157

Domcke, S., Bardet, A. F., Adrian Ginno, P., Hartl, D., Burger, L., & Schübeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature*, *528*(7583), 575–579. https://doi.org/10.1038/nature16462

Doskočil, J., & Šorm, F. (1962). Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. *Biochimica et Biophysica Acta*, *55*(6), 953–959. https://doi.org/10.1016/0006-3002(62)90909-5

Dreyer, C. (1987). Differential accumulation of oocyte nuclear proteins by embryonic nuclei of Xenopus. *Development* , *101*(4), 829–846. https://doi.org/10.1242/dev.101.4.829

Duan, J. E., Jiang, Z. C., Alqahtani, F., Mandoiu, I., Dong, H., Zheng, X., Marjani, S. L., Chen, J., & Tian, X. C. (2019). Methylome Dynamics of Bovine Gametes and in vivo Early Embryos. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.00512

Dubini, R. C. A., Schön, A., Müller, M., Carell, T., & Rovó, P. (2020). Impact of 5-formylcytosine on the melting kinetics of DNA by 1H NMR chemical exchange. *Nucleic Acids Research*, *48*(15), 8796–8807. https://doi.org/10.1093/nar/gkaa589

Du, J., Johnson, L. M., Jacobsen, S. E., & Patel, D. J. (2015). DNA methylation pathways and their crosstalk with histone methylation. *Nature Reviews Molecular Cell Biology*, *16*(9), 519–532. https://doi.org/10.1038/nrm4043

Engelke, D. R., Ng, S. Y., Shastry, B. S., & Roeder, R. G. (1980). Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. *Cell*, *19*(3), 717–728. https://doi.org/10.1016/S0092-8674(80)80048-1

Estève, P.-O., Chin, H. G., Smallwood, A., Feehery, G. R., Gangisetty, O., Karpf, A. R., Carey, M. F., & Pradhan, S. (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes & Development*, *20*(22), 3089–3103. https://doi.org/10.1101/gad.1463706

Ferguson-Smith, A. C., & Bourchis, D. (2018). The discovery and importance of genomic imprinting. *ELife*, *7*. https://doi.org/10.7554/eLife.42368

Ficz, G., Branco, M. R., Seisenberger, S., Santos, F., Krueger, F., Hore, T. A., Marques, C. J., Andrews, S., & Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*, *473*(7347), 398–402. https://doi.org/10.1038/nature10008

Ficz, G., Hore, T. A., Santos, F., Lee, H. J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.-L., Walter, J., Cook, S. J., Andrews, S., Branco, M. R., & Reik, W. (2013). FGF Signaling Inhibition in ESCs Drives Rapid Genome-wide Demethylation to the Epigenetic Ground State of Pluripotency. *Cell Stem Cell*, *13*(3), 351–359. https://doi.org/10.1016/j.stem.2013.06.004

Filion, G. J. P., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., & Defossez, P.-A. (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Molecular and Cellular Biology*, *26*(1), 169–181. https://doi.org/10.1128/MCB.26.1.169-181.2006

Fisher, M., James-Zorn, C., Ponferrada, V., Bell, A. J., Sundararaj, N., Segerdell, E., Chaturvedi, P., Bayyari, N., Chu, S., Pells, T., Lotay, V., Agalakov, S., Wang, D. Z., Arshinoff, B. I., Foley, S., Karimi, K., Vize, P. D., & Zorn, A. M. (2023). Xenbase: key features and resources of the Xenopus model organism knowledgebase. *Genetics*, *224*(1). https://doi.org/10.1093/genetics/iyad018

French, S. L., Osheim, Y. N., Schneider, D. A., Sikes, M. L., Fernandez, C. F., Copela, L. A., Misra, V. A., Nomura, M., Wolin, S. L., & Beyer, A. L. (2008). Visual Analysis of the Yeast

5S rRNA Gene Transcriptome: Regulation and Role of La Protein. *Molecular and Cellular Biology*, *28*(14), 4576–4587. https://doi.org/10.1128/MCB.00127-08

Fu, L., Guerrero, C. R., Zhong, N., Amato, N. J., Liu, Y., Liu, S., Cai, Q., Ji, D., Jin, S.-G., Niedernhofer, L. J., Pfeifer, G. P., Xu, G.-L., & Wang, Y. (2014). Tet-mediated formation of 5-hydroxymethylcytosine in RNA. *Journal of the American Chemical Society*, *136*(33), 11582–11585. https://doi.org/10.1021/ja505305z

Fu, T., Liu, L., Yang, Q.-L., Wang, Y., Xu, P., Zhang, L., Liu, S., Dai, Q., Ji, Q., Xu, G.-L., He, C., Luo, C., & Zhang, L. (2019). Thymine DNA glycosylase recognizes the geometry alteration of minor grooves induced by 5-formylcytosine and 5-carboxylcytosine. *Chemical Science*, *10*(31), 7407–7417. https://doi.org/10.1039/c9sc02807b

Gao, J., & Shen, W. (2021). Xenopus in revealing developmental toxicity and modeling human diseases. *Environmental Pollution*, *268*, 115809. https://doi.org/10.1016/j.envpol.2020.115809

Gao, Y., Li, L., Yuan, P., Zhai, F., Ren, Y., Yan, L., Li, R., Lian, Y., Zhu, X., Wu, X., Kee, K., Wen, L., Qiao, J., & Tang, F. (2020). 5-Formylcytosine landscapes of human preimplantation embryos at single-cell resolution. *PLoS Biology*, *18*(7), e3000799. https://doi.org/10.1371/journal.pbio.3000799

Gentsch, G. E., Owens, N. D. L., & Smith, J. C. (2019). The Spatiotemporal Control of Zygotic Genome Activation. *IScience*, *16*, 485–498. https://doi.org/10.1016/j.isci.2019.06.013

Gerber, A., Ito, K., Chu, C.-S., & Roeder, R. G. (2020). Gene-Specific Control of tRNA Expression by RNA Polymerase II. *Molecular Cell*, *78*(4), 765-778.e7. https://doi.org/10.1016/j.molcel.2020.03.023

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., & Carell, T. (2010). Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. *PLoS ONE*, *5*(12), e15367. https://doi.org/10.1371/journal.pone.0015367

Gold, M., Hurwitz, J., & Anders, M. (1963). THE ENZYMATIC METHYLATION OF RNA AND DNA, II. ON THE SPECIES SPECIFICITY OF THE METHYLATION ENZYMES. *Proceedings of the National Academy of Sciences*, *50*(1), 164–169. https://doi.org/10.1073/pnas.50.1.164

Golla, J. P., Zhao, J., Mann, I. K., Sayeed, S. K., Mandal, A., Rose, R. B., & Vinson, C. (2014). Carboxylation of cytosine (5caC) in the CG dinucleotide in the E-box motif (CGCAG|GTG) increases binding of the Tcf3|Ascl1 helix-loop-helix heterodimer 10-fold. *Biochemical and Biophysical Research Communications*, *449*(2), 248–255. https://doi.org/10.1016/j.bbrc.2014.05.018

Gowher, H., & Jeltsch, A. (2001). Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. *Journal of Molecular Biology*, *309*(5), 1201–1208. https://doi.org/10.1006/jmbi.2001.4710

Graf, J. D., & Kobel, H. R. (1991). Genetics of Xenopus laevis. *Methods in Cell Biology*, *36*, 19–34. https://doi.org/10.1016/s0091-679x(08)60270-8

Gruenbaum, Y., Naveh-Many, T., Cedar, H., & Razin, A. (1981). Sequence specificity of methylation in higher plant DNA. *Nature*, *292*(5826), 860–862. https://doi.org/10.1038/292860a0

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.-P., Hu, B., Walsh, C. P., Li, J., Tang, F., & Xu, G.-L. (2014). Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. *Cell Stem Cell*, *15*(4), 447–459. https://doi.org/10.1016/j.stem.2014.08.003

Gupta, M., Sonnett, M., Ryazanova, L., Presler, M., & Wühr, M. (2018). *Quantitative Proteomics of Xenopus Embryos I, Sample Preparation* (pp. 175–194). https://doi.org/10.1007/978-1-4939-8784-9_13

Gupta, R., Wills, A., Ucar, D., & Baker, J. (2014). Developmental enhancers are marked independently of zygotic Nodal signals in Xenopus. *Developmental Biology*, *395*(1), 38–49. https://doi.org/10.1016/j.ydbio.2014.08.034

Gurdon, J. B. (1962). The Developmental Capacity of Nuclei taken from Intestinal Epithelium Cells of Feeding Tadpoles. *Development*, *10*(4), 622–640. https://doi.org/10.1242/dev.10.4.622

Gurdon, J. B., & Hopwood, N. (2000). The introduction of Xenopus laevis into developmental biology: of empire, pregnancy testing and ribosomal genes. *The International Journal of Developmental Biology*, *44*(1), 43–50.

Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S., Iqbal, K., Shi, Y. G., Deng, Z., Szabó, P. E., Pfeifer, G. P., Li, J., & Xu, G.-L. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, *477*(7366), 606–610. https://doi.org/10.1038/nature10443

Guven-Ozkan, T., Nishi, Y., Robertson, S. M., & Lin, R. (2008). Global Transcriptional Repression in C. elegans Germline Precursors by Regulated Sequestration of TAF-4. *Cell*, *135*(1), 149–160. https://doi.org/10.1016/j.cell.2008.07.040

Haeusler, R. A., & Engelke, D. R. (2006). Spatial organization of transcription by RNA polymerase III. *Nucleic Acids Research*, *34*(17), 4826–4836. https://doi.org/10.1093/nar/gkl656

Hahn, S., & Roberts, S. (2000). The zinc ribbon domains of the general transcription factors TFIIB and Brf: conserved functional surfaces but different roles in transcription initiation. *Genes & Development*, *14*(6), 719–730. http://www.ncbi.nlm.nih.gov/pubmed/10733531

Hamatani, T., Carter, M. G., Sharov, A. A., & Ko, M. S. H. (2004). Dynamics of Global Gene Expression Changes during Mouse Preimplantation Development. *Developmental Cell*, *6*(1), 117–131. https://doi.org/10.1016/S1534-5807(03)00373-3

Han, D., Schomacher, L., Schüle, K. M., Mallick, M., Musheev, M. U., Karaulanov, E., Krebs, L., von Seggern, A., & Niehrs, C. (2019). NEIL1 and NEIL2 DNA glycosylases protect neural crest development against mitochondrial oxidative stress. *ELife*, *8*. https://doi.org/10.7554/eLife.49044

Hardeland, U., Bentele, M., Jiricny, J., & Schär, P. (2000). Separating Substrate Recognition from Base Hydrolysis in Human Thymine DNA Glycosylase by Mutational Analysis. *Journal of Biological Chemistry*, *275*(43), 33449–33456. https://doi.org/10.1074/jbc.M005095200

Hardwick, J. S., Ptchelkine, D., El-Sagheer, A. H., Tear, I., Singleton, D., Phillips, S. E. V, Lane, A. N., & Brown, T. (2017). 5-Formylcytosine does not change the global structure of DNA. *Nature Structural & Molecular Biology*, *24*(6), 544–552. https://doi.org/10.1038/nsmb.3411

Harland, R. M., & Grainger, R. M. (2011). Xenopus research: metamorphosed by genetics and genomics. *Trends in Genetics : TIG*, *27*(12), 507–515. https://doi.org/10.1016/j.tig.2011.08.003

Harper, M. E., Price, J., & Korn, L. J. (1983). Chromosomal mapping of Xenopus 5S genes: somatic-type versus oocyte-type. *Nucleic Acids Research*, *11*(8), 2313–2323. https://doi.org/10.1093/nar/11.8.2313

Hashimoto, H., Pais, J. E., Zhang, X., Saleh, L., Fu, Z.-Q., Dai, N., Corrêa, I. R., Zheng, Y., & Cheng, X. (2014). Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature*, *506*(7488), 391–395. https://doi.org/10.1038/nature12905

Hellman, A., & Chess, A. (2007). Gene Body-Specific Methylation on the Active X Chromosome. *Science*, *315*(5815), 1141–1143. https://doi.org/10.1126/science.1136352

Hellsten, U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N. H., Shu, S., Taher, L., Blitz, I. L., Blumberg, B., Dichmann, D. S., Dubchak, I., Amaya, E., Detter, J. C., Fletcher, R., Gerhard, D. S., Goodstein, D., … Rokhsar, D. S. (2010). The genome of the Western clawed frog Xenopus tropicalis. *Science*, *328*(5978), 633–636. https://doi.org/10.1126/science.1183670

Hendrich, B., & Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Molecular and Cellular Biology*, *18*(11), 6538–6547. https://doi.org/10.1128/MCB.18.11.6538

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.-X., Zhang, K., He, C., & Xu, G.-L. (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science*, *333*(6047), 1303–1307. https://doi.org/10.1126/science.1210944

Heyn, P., Kircher, M., Dahl, A., Kelso, J., Tomancak, P., Kalinka, A. T., & Neugebauer, K. M. (2014). The Earliest Transcribed Zygotic Genes Are Short, Newly Evolved, and Different across Species. *Cell Reports*, *6*(2), 285–292. https://doi.org/10.1016/j.celrep.2013.12.030

Highett, M. I., Beven, A. F., & Shaw, P. J. (1993). Localization of 5 S genes and transcripts in Pisum sativum nuclei. *Journal of Cell Science*, *105*(4), 1151–1158. https://doi.org/10.1242/jcs.105.4.1151

Hirsch, N., Zimmerman, L. B., & Grainger, R. M. (2002). Xenopus, the next generation:X. Tropicalis genetics and genomics. *Developmental Dynamics*, *225*(4), 422–433. https://doi.org/10.1002/dvdy.10178

Hoffmann, N. A., Jakobi, A. J., Moreno-Morcillo, M., Glatt, S., Kosinski, J., Hagen, W. J. H., Sachse, C., & Müller, C. W. (2015). Molecular structures of unbound and transcribing RNA polymerase III. *Nature*, *528*(7581), 231–236. https://doi.org/10.1038/nature16143

Honda, S., Shigematsu, M., Morichika, K., Telonis, A. G., & Kirino, Y. (2015). Four-leaf clover qRT-PCR: A convenient method for selective quantification of mature tRNA. *RNA Biology*, *12*(5), 501–508. https://doi.org/10.1080/15476286.2015.1031951

Hon, G. C., Song, C.-X., Du, T., Jin, F., Selvaraj, S., Lee, A. Y., Yen, C., Ye, Z., Mao, S.-Q., Wang, B.-A., Kuan, S., Edsall, L. E., Zhao, B. S., Xu, G.-L., He, C., & Ren, B. (2014). 5mC Oxidation by Tet2 Modulates Enhancer Activity and Timing of Transcriptome Reprogramming during Differentiation. *Molecular Cell*, *56*(2), 286–297. https://doi.org/10.1016/j.molcel.2014.08.026

Hontelez, S., van Kruijsbergen, I., Georgiou, G., van Heeringen, S. J., Bogdanovic, O., Lister, R., & Veenstra, G. J. C. (2015). Embryonic transcription is controlled by maternally defined chromatin state. *Nature Communications*, *6*(1), 10148. https://doi.org/10.1038/ncomms10148

Hopper, A. K., Pai, D. A., & Engelke, D. R. (2010). Cellular dynamics of tRNAs and their genes. *FEBS Letters*, *584*(2), 310–317. https://doi.org/10.1016/j.febslet.2009.11.053

Horvath, S., & Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, *19*(6), 371–384. https://doi.org/10.1038/s41576-018-0004-3

Hotchkiss, R. D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *The Journal of Biological Chemistry*, *175*(1), 315–332.

Howlett, S. K., & Reik, W. (1991). Methylation levels of maternal and paternal genomes during preimplantation development. *Development*, *113*(1), 119–127. https://doi.org/10.1242/dev.113.1.119

Hsieh, Y. J., Kundu, T. K., Wang, Z., Kovelman, R., & Roeder, R. G. (1999). The TFIIIC90 subunit of TFIIIC interacts with multiple components of the RNA polymerase III machinery and contains a histone-specific acetyltransferase activity. *Molecular and Cellular Biology*, *19*(11), 7697–7704. https://doi.org/10.1128/MCB.19.11.7697

Huang, S., Deerinck, T. J., Ellisman, M. H., & Spector, D. L. (1997). The Dynamic Organization of the Perinucleolar Compartment in the Cell Nucleus. *The Journal of Cell Biology*, *137*(5), 965–974. https://doi.org/10.1083/jcb.137.5.965

Huang, W., Lan, M.-D., Qi, C.-B., Zheng, S.-J., Wei, S.-Z., Yuan, B.-F., & Feng, Y.-Q. (2016). Formation and determination of the oxidation products of 5-methylcytosine in RNA. *Chemical Science*, *7*(8), 5495–5502. https://doi.org/10.1039/c6sc01589a

Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W. A., Kang, J., Zepeda-Martínez, J. A., Pape, U. J., Jacobsen, S. E., Peters, B., & Rao, A. (2014). Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences*, *111*(4), 1361–1366. https://doi.org/10.1073/pnas.1322921111

Huang, Y., Kim, J. K., Do, D. V., Lee, C., Penfold, C. A., Zylicz, J. J., Marioni, J. C., Hackett, J. A., & Surani, M. A. (2017). Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition. *ELife*, *6*. https://doi.org/10.7554/eLife.22345

Huber, S. M., van Delft, P., Mendil, L., Bachman, M., Smollett, K., Werner, F., Miska, E. A., & Balasubramanian, S. (2015). Formation and abundance of 5-hydroxymethylcytosine in RNA. *Chembiochem*, *16*(5), 752–755. https://doi.org/10.1002/cbic.201500013

Hug, C. B., Grimaldi, A. G., Kruse, K., & Vaquerizas, J. M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell*, *169*(2), 216-228.e19. https://doi.org/10.1016/j.cell.2017.03.024

Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y. G., Zhu, J., Wang, P., & Xu, Y. (2013). Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell*, *155*(7), 1545–1555. https://doi.org/10.1016/j.cell.2013.11.020

Hu, L., Lu, J., Cheng, J., Rao, Q., Li, Z., Hou, H., Lou, Z., Zhang, L., Li, W., Gong, W., Liu, M., Sun, C., Yin, X., Li, J., Tan, X., Wang, P., Wang, Y., Fang, D., Cui, Q., … Xu, Y. (2015). Structural insight into substrate preference for TET-mediated oxidation. *Nature*, *527*(7576), 118–122. https://doi.org/10.1038/nature15713

Hunt, T. (2002). NOBEL LECTURE: Protein Synthesis, Proteolysis, and Cell Cycle Transitions. *Bioscience Reports*, *22*(5–6), 465–486. https://doi.org/10.1023/A:1022077317801

Inoue, A., & Zhang, Y. (2011). Replication-Dependent Loss of 5-Hydroxymethylcytosine in Mouse Preimplantation Embryos. *Science*, *334*(6053), 194–194. https://doi.org/10.1126/science.1212483

Iqbal, K., Jin, S.-G., Pfeifer, G. P., & Szabó, P. E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proceedings of the National Academy of Sciences*, *108*(9), 3642–3647. https://doi.org/10.1073/pnas.1014033108

Ito, S., D'Alessio, A. C., Taranova, O. V, Hong, K., Sowers, L. C., & Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, *466*(7310), 1129–1133. https://doi.org/10.1038/nature09303

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., & Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* , *333*(6047), 1300–1303. https://doi.org/10.1126/science.1210597

Iurlaro, M., Ficz, G., Oxley, D., Raiber, E. A., Bachman, M., Booth, M. J., Andrews, S., Balasubramanian, S., & Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biology*, *14*(10). https://doi.org/10.1186/gb-2013-14-10-r119

Iurlaro, M., McInroy, G. R., Burgess, H. E., Dean, W., Raiber, E. A., Bachman, M., Beraldi, D., Balasubramanian, S., & Reik, W. (2016). In vivo genome-wide profiling reveals a tissue-specific role for 5-formylcytosine. *Genome Biology*, *17*(1). https://doi.org/10.1186/s13059-016-1001-5

Jaisal, M., Sannapureddi, R. K. R., Rana, A., & Sathyamoorthy, B. (2023). 5-Formylcytosine weakens the G–C pair and imparts local conformational fluctuations to DNA duplexes. *Physical Chemistry Chemical Physics*, *25*(1), 241–254. https://doi.org/10.1039/D2CP04837J

Jansz, N. (2019). DNA methylation dynamics at transposable elements in mammals. *Essays in Biochemistry*, *63*(6), 677–689. https://doi.org/10.1042/EBC20190039

Jevtić, P., & Levy, D. L. (2015). Nuclear size scaling during Xenopus early development contributes to midblastula transition timing. *Current Biology*, *25*(1), 45–52. https://doi.org/10.1016/j.cub.2014.10.051

Jevtić, P., & Levy, D. L. (2017). Both Nuclear Size and DNA Amount Contribute to Midblastula Transition Timing in Xenopus laevis. *Scientific Reports*, *7*(1), 7908. https://doi.org/10.1038/s41598-017-08243-z

Jiang, D., Wei, S., Chen, F., Zhang, Y., & Li, J. (2017). TET3-mediated DNA oxidation promotes ATR-dependent DNA damage response. *EMBO Reports*, *18*(5), 781–796. https://doi.org/10.15252/embr.201643179

Jiang, L., Zhang, J., Wang, J.-J., Wang, L., Zhang, L., Li, G., Yang, X., Ma, X., Sun, X., Cai, J., Zhang, J., Huang, X., Yu, M., Wang, X., Liu, F., Wu, C.-I., He, C., Zhang, B., Ci, W., & Liu, J. (2013). Sperm, but Not Oocyte, DNA Methylome Is Inherited by Zebrafish Early Embryos. *Cell*, *153*(4), 773–784. https://doi.org/10.1016/j.cell.2013.04.041

Ji, S., Shao, H., Han, Q., Seiler, C. L., & Tretyakova, N. Y. (2017). Reversible DNA–Protein Cross-Linking at Epigenetic DNA Marks. *Angewandte Chemie - International Edition*, *56*(45), 14130–14134. https://doi.org/10.1002/anie.201708286

Johnson, T. B., & Coghill, R. D. (1925). Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus1. *Journal of the American Chemical Society*, *47*(11), 2838–2844.

Joseph, S. R., Pálfy, M., Hilbert, L., Kumar, M., Karschau, J., Zaburdaev, V., Shevchenko, A., & Vastenhouw, N. L. (2017). Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. *ELife*, *6*. https://doi.org/10.7554/eLife.23326

Kafer, G. R., Li, X., Horii, T., Suetake, I., Tajima, S., Hatada, I., & Carlton, P. M. (2016). 5-Hydroxymethylcytosine Marks Sites of DNA Damage and Promotes Genome Stability. *Cell Reports*, *14*(6), 1283–1292. https://doi.org/10.1016/j.celrep.2016.01.035

Kafri, T., Ariel, M., Brandeis, M., Shemer, R., Urven, L., McCarrey, J., Cedar, H., & Razin, A. (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes & Development*, *6*(5), 705–714. https://doi.org/10.1101/gad.6.5.705

Kashiwagi, K., Kashiwagi, A., Kurabayashi, A., Hanada, H., Nakajima, K., Okada, M., Takase, M., & Yaoita, Y. (2010). Xenopus tropicalis: an ideal experimental animal in amphibia. *Experimental Animals*, *59*(4), 395–405. https://doi.org/10.1538/expanim.59.395

Kassavetis, G. A., Braun, B. R., Nguyen, L. H., & Peter Geiduschek, E. (1990). S. cerevisiae TFIIIB is the transcription initiation factor proper of RNA polymerase III, while TFIIIA and TFIIIC are assembly factors. *Cell*, *60*(2), 235–245. https://doi.org/10.1016/0092-8674(90)90739-2

Kassavetis, G. A., Kumar, A., Ramirez, E., & Geiduschek, E. P. (1998). Functional and Structural Organization of Brf, the TFIIB-Related Component of the RNA Polymerase III

Transcription Initiation Complex. *Molecular and Cellular Biology*, *18*(9), 5587–5599. https://doi.org/10.1128/MCB.18.9.5587

Kellinger, M. W., Song, C. X., Chong, J., Lu, X. Y., He, C., & Wang, D. (2012). 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nature Structural and Molecular Biology*, *19*(8), 831–833. https://doi.org/10.1038/nsmb.2346

Khrapunov, S., Warren, C., Cheng, H., Berko, E. R., Greally, J. M., & Brenowitz, M. (2014). Unusual Characteristics of the DNA Binding Domain of Epigenetic Regulatory Protein MeCP2 Determine Its Binding Specificity. *Biochemistry*, *53*(21), 3379–3391. https://doi.org/10.1021/bi500424z

Kim, J., Kollhoff, A., Bergmann, A., & Stubbs, L. (2003). Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3. *Human Molecular Genetics*, *12*(3), 233–245. https://doi.org/10.1093/hmg/ddg028

Kitsera, N., Allgayer, J., Parsa, E., Geier, N., Rossa, M., Carell, T., & 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*(19), 11033–11042. https://doi.org/10.1093/nar/gkx718

Klemm, S. L., Shipony, Z., & Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics*, *20*(4), 207–220. https://doi.org/10.1038/s41576-018-0089-8

Kloc, M. (2014). Polarity of Xenopus Oocytes and Early Embryos. In *Xenopus Development* (pp. 60–74). Wiley. https://doi.org/10.1002/9781118492833.ch4

Klutstein, M., Nejman, D., Greenfield, R., & Cedar, H. (2016). DNA Methylation in Cancer and Aging. *Cancer Research*, *76*(12), 3446–3450. https://doi.org/10.1158/0008-5472.CAN-15-3278

Knowland, J. S. (1970). Polyacrylamide gel electrophoresis of nucleic acids synthesised during the early development of Xenopus laevis Daudin. *Biochimica et Biophysica Acta*, *204*(2), 416–429. https://doi.org/10.1016/0005-2787(70)90162-0

Koh, K. P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C. A., Mostoslavsky, G., Lahesmaa, R., Orkin, S. H., Rodig, S. J., Daley, G. Q., & Rao, A. (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*, *8*(2), 200–213. https://doi.org/10.1016/j.stem.2011.01.008

Konovalov, K. A., Wang, W., Wang, G., Goonetilleke, E. C., Gao, X., Wang, D., & Huang, X. (2021). A comprehensive mechanism for 5-carboxylcytosine-induced transcriptional pausing revealed by Markov state models. *The Journal of Biological Chemistry*, *296*, 100735. https://doi.org/10.1016/j.jbc.2021.100735

Kopp, K., & Huang, S. (2005). Perinucleolar compartment and transformation. *Journal of Cellular Biochemistry*, *95*(2), 217–225. https://doi.org/10.1002/jcb.20403

Kor, S. D., Chowdhury, N., Keot, A. K., Yogendra, K., Chikkaputtaiah, C., & Sudhakar Reddy, P. (2023). RNA Pol III promoters—key players in precisely targeted plant genome editing. *Frontiers in Genetics*, *13*. https://doi.org/10.3389/fgene.2022.989199

Kramerov, D. A., & Vassetzky, N. S. (2011). Origin and evolution of SINEs in eukaryotic genomes. *Heredity*, *107*(6), 487–495. https://doi.org/10.1038/hdy.2011.43

Kress, H., Bechler, K., Swida, U., & Maletz, S. (2001). Evolution of 5S rRNA gene families in Drosophila. *Chromosome Research*, *9*(5), 403–415. https://doi.org/10.1023/A:1016787602583

Kriaucionis, S., & Heintz, N. (2009). The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science*, *324*(5929), 929–930. https://doi.org/10.1126/science.1169786

Kuo, M. T., Mandel, J. L., & Chambon, P. (1979). DNA methylation: correlation with DNase I sensitivity of chicken ovalbumin and conalbumin chromatin. *Nucleic Acids Research*, *7*(8), 2105–2113. https://doi.org/10.1093/nar/7.8.2105

Kwasnieski, J. C., Orr-Weaver, T. L., & Bartel, D. P. (2019). Early genome activation in Drosophila is extensive with an initial tendency for aborted transcripts and retained introns. *Genome Research*, *29*(7), 1188–1197. https://doi.org/10.1101/gr.242164.118

Kwon, H., Imbalzano, A. N., Khavari, P. A., Kingston, R. E., & Green, M. R. (1994). Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature*, *370*(6489), 477–481. https://doi.org/10.1038/370477a0

Lan, J., Rajan, N., Bizet, M., Penning, A., Singh, N. K., Guallar, D., Calonne, E., Li Greci, A., Bonvin, E., Deplus, R., Hsu, P. J., Nachtergaele, S., Ma, C., Song, R., Fuentes-Iglesias, A., Hassabi, B., Putmans, P., Mies, F., Menschaert, G., … Fuks, F. (2020). Functional role of Tet-mediated RNA hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature Communications*, *11*(1), 4956. https://doi.org/10.1038/s41467-020-18729-6

Laue, K., Rajshekar, S., Courtney, A. J., Lewis, Z. A., & Goll, M. G. (2019). The maternal to zygotic transition regulates genome-wide heterochromatin establishment in the zebrafish embryo. *Nature Communications*, *10*(1), 1551. https://doi.org/10.1038/s41467-019-09582-3

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C. T., Low, H. M., Kin Sung, K. W., Rigoutsos, I., Loring, J., & Wei, C.-L. (2010). Dynamic changes in the human methylome during differentiation. *Genome Research*, *20*(3), 320–331. https://doi.org/10.1101/gr.101907.109

Lee, H. J., Hore, T. A., & Reik, W. (2014). Reprogramming the Methylome: Erasing Memory and Creating Diversity. *Cell Stem Cell*, *14*(6), 710–719. https://doi.org/10.1016/j.stem.2014.05.008

Lee, M. T., Bonneau, A. R., Takacs, C. M., Bazzini, A. A., DiVito, K. R., Fleming, E. S., & Giraldez, A. J. (2013). Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature*, *503*(7476), 360–364. https://doi.org/10.1038/nature12632

Lehnertz, B., Ueda, Y., Derijck, A. A. H. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., & Peters, A. H. F. M. (2003). Suv39h-Mediated Histone H3

Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Current Biology*, *13*(14), 1192–1200. https://doi.org/10.1016/S0960-9822(03)00432-9

Lessing, D., & Lee, J. T. (2013). X Chromosome Inactivation and Epigenetic Responses to Cellular Reprogramming. *Annual Review of Genomics and Human Genetics*, *14*(1), 85–110. https://doi.org/10.1146/annurev-genom-091212-153530

Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F., & Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell*, *69*(6), 905–914. https://doi.org/10.1016/0092-8674(92)90610-o

Lewis, L. C., Lo, P. C. K., Foster, J. M., Dai, N., Corrêa, I. R., Durczak, P. M., Duncan, G., Ramsawhook, A., Aithal, G. P., Denning, C., Hannan, N. R. F., & Ruzov, A. (2017). Dynamics of 5-carboxylcytosine during hepatic differentiation: Potential general role for active demethylation by DNA repair in lineage specification. *Epigenetics*, *12*(4), 277–286. https://doi.org/10.1080/15592294.2017.1292189

Ličytė, J., Gibas, P., Skardžiūtė, K., Stankevičius, V., Rukšėnaitė, A., & Kriukienė, E. (2020). A Bisulfite-free Approach for Base-Resolution Analysis of Genomic 5-Carboxylcytosine. *Cell Reports*, *32*(11), 108155. https://doi.org/10.1016/j.celrep.2020.108155

Li, E., Beard, C., & Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature*, *366*(6453), 362–365. https://doi.org/10.1038/366362a0

Li, E., & Zhang, Y. (2014). DNA Methylation in Mammals. *Cold Spring Harbor Perspectives in Biology*, *6*(5), a019133–a019133. https://doi.org/10.1101/cshperspect.a019133

Li, F., Zhang, Y., Bai, J., Greenberg, M. M., Xi, Z., & Zhou, C. (2017). 5-Formylcytosine Yields DNA-Protein Cross-Links in Nucleosome Core Particles. *Journal of the American Chemical Society*, *139*(31), 10617–10620. https://doi.org/10.1021/jacs.7b05495

Li, H., Rauch, T., Chen, Z.-X., Szabó, P. E., Riggs, A. D., & Pfeifer, G. P. (2006). The Histone Methyltransferase SETDB1 and the DNA Methyltransferase DNMT3A Interact Directly and Localize to Promoters Silenced in Cancer Cells. *Journal of Biological Chemistry*, *281*(28), 19489–19500. https://doi.org/10.1074/jbc.M513249200

Lin, C., & Todo, T. (2005). The cryptochromes. *Genome Biology*, *6*(5), 220. https://doi.org/10.1186/gb-2005-6-5-220

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–322. https://doi.org/10.1038/nature08514

Liu, G., Wang, W., Hu, S., Wang, X., & Zhang, Y. (2018). Inherited DNA methylation primes the establishment of accessible chromatin during genome activation. *Genome Research*, *28*(7), 998–1007. https://doi.org/10.1101/gr.228833.117

Liu, Y., Hu, Z., Cheng, J., Siejka-Zielińska, P., Chen, J., Inoue, M., Ahmed, A. A., & Song, C.-X. (2021). Subtraction-free and bisulfite-free specific sequencing of 5-methylcytosine and its

oxidized derivatives at base resolution. *Nature Communications*, *12*(1), 618. https://doi.org/10.1038/s41467-021-20920-2

Lombard-Banek, C., Reddy, S., Moody, S. A., & Nemes, P. (2016). Label-free Quantification of Proteins in Single Embryonic Cells with Neural Fate in the Cleavage-Stage Frog (Xenopus laevis) Embryo using Capillary Electrophoresis Electrospray Ionization High-Resolution Mass Spectrometry (CE-ESI-HRMS). *Molecular & Cellular Proteomics* , *15*(8), 2756–2768. https://doi.org/10.1074/mcp.M115.057760

Lorch, Y., LaPointe, J. W., & Kornberg, R. D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell*, *49*(2), 203–210. https://doi.org/10.1016/0092-8674(87)90561-7

Lorsbach, R. B., Moore, J., Mathew, S., Raimondi, S. C., Mukatira, S. T., & Downing, J. R. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia*, *17*(3), 637–641. https://doi.org/10.1038/sj.leu.2402834

Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., & Zhang, Y. (2014). Role of Tet proteins in enhancer activity and telomere elongation. *Genes & Development*, *28*(19), 2103–2119. https://doi.org/10.1101/gad.248005.114

Lühnsdorf, B., Kitsera, N., Warken, D., Lingg, T., Epe, B., & Khobta, A. (2012). Generation of reporter plasmids containing defined base modifications in the DNA strand of choice. *Analytical Biochemistry*, *425*(1), 47–53. https://doi.org/10.1016/j.ab.2012.03.001

Lu, X., Han, D., Zhao, B. S., Song, C.-X., Zhang, L.-S., Doré, L. C., & He, C. (2015). Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. *Cell Research*, *25*(3), 386–389. https://doi.org/10.1038/cr.2015.5

Macleod, D., Clark, V. H., & Bird, A. (1999). Absence of genome-wide changes in DNA methylation during development of the zebrafish. *Nature Genetics*, *23*(2), 139–140. https://doi.org/10.1038/13767

Mahfoudhi, E., Talhaoui, I., Cabagnols, X., Della Valle, V., Secardin, L., Rameau, P., Bernard, O. A., Ishchenko, A. A., Abbes, S., Vainchenker, W., Saparbaev, M., & Plo, I. (2016). TET2-mediated 5-hydroxymethylcytosine induces genetic instability and mutagenesis. *DNA Repair*, *43*, 78–88. https://doi.org/10.1016/j.dnarep.2016.05.031

Maiti, A., & Drohat, A. C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. *Journal of Biological Chemistry*, *286*(41), 35334–35338. https://doi.org/10.1074/jbc.C111.284620

Maryon, E., & Carroll, D. (1989). Degradation of linear DNA by a strand-specific exonuclease activity in Xenopus laevis oocytes. *Molecular and Cellular Biology*, *9*(11), 4862–4871. https://doi.org/10.1128/MCB.9.11.4862

Marzouki, N., Camier, S., Ruet, A., Moenne, A., & Sentenac, A. (1986). Selective proteolysis defines two DNA binding domains in yeast transcription factor tau. *Nature*, *323*(6084), 176–178. https://doi.org/10.1038/323176a0

Matera, A. G., Frey, M. R., Margelot, K., & 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*(5), 1181–1193. https://doi.org/10.1083/jcb.129.5.1181

Mayer, W., Niveleau, A., Walter, J., Fundele, R., & Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature*, *403*(6769), 501–502. https://doi.org/10.1038/35000656

McGhee, J. D., & Ginder, G. D. (1979). Specific DNA methylation sites in the vicinity of the chicken β-globin genes. *Nature*, *280*(5721), 419–420. https://doi.org/10.1038/280419a0

Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L., & Bird, A. P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell*, *58*(3), 499–507. https://doi.org/10.1016/0092-8674(89)90430-3

Melamed, P., Yosefzon, Y., David, C., Tsukerman, A., & Pnueli, L. (2018). Tet Enzymes, Variants, and Differential Effects on Function. *Frontiers in Cell and Developmental Biology*, *6*, 22. https://doi.org/10.3389/fcell.2018.00022

Miller, M., Reddy, B. A., Kloc, M., Li, X. X., Dreyer, C., & Etkin, L. D. (1991). The nuclear-cytoplasmic distribution of the Xenopus nuclear factor, xnf7, coincides with its state of phosphorylation during early development. *Development* , *113*(2), 569–575. https://doi.org/10.1242/dev.113.2.569

Mishima, Y., & Tomari, Y. (2016). Codon Usage and 3′ UTR Length Determine Maternal mRNA Stability in Zebrafish. *Molecular Cell*, *61*(6), 874–885. https://doi.org/10.1016/j.molcel.2016.02.027

Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., Bibel, M., & Schübeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular Cell*, *30*(6), 755–766. https://doi.org/10.1016/j.molcel.2008.05.007

Moir, R. D., & Willis, I. M. (2013). Regulation of pol III transcription by nutrient and stress signaling pathways. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1829*(3–4), 361–375. https://doi.org/10.1016/j.bbagrm.2012.11.001

Monk, M., Boubelik, M., & Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development*, *99*(3), 371–382. https://doi.org/10.1242/dev.99.3.371

Müller, F., Clarkson, S. G., & Galas, D. J. (1987). Sequence of a 3.18 kb tandem repeat of Xenopus laevis DNA containing 8 tRNA genes. *Nucleic Acids Research*, *15*(17), 7191. https://doi.org/10.1093/nar/15.17.7191

Müller, N., Ponkkonen, E., Carell, T., & 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*(20), 11025. https://doi.org/10.3390/ijms222011025

Mutlu, B., Chen, H.-M., Moresco, J. J., Orelo, B. D., Yang, B., Gaspar, J. M., Keppler-Ross, S., Yates, J. R., Hall, D. H., Maine, E. M., & Mango, S. E. (2018). Regulated nuclear

accumulation of a histone methyltransferase times the onset of heterochromatin formation in C. elegans embryos. *Science Advances*, *4*(8). https://doi.org/10.1126/sciadv.aat6224

Nagae, G., Isagawa, T., Shiraki, N., Fujita, T., Yamamoto, S., Tsutsumi, S., Nonaka, A., Yoshiba, S., Matsusaka, K., Midorikawa, Y., Ishikawa, S., Soejima, H., Fukayama, M., Suemori, H., Nakatsuji, N., Kume, S., & Aburatani, H. (2011). Tissue-specific demethylation in CpG-poor promoters during cellular differentiation. *Human Molecular Genetics*, *20*(14), 2710–2721. https://doi.org/10.1093/hmg/ddr170

Nanan, K. K., Sturgill, D. M., Prigge, M. F., Thenoz, M., Dillman, A. A., Mandler, M. D., & Oberdoerffer, S. (2019). TET-Catalyzed 5-Carboxylcytosine Promotes CTCF Binding to Suboptimal Sequences Genome-wide. *IScience*, *19*, 326–339. https://doi.org/10.1016/j.isci.2019.07.041

Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., & Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, *393*(6683), 386–389. https://doi.org/10.1038/30764

Nenni, M. J., Fisher, M. E., James-Zorn, C., Pells, T. J., Ponferrada, V., Chu, S., Fortriede, J. D., Burns, K. A., Wang, Y., Lotay, V. S., Wang, D. Z., Segerdell, E., Chaturvedi, P., Karimi, K., Vize, P. D., & Zorn, A. M. (2019). Xenbase: Facilitating the Use of Xenopus to Model Human Disease. *Frontiers in Physiology*, *10*. https://doi.org/10.3389/fphys.2019.00154

Neri, F., Incarnato, D., Krepelova, A., Parlato, C., & 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*(7), 1191–1205. https://doi.org/10.1038/nprot.2016.063

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Anselmi, F., Parlato, C., Medana, C., DalBello, F., & Oliviero, S. (2015). Single-Base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter DNA Methylation Dynamics. *Cell Reports*, *10*(5), 674–683. https://doi.org/10.1016/j.celrep.2015.01.008

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., & Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, *543*(7643), 72–77. https://doi.org/10.1038/nature21373

Nestor, C. E., Ottaviano, R., Reddington, J., Sproul, D., Reinhardt, D., Dunican, D., Katz, E., Dixon, J. M., Harrison, D. J., & Meehan, R. R. (2012). Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Research*, *22*(3), 467–477. https://doi.org/10.1101/gr.126417.111

Newport, J., & Kirschner, M. (1982a). A major developmental transition in early xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell*, *30*(3), 675–686. https://doi.org/10.1016/0092-8674(82)90272-0

Newport, J., & Kirschner, M. (1982b). A major developmental transition in early xenopus embryos: II. control of the onset of transcription. *Cell*, *30*(3), 687–696. https://doi.org/10.1016/0092-8674(82)90273-2

Ng, H. H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., & Bird, A. (1999). MBD2 is a transcriptional repressor belonging

to the MeCP1 histone deacetylase complex. *Nature Genetics*, *23*(1), 58–61. https://doi.org/10.1038/12659

Ngo, T. T. M., Yoo, J., Dai, Q., Zhang, Q., He, C., Aksimentiev, A., & Ha, T. (2016). Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms10813

Nieuwkoop, P. D., & Faber, J. (Eds.). (1994). *Normal Table of Xenopus Laevis (Daudin)*. Garland Science. https://doi.org/10.1201/9781003064565

Niu, L., Shen, W., Shi, Z., Tan, Y., He, N., Wan, J., Sun, J., Zhang, Y., Huang, Y., Wang, W., Fang, C., Li, J., Zheng, P., Cheung, E., Chen, Y., Li, L., & Hou, C. (2021). Three-dimensional folding dynamics of the Xenopus tropicalis genome. *Nature Genetics*, *53*(7), 1075–1087. https://doi.org/10.1038/s41588-021-00878-z

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, *99*(3), 247–257. https://doi.org/10.1016/s0092-8674(00)81656-6

Okano, M., Xie, S., & Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genetics*, *19*(3), 219–220. https://doi.org/10.1038/890

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., Graves, B. J., & Cairns, B. R. (2010). Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nature Structural & Molecular Biology*, *17*(5), 620–628. https://doi.org/10.1038/nsmb.1801

Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H., & Hayashi, Y. (2002). LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Research*, *62*(14), 4075–4080.

Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C. D., Cheng, X., & Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, *448*(7154), 714–717. https://doi.org/10.1038/nature05987

Orioli, A., Praz, V., Lhôte, P., & Hernandez, N. (2016). Human MAF1 targets and represses active RNA polymerase III genes by preventing recruitment rather than inducing long-term transcriptional arrest. *Genome Research*, *26*(5), 624–635. https://doi.org/10.1101/gr.201400.115

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., & Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. *Current Biology*, *10*(8), 475–478. https://doi.org/10.1016/S0960-9822(00)00448-6

Owens, N. D. L., Blitz, I. L., Lane, M. A., Patrushev, I., Overton, J. D., Gilchrist, M. J., Cho, K. W. Y., & Khokha, M. K. (2016). Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. *Cell Reports*, *14*(3), 632–647. https://doi.org/10.1016/J.CELREP.2015.12.050

Palancade, B., Bellier, S., Almouzni, G., & Bensaude, O. (2001). Incomplete RNA polymerase II phosphorylation in Xenopus laevis early embryos. *Journal of Cell Science*, *114*(Pt 13), 2483–2489. https://doi.org/10.1242/jcs.114.13.2483

Paranjpe, S. S., Jacobi, U. G., van Heeringen, S. J., & Veenstra, G. J. C. (2013). A genome-wide survey of maternal and embryonic transcripts during Xenopus tropicalis development. *BMC Genomics*, *14*, 762. https://doi.org/10.1186/1471-2164-14-762

Pardue, M. L., Brown, D. D., & Birnstiel, M. L. (1973). Location of the genes for 5S ribosomal RNA in Xenopus laevis. *Chromosoma*, *42*(2), 191–203. https://doi.org/10.1007/BF00320940

Pastor, W. A., Pape, U. J., Huang, Y., Henderson, H. R., Lister, R., Ko, M., McLoughlin, E. M., Brudno, Y., Mahapatra, S., Kapranov, P., Tahiliani, M., Daley, G. Q., Liu, X. S., Ecker, J. R., Milos, P. M., Agarwal, S., & Rao, A. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*, *473*(7347), 394–397. https://doi.org/10.1038/nature10102

Pelechano, V., Chavez, S., & Perez-Ortin, J. E. (2010). A Complete Set of Nascent Transcription Rates for Yeast Genes. *PloS One*, *5*(11), 1–10. https://doi.org/https://doi.org/10.1371/journal.pone.0015442

Pendina, A. A., Efimova, O. A., Krapivin, M. I., Mekina, I. D., Tikhonov, A. V., Koltsova, A. S., Petrovskaia-Kaminskaia, A. V., Chiryaeva, O. G., Kogan, I. Y., Gzgzyan, A. M., & Baranov, V. S. (2018). Genomic distribution of 5-formylcytosine and 5-carboxylcytosine in human preimplantation embryos. *Molecular Reproduction and Development*, *85*(12), 893–895. https://doi.org/10.1002/mrd.23074

Pérez-Montero, S., Carbonell, A., Morán, T., Vaquero, A., & Azorín, F. (2013). The Embryonic Linker Histone H1 Variant of Drosophila, dBigH1, Regulates Zygotic Genome Activation. *Developmental Cell*, *26*(6), 578–590. https://doi.org/10.1016/j.devcel.2013.08.011

Peshkin, L., Wühr, M., Pearl, E., Haas, W., Freeman, R. M., Gerhart, J. C., Klein, A. M., Horb, M., Gygi, S. P., & Kirschner, M. W. (2015). On the Relationship of Protein and mRNA Dynamics in Vertebrate Embryonic Development. *Developmental Cell*, *35*(3), 383–394. https://doi.org/10.1016/j.devcel.2015.10.010

Peterson, R. C., Doering, J. L., & Brown, D. D. (1980). Characterization of two xenopus somatic 5S DNAs and one minor oocyte-specific 5S DNA. *Cell*, *20*(1), 131–141. https://doi.org/10.1016/0092-8674(80)90241-X

Phillips, J. E., & Corces, V. G. (2009). CTCF: Master Weaver of the Genome. *Cell*, *137*(7), 1194–1211. https://doi.org/10.1016/j.cell.2009.06.001

Potapov, V., Fu, X., Dai, N., Corrêa, I. R., Tanner, N. A., & Ong, J. L. (2018). Base modifications affecting RNA polymerase and reverse transcriptase fidelity. *Nucleic Acids Research*, *46*(11), 5753–5763. https://doi.org/10.1093/nar/gky341

Potok, M. E., Nix, D. A., Parnell, T. J., & Cairns, B. R. (2013). Reprogramming the Maternal Zebrafish Genome after Fertilization to Match the Paternal Methylation Pattern. *Cell*, *153*(4), 759–772. https://doi.org/10.1016/j.cell.2013.04.030

Procunier, J. D., & Tartof, K. D. (1975). Genetic analysis of the 5S RNA genes in Drosophila Melanogaster. *Genetics*, *81*(3), 515–523. https://doi.org/10.1093/genetics/81.3.515

Prokhortchouk, A., Hendrich, B., Jørgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., & Prokhortchouk, E. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes & Development*, *15*(13), 1613–1618. https://doi.org/10.1101/gad.198501

Raab, J. R., Chiu, J., Zhu, J., Katzman, S., Kurukuti, S., Wade, P. A., Haussler, D., & Kamakaka, R. T. (2012). Human tRNA genes function as chromatin insulators. *The EMBO Journal*, *31*(2), 330–350. https://doi.org/10.1038/emboj.2011.406

Raiber, E. A., Beraldi, D., Ficz, G., Burgess, H. E., Branco, M. R., Murat, P., Oxley, D., Booth, M. J., Reik, W., & Balasubramanian, S. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biology*, *13*(8), R69. https://doi.org/10.1186/gb-2012-13-8-r69

Raiber, E. A., Murat, P., Chirgadze, D. Y., Beraldi, D., Luisi, B. F., & Balasubramanian, S. (2015). 5-formylcytosine alters the structure of the DNA double helix. *Nature Structural and Molecular Biology*, *22*(1), 44–49. https://doi.org/10.1038/nsmb.2936

Raiber, E. A., Portella, G., Martínez Cuesta, S., Hardisty, R., Murat, P., Li, Z., Iurlaro, M., Dean, W., Spindel, J., Beraldi, D., Liu, Z., Dawson, M. A., Reik, W., & Balasubramanian, S. (2018). 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. *Nature Chemistry*, *10*(12), 1258–1266. https://doi.org/10.1038/s41557-018-0149-x

Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P., & Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(10), 5237–5242. https://doi.org/10.1073/pnas.97.10.5237

Ramsay, E. P., Abascal-Palacios, G., Daiß, J. L., King, H., Gouge, J., Pilsl, M., Beuron, F., Morris, E., Gunkel, P., Engel, C., & Vannini, A. (2020). Structure of human RNA polymerase III. *Nature Communications*, *11*(1), 6409. https://doi.org/10.1038/s41467-020-20262-5

Roberts, D. N., Stewart, A. J., Huff, J. T., & Cairns, B. R. (2003). The RNA polymerase III transcriptome revealed by genome-wide localization and activity-occupancy relationships. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(25), 14695–14700. https://doi.org/10.1073/pnas.2435566100

Roeder, R. G. (1974). Multiple forms of deoxyribonucleic acid-dependent ribonucleic acid polymerase in Xenopus laevis. Levels of activity during oocyte and embryonic development. *The Journal of Biological Chemistry*, *249*(1), 249–256.

Rosen, A., Sarid, S., & Daniel, V. (1984). Genes and pseudogenes in a reiterated rat tRNA gene cluster. *Nucleic Acids Research*, *12*(12), 4893–4906. https://doi.org/10.1093/nar/12.12.4893

Rougier, N., Bourc'his, D., Gomes, D. M., Niveleau, A., Plachot, M., Pàldi, A., & Viegas-Péquignot, E. (1998). Chromosome methylation patterns during mammalian

preimplantation development. *Genes & Development*, *12*(14), 2108–2113. https://doi.org/10.1101/gad.12.14.2108

Ruzov, A., Dunican, D. S., Prokhortchouk, A., Pennings, S., Stancheva, I., Prokhortchouk, E., & Meehan, R. R. (2004). Kaiso is a genome-wide repressor of transcription that is essential for amphibian development. *Development*, *131*(24), 6185–6194. https://doi.org/10.1242/dev.01549

Ruzov, A., Tsenkina, Y., Serio, A., Dudnakova, T., Fletcher, J., Bai, Y., Chebotareva, T., Pells, S., Hannoun, Z., Sullivan, G., Chandran, S., Hay, D. C., Bradley, M., Wilmut, I., & De Sousa, P. (2011). Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell Research*, *21*(9), 1332–1342. https://doi.org/10.1038/cr.2011.113

Saeki, H., Ohsumi, K., Aihara, H., Ito, T., Hirose, S., Ura, K., & Kaneda, Y. (2005). Linker histone variants control chromatin dynamics during early embryogenesis. *Proceedings of the National Academy of Sciences*, *102*(16), 5697–5702. https://doi.org/10.1073/pnas.0409824102

Sanford, J. P., Clark, H. J., Chapman, V. M., & Rossant, J. (1987). Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. *Genes & Development*, *1*(10), 1039–1046. https://doi.org/10.1101/gad.1.10.1039

Santos, F., Hendrich, B., Reik, W., & Dean, W. (2002). Dynamic Reprogramming of DNA Methylation in the Early Mouse Embryo. *Developmental Biology*, *241*(1), 172–182. https://doi.org/10.1006/dbio.2001.0501

Santos, F., Peat, J., Burgess, H., Rada, C., Reik, W., & Dean, W. (2013). Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics & Chromatin*, *6*(1), 39. https://doi.org/10.1186/1756-8935-6-39

Schmidt-Zachmann, M. S., Hügle, B., Scheer, U., & Franke, W. W. (1984). Identification and localization of a novel nucleolar protein of high molecular weight by a monoclonal antibody. *Experimental Cell Research*, *153*(2), 327–346. http://www.ncbi.nlm.nih.gov/pubmed/6539710

Schomacher, L., Han, D., Musheev, M. U., Arab, K., Kienhöfer, S., von Seggern, A., & Niehrs, C. (2016). Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nature Structural & Molecular Biology*, *23*(2), 116–124. https://doi.org/10.1038/nsmb.3151

Schomacher, L., & Niehrs, C. (2017a). DNA repair and erasure of 5-methylcytosine in vertebrates. *BioEssays*, *39*(3), 1600218. https://doi.org/10.1002/bies.201600218

Schomacher, L., & Niehrs, C. (2017b). DNA repair and erasure of 5-methylcytosine in vertebrates. *BioEssays*, *39*(3), 1600218. https://doi.org/10.1002/bies.201600218

Schramm, L., & Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. *Genes & Development*, *16*(20), 2593–2620. https://doi.org/10.1101/gad.1018902

Schramm, L., Pendergrast, P. S., Sun, Y., & Hernandez, N. (2000). Different human TFIIIB activities direct RNA polymerase III transcription from TATA-containing and TATA-less promoters. *Genes & Development*, *14*(20), 2650–2663. https://doi.org/10.1101/gad.836400

Schultz, P., Marzouki, N., Marck, C., Ruet, A., Oudet, P., & Sentenac, A. (1989). The two DNA-binding domains of yeast transcription factor tau as observed by scanning transmission electron microscopy. *The EMBO Journal*, *8*(12), 3815–3824. https://doi.org/10.1002/j.1460-2075.1989.tb08559.x

Schulz, K. N., & Harrison, M. M. (2019). Mechanisms regulating zygotic genome activation. *Nature Reviews Genetics*, *20*(4), 221–234. https://doi.org/10.1038/s41576-018-0087-x

Schwartz, C. M., Hussain, M. S., Blenner, M., & Wheeldon, I. (2016). Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR–Cas9-Mediated Genome Editing in Yarrowia lipolytica. *ACS Synthetic Biology*, *5*(4), 356–359. https://doi.org/10.1021/acssynbio.5b00162

Session, A. M., Uno, Y., Kwon, T., Chapman, J. A., Toyoda, A., Takahashi, S., Fukui, A., Hikosaka, A., Suzuki, A., Kondo, M., van Heeringen, S. J., Quigley, I., Heinz, S., Ogino, H., Ochi, H., Hellsten, U., Lyons, J. B., Simakov, O., Putnam, N., … Rokhsar, D. S. (2016). Genome evolution in the allotetraploid frog Xenopus laevis. *Nature*, *538*(7625), 336–343. https://doi.org/10.1038/nature19840

Sha, Q.-Q., Zhu, Y.-Z., Li, S., Jiang, Y., Chen, L., Sun, X.-H., Shen, L., Ou, X.-H., & Fan, H.-Y. (2020). Characterization of zygotic genome activation-dependent maternal mRNA clearance in mouse. *Nucleic Acids Research*, *48*(2), 879–894. https://doi.org/10.1093/nar/gkz1111

Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., & Zhang, Y. (2014). Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes. *Cell Stem Cell*, *15*(4), 459–471. https://doi.org/10.1016/j.stem.2014.09.002

Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A. C., Fung, H. L., Zhang, K., & Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell*, *153*(3), 692–706. https://doi.org/10.1016/j.cell.2013.04.002

Shen, L., & Zhang, Y. (2013). 5-Hydroxymethylcytosine: generation, fate, and genomic distribution. *Current Opinion in Cell Biology*, *25*(3), 289–296. https://doi.org/10.1016/j.ceb.2013.02.017

Shen, Q., Zhang, Q., Shi, Y., Shi, Q., Jiang, Y., Gu, Y., Li, Z., Li, X., Zhao, K., Wang, C., Li, N., & Cao, X. (2018). Tet2 promotes pathogen infection-induced myelopoiesis through mRNA oxidation. *Nature*, *554*(7690), 123–127. https://doi.org/10.1038/nature25434

Shibutani, T., Ito, S., Toda, M., Kanao, R., Collins, L. B., Shibata, M., Urabe, M., Koseki, H., Masuda, Y., Swenberg, J. A., Masutani, C., Hanaoka, F., Iwai, S., & Kuraoka, I. (2014). Guanine- 5-carboxylcytosine base pairs mimic mismatches during DNA replication. *Scientific Reports*, *4*(1), 5220. https://doi.org/10.1038/srep05220

Shi, D.-Q., Ali, I., Tang, J., & Yang, W.-C. (2017). New Insights into 5hmC DNA Modification: Generation, Distribution and Function. *Frontiers in Genetics*, *8*. https://doi.org/10.3389/fgene.2017.00100

Shortridge, R. D., Johnson, G. D., Craig, L. C., Pirtle, I. L., & Pirtle, R. M. (1989). A human tRNA gene heterocluster encoding threonine, proline and valine tRNAs. *Gene*, *79*(2), 309–324. https://doi.org/10.1016/0378-1119(89)90213-8

Sinsheimer, R. L. (1954). The action of pancreatic desoxyribonuclease. I. Isolation of mono- and dinucleotides. *The Journal of Biological Chemistry*, *208*(1), 445–459.

Skvortsova, K., Stirzaker, C., & Taberlay, P. (2019). The DNA methylation landscape in cancer. *Essays in Biochemistry*, *63*(6), 797–811. https://doi.org/10.1042/EBC20190037

Slenn, T. J., Morris, B., Havens, C. G., Freeman, R. M., Takahashi, T. S., & Walter, J. C. (2014). Thymine DNA Glycosylase Is a CRL4Cdt2 Substrate. *Journal of Biological Chemistry*, *289*(33), 23043–23055. https://doi.org/10.1074/jbc.M114.574194

Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H., Gnirke, A., Regev, A., & Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, *484*(7394), 339–344. https://doi.org/10.1038/nature10960

Song, C.-X., Diao, J., Brunger, A. T., & Quake, S. R. (2016). Simultaneous single-molecule epigenetic imaging of DNA methylation and hydroxymethylation. *Proceedings of the National Academy of Sciences*, *113*(16), 4338–4343. https://doi.org/10.1073/pnas.1600223113

Song, C. X., Szulwach, K. E., Dai, Q., Fu, Y., Mao, S. Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., Gao, J., Liu, P., Li, L., Xu, G. L., Jin, P., & He, C. (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell*, *153*(3), 678–691. https://doi.org/10.1016/j.cell.2013.04.001

Song, C.-X., Szulwach, K. E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.-H., Zhang, W., Jian, X., Wang, J., Zhang, L., Looney, T. J., Zhang, B., Godley, L. A., Hicks, L. M., Lahn, B. T., Jin, P., & He, C. (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature Biotechnology*, *29*(1), 68–72. https://doi.org/10.1038/nbt.1732

Song, J., & Pfeifer, G. P. (2016). Are there specific readers of oxidized 5-methylcytosine bases? *BioEssays*, *38*(10), 1038–1047. https://doi.org/10.1002/bies.201600126

Soragni, E., & Kassavetis, G. A. (2008). Absolute Gene Occupancies by RNA Polymerase III, TFIIIB, and TFIIIC in Saccharomyces cerevisiae. *Journal of Biological Chemistry*, *283*(39), 26568–26576. https://doi.org/10.1074/jbc.M803769200

Sørensen, P. D., & Frederiksen, S. (1991). Characterization of human 5S rRNA genes. *Nucleic Acids Research*, *19*(15), 4147–4151. https://doi.org/10.1093/nar/19.15.4147

Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W. T. C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., Eberl, H. C., Mensinga, A., Brinkman, A. B., Lephikov, K., Müller, U., Walter, J., Boelens, R., van Ingen, H., Leonhardt, H., … Vermeulen, M. (2013). Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell*, *152*(5), 1146–1159. https://doi.org/10.1016/j.cell.2013.02.004

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., Tiwari, V. K., & Schübeler, D. (2011). DNA-

binding factors shape the mouse methylome at distal regulatory regions. *Nature*, *480*(7378), 490–495. https://doi.org/10.1038/nature10716

Stapel, L. C., Zechner, C., & Vastenhouw, N. L. (2017). Uniform gene expression in embryos is achieved by temporal averaging of transcription noise. *Genes & Development*, *31*(16), 1635–1640. https://doi.org/10.1101/gad.302935.117

Steffensen, D. M., Duffey, P., & Prensky, W. (1974). Localisation of 5S ribosomal RNA genes on human chromosome 1. *Nature*, *252*(5485), 741–743. https://doi.org/10.1038/252741a0

Stolz, P., Mantero, A. S., Tvardovskiy, A., Ugur, E., Wange, L. E., Mulholland, C. B., Cheng, Y., Wierer, M., Enard, W., Schneider, R., Bartke, T., Leonhardt, H., Elsässer, S. J., & Bultmann, S. (2022). TET1 regulates gene expression and repression of endogenous retroviruses independent of DNA demethylation. *Nucleic Acids Research*, *50*(15), 8491–8511. https://doi.org/10.1093/nar/gkac642

Stutz, F., Gouilloud, E., & Clarkson, S. G. (1989). Oocyte and somatic tyrosine tRNA genes in Xenopus laevis. *Genes & Development*, *3*(8), 1190–1198. https://doi.org/10.1101/gad.3.8.1190

Sun, L., Champion, M. M., Huber, P. W., & Dovichi, N. J. (2016). Proteomics of Xenopus development. *Molecular Human Reproduction*, *22*(3), 193–199. https://doi.org/10.1093/molehr/gav052

Sun, Z., Dai, N., Borgaro, J. G., Quimby, A., Sun, D., Corrêa, I. R., Zheng, Y., Zhu, Z., & Guan, S. (2015). A Sensitive approach to map genome-wide 5-Hydroxymethylcytosine and 5-Formylcytosine at single-base resolution. *Molecular Cell*, *57*(4), 750–761. https://doi.org/10.1016/j.molcel.2014.12.035

Szulwach, K. E., Li, X., Li, Y., Song, C.-X., Han, J. W., Kim, S., Namburi, S., Hermetz, K., Kim, J. J., Rudd, M. K., Yoon, Y.-S., Ren, B., He, C., & Jin, P. (2011). Integrating 5-Hydroxymethylcytosine into the Epigenomic Landscape of Human Embryonic Stem Cells. *PLoS Genetics*, *7*(6), e1002154. https://doi.org/10.1371/journal.pgen.1002154

Szwagierczak, A., Bultmann, S., Schmidt, C. S., Spada, F., & Leonhardt, H. (2010). Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Research*, *38*(19). https://doi.org/10.1093/nar/gkq684

Tadros, W., & Lipshitz, H. D. (2009). The maternal-to-zygotic transition: a play in two acts. *Development*, *136*(18), 3033–3042. https://doi.org/10.1242/dev.033183

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., & Rao, A. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science*, *324*(5929), 930–935. https://doi.org/10.1126/science.1170116

Tamanaha, E., Guan, S., Marks, K., & Saleh, L. (2016). Distributive Processing by the Iron(II)/α-Ketoglutarate-Dependent Catalytic Domains of the TET Enzymes Is Consistent with Epigenetic Roles for Oxidized 5-Methylcytosine Bases. *Journal of the American Chemical Society*, *138*(30), 9345–9348. https://doi.org/10.1021/jacs.6b03243

Tandon, P., Conlon, F., Furlow, J. D., & Horb, M. E. (2017). Expanding the genetic toolkit in Xenopus: Approaches and opportunities for human disease modeling. *Developmental Biology*, *426*(2), 325–335. https://doi.org/10.1016/j.ydbio.2016.04.009

Tani, H., & Akimitsu, N. (2012). Genome-wide technology for determining RNA stability in mammalian cells. *RNA Biology*, *9*(10), 1233–1238. https://doi.org/10.4161/rna.22036

Tan, M. H., Au, K. F., Yablonovitch, A. L., Wills, A. E., Chuang, J., Baker, J. C., Wong, W. H., & Li, J. B. (2013). RNA sequencing reveals a diverse and dynamic repertoire of the Xenopus tropicalis transcriptome over development. *Genome Research*, *23*(1), 201–216. https://doi.org/10.1101/gr.141424.112

Tesařék, J., Kopečný, V., Plachot, M., & Mandelbaum, J. (1988). Early morphological signs of embryonic genome expression in human preimplantation development as revealed by quantitative electron microscopy. *Developmental Biology*, *128*(1), 15–20. https://doi.org/10.1016/0012-1606(88)90261-8

Thompson, M., Haeusler, R. A., Good, P. D., & Engelke, D. R. (2003). Nucleolar Clustering of Dispersed tRNA Genes. *Science*, *302*(5649), 1399–1401. https://doi.org/10.1126/science.1089814

Tremblay, K. D., Duran, K. L., & Bartolomei, M. S. (1997). A 5' 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development. *Molecular and Cellular Biology*, *17*(8), 4322–4329. https://doi.org/10.1128/MCB.17.8.4322

Turowski, T. W., Leśniewska, E., Delan-Forino, C., Sayou, C., Boguta, M., & Tollervey, D. (2016). Global analysis of transcriptionally engaged yeast RNA polymerase III reveals extended tRNA transcripts. *Genome Research*, *26*(7), 933–944. https://doi.org/10.1101/gr.205492.116

Turowski, T. W., & Tollervey, D. (2016). Transcription by RNA polymerase III: Insights into mechanism and regulation. *Biochemical Society Transactions*, *44*(5), 1367–1375. https://doi.org/10.1042/BST20160062

Ura, K., Nightingale, K., & Wolffe, A. P. (1996). Differential association of HMG1 and linker histones B4 and H1 with dinucleosomal DNA: structural transitions and transcriptional repression. *The EMBO Journal*, *15*(18), 4959–4969.

Van Bortle, K., & Corces, V. G. (2012). tDNA insulators and the emerging role of TFIIIC in genome organization. *Transcription*, *3*(6), 277–284. https://doi.org/10.4161/trns.21579

van Heeringen, S. J., Akkers, R. C., van Kruijsbergen, I., Arif, M. A., Hanssen, L. L. P., Sharifi, N., & Veenstra, G. J. C. (2014). Principles of nucleation of H3K27 methylation during embryonic development. *Genome Research*, *24*(3), 401–410. https://doi.org/10.1101/gr.159608.113

Vassena, R., Boué, S., González-Roca, E., Aran, B., Auer, H., Veiga, A., & Belmonte, J. C. I. (2011). Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development*, *138*(17), 3699–3709. https://doi.org/10.1242/dev.064741

Vastenhouw, N. L., Cao, W. X., & Lipshitz, H. D. (2019). The maternal-to-zygotic transition revisited. *Development*, *146*(11). https://doi.org/10.1242/dev.161471

Veenstra, G. J. C., & Wolffe, A. P. (2001). Constitutive genomic methylation during embryonic development of Xenopus. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, *1521*(1–3), 39–44. https://doi.org/10.1016/S0167-4781(01)00280-9

Veenstra, G. J., Destrée, O. H., & Wolffe, A. P. (1999). Translation of maternal TATA-binding protein mRNA potentiates basal but not activated transcription in Xenopus embryos at the midblastula transition. *Molecular and Cellular Biology*, *19*(12), 7972–7982. https://doi.org/10.1128/MCB.19.12.7972

Veenstra, G. J., Mathu, M. T., & Destrée, O. H. (1999). The Oct-1 POU domain directs developmentally regulated nuclear translocation in Xenopus embryos. *Biological Chemistry*, *380*(2), 253–257. https://doi.org/10.1515/BC.1999.033

Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K., & Pasini, D. (2013). Tet Proteins Connect the O-Linked N-acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells. *Molecular Cell*, *49*(4), 645–656. https://doi.org/10.1016/j.molcel.2012.12.019

Vize, P. D., & Zorn, A. M. (2017). Xenopus genomic data and browser resources. *Developmental Biology*, *426*(2), 194–199. https://doi.org/10.1016/j.ydbio.2016.03.030

Wagner, G., Singhal, N., Nicetto, D., Straub, T., Kremmer, E., & Rupp, R. A. W. (2017). Brg1 chromatin remodeling ATPase balances germ layer patterning by amplifying the transcriptional burst at midblastula transition. *PLOS Genetics*, *13*(5), e1006757. https://doi.org/10.1371/journal.pgen.1006757

Wagner, M., Steinbacher, J., Kraus, T. F. J., Michalakis, S., Hackner, B., Pfaffeneder, T., Perera, A., Müller, M., Giese, A., Kretzschmar, H. A., & Carell, T. (2015). Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. *Angewandte Chemie*, *54*(42), 12511–12514. https://doi.org/10.1002/anie.201502722

Wang, C., Politz, J. C., Pederson, T., & 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*(6), 2425–2435. https://doi.org/10.1091/mbc.e02-12-0818

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., Li, W., Zhou, Q., Aluru, N., Tang, F., He, C., Huang, X., & Liu, J. (2014). Programming and Inheritance of Parental DNA Methylomes in Mammals. *Cell*, *157*(4), 979–991. https://doi.org/10.1016/j.cell.2014.04.017

Wang, L., Zhou, Y., Xu, L., Xiao, R., Lu, X., Chen, L., Chong, J., Li, H., He, C., Fu, X. D., & Wang, D. (2015). Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature*, *523*(7562), 621–625. https://doi.org/10.1038/nature14482

Wang, W., Côté, J., Xue, Y., Zhou, S., Khavari, P. A., Biggar, S. R., Muchardt, C., Kalpana, G. V, Goff, S. P., Yaniv, M., Workman, J. L., & Crabtree, G. R. (1996). Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. *The EMBO Journal*, *15*(19), 5370–5382.

Watson, J. D., & Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, *171*(4356), 737–738. https://doi.org/10.1038/171737a0

Watt, F., & Molloy, P. L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes & Development*, *2*(9), 1136–1143. https://doi.org/10.1101/gad.2.9.1136

Weber, A. R., Krawczyk, C., Robertson, A. B., Kuśnierczyk, A., Vågbø, C. B., Schuermann, D., Klungland, A., & Schär, P. (2016). Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nature Communications*, *7*, 10806. https://doi.org/10.1038/ncomms10806

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M., & Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics*, *39*(4), 457–466. https://doi.org/10.1038/ng1990

Wheldon, L. M., Abakir, A., Ferjentsik, Z., Dudnakova, T., Strohbuecker, S., Christie, D., Dai, N., Guan, S., Foster, J. M., Corrêa, I. R., Loose, M., Dixon, J. E., Sottile, V., Johnson, A. D., & Ruzov, A. (2014). Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. *Cell Reports*, *7*(5), 1353–1361. https://doi.org/10.1016/j.celrep.2014.05.003

Wickham, H. (2016). *Ggplot2 - Elegant Graphics for Data Analysis*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4

Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. A. C., Rappsilber, J., & Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*, *473*(7347), 343–349. https://doi.org/10.1038/nature10066

Wilson, C. J., Chao, D. M., Imbalzano, A. N., Schnitzler, G. R., Kingston, R. E., & Young, R. A. (1996). RNA Polymerase II Holoenzyme Contains SWI/SNF Regulators Involved in Chromatin Remodeling. *Cell*, *84*(2), 235–244. https://doi.org/10.1016/S0092-8674(00)80978-2

Woodland, H. (1982). The translational control phase of early development. *Bioscience Reports*, *2*(7), 471–491. https://doi.org/10.1007/BF01115246

Woodland, H. R., & Gurdon, J. B. (1969). RNA synthesis in an amphibian nuclear-transplant hybrid. *Developmental Biology*, *20*(2), 89–104. https://doi.org/10.1016/0012-1606(69)90007-4

Wood, L., Hatzenbuhler, N., Peterson, R., & Vogeli, G. (1991). Isolation of a mouse genomic clone containing four tRNACys-encoding genes. *Gene*, *98*(2), 249–252. https://doi.org/10.1016/0378-1119(91)90181-a

Workman, J. L., & Kingston, R. E. (1998). ALTERATION OF NUCLEOSOME STRUCTURE AS A MECHANISM OF TRANSCRIPTIONAL REGULATION. *Annual Review of Biochemistry*, *67*(1), 545–579. https://doi.org/10.1146/annurev.biochem.67.1.545

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C. J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., & Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian

zygote is linked with epigenetic reprogramming. *Nature Communications*, *2*(1), 241. https://doi.org/10.1038/ncomms1240

Wu, F., Li, X., Looso, M., Liu, H., Ding, D., Günther, S., Kuenne, C., Liu, S., Weissmann, N., Boettger, T., Atzberger, A., Kolahian, S., Renz, H., Offermanns, S., Gärtner, U., Potente, M., Zhou, Y., Yuan, X., & Braun, T. (2023). Spurious transcription causing innate immune responses is prevented by 5-hydroxymethylcytosine. *Nature Genetics*, *55*(1), 100–111. https://doi.org/10.1038/s41588-022-01252-3

Wu, H., D'Alessio, A. C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y. E., & 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*(7), 679–684. https://doi.org/10.1101/gad.2036011

Wu, H., Wu, X., Shen, L., & Zhang, Y. (2014). Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nature Biotechnology*, *32*(12), 1231–1240. https://doi.org/10.1038/nbt.3073

Wu, S. C., & Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. *Nature Reviews Molecular Cell Biology*, *11*(9), 607–620. https://doi.org/10.1038/nrm2950

Wu, X., & Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nature Reviews Genetics*, *18*(9), 517–534. https://doi.org/10.1038/nrg.2017.33

Wyatt, G. R. (1950). Occurrence of 5-Methyl-Cytosine in Nucleic Acids. *Nature*, *166*(4214), 237–238. https://doi.org/10.1038/166237b0

Wyatt, G. R. (1951). Recognition and estimation of 5-methylcytosine in nucleic acids. *Biochemical Journal*, *48*(5), 581–584. https://doi.org/10.1042/bj0480581

Xia, B., Han, D., Lu, X., Sun, Z., Zhou, A., Yin, Q., Zeng, H., Liu, M., Jiang, X., Xie, W., He, C., & Yi, C. (2015). Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. *Nature Methods*, *12*(11), 1047–1050. https://doi.org/10.1038/nmeth.3569

Xiao, L., Parolia, A., Qiao, Y., Bawa, P., Eyunni, S., Mannan, R., Carson, S. E., Chang, Y., Wang, X., Zhang, Y., Vo, J. N., Kregel, S., Simko, S. A., Delekta, A. D., Jaber, M., Zheng, H., Apel, I. J., McMurry, L., Su, F., … Chinnaiyan, A. M. (2022). Targeting SWI/SNF ATPases in enhancer-addicted prostate cancer. *Nature*, *601*(7893), 434–439. https://doi.org/10.1038/s41586-021-04246-z

Xie, W., Barr, C. L., Kim, A., Yue, F., Lee, A. Y., Eubanks, J., Dempster, E. L., & Ren, B. (2012). Base-Resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. *Cell*, *148*(4), 816–831. https://doi.org/10.1016/j.cell.2011.12.035

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A. J., Zheng, L., Zhang, H., Huang, S., Min, J., Nicholson, T., Chen, T., Xu, G., Shi, Y., Zhang, K., & Shi, Y. G. (2011). Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. *Molecular Cell*, *42*(4), 451–464. https://doi.org/10.1016/j.molcel.2011.04.005

Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J. G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., Diao, J., Wu, F., He, H. H., Cui, Q., Clark, E., Ma, C., Barbara, A., Veenstra, G. J. C., Xu, G., … Shi, Y. G. (2012). Tet3 CXXC Domain and Dioxygenase Activity Cooperatively

Regulate Key Genes for Xenopus Eye and Neural Development. *Cell*, *151*(6), 1200–1213. https://doi.org/10.1016/j.cell.2012.11.014

Yanai, I., Peshkin, L., Jorgensen, P., & Kirschner, M. W. (2011). Mapping Gene Expression in Two Xenopus Species: Evolutionary Constraints and Developmental Flexibility. *Developmental Cell*, *20*(4), 483–496. https://doi.org/10.1016/j.devcel.2011.03.015

Yan, R., Cheng, X., Gu, C., Xu, Y., Long, X., Zhai, J., Sun, F., Qian, J., Du, Y., Wang, H., & Guo, F. (2023). Dynamics of DNA hydroxymethylation and methylation during mouse embryonic and germline development. *Nature Genetics*, *55*(1), 130–143. https://doi.org/10.1038/s41588-022-01258-x

Yildirim, O., Li, R., Hung, J.-H., Chen, P. B., Dong, X., Ee, L.-S., Weng, Z., Rando, O. J., & Fazzio, T. G. (2011). Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. *Cell*, *147*(7), 1498–1510. https://doi.org/10.1016/j.cell.2011.11.054

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P. K., Kivioja, T., Dave, K., Zhong, F., Nitta, K. R., Taipale, M., Popov, A., Ginno, P. A., Domcke, S., Yan, J., Schübeler, D., Vinson, C., & Taipale, J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, *356*(6337). https://doi.org/10.1126/science.aaj2239

Yoon, H.-G., Chan, D. W., Reynolds, A. B., Qin, J., & Wong, J. (2003). N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. *Molecular Cell*, *12*(3), 723–734. https://doi.org/10.1016/j.molcel.2003.08.008

You, C., Ji, D., Dai, X., & Wang, Y. (2014). Effects of Tet-mediated Oxidation Products of 5-Methylcytosine on DNA Transcription in vitro and in Mammalian Cells. *Scientific Reports*, *4*(1), 7052. https://doi.org/10.1038/srep07052

Yu, Z., Genest, P.-A., ter Riet, B., Sweeney, K., DiPaolo, C., Kieft, R., Christodoulou, E., Perrakis, A., Simmons, J. M., Hausinger, R. P., van Luenen, H. G. A. M., Rigden, D. J., Sabatini, R., & Borst, P. (2007). The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acids Research*, *35*(7), 2107–2115. https://doi.org/10.1093/nar/gkm049

Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G.-L., Luo, C., Jiang, H., & He, C. (2012). Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nature Chemical Biology*, *8*(4), 328–330. https://doi.org/10.1038/nchembio.914

Zhang, R., Lahens, N. F., Ballance, H. I., Hughes, M. E., & Hogenesch, J. B. (2014). A circadian gene expression atlas in mammals: Implications for biology and medicine. *Proceedings of the National Academy of Sciences*, *111*(45), 16219–16224. https://doi.org/10.1073/pnas.1408886111

Zhang, Y., Anderson, S. J., French, S. L., Sikes, M. L., Viktorovskaya, O. V., Huband, J., Holcomb, K., Hartman, J. L., Beyer, A. L., & Schneider, D. A. (2013). The SWI/SNF Chromatin Remodeling Complex Influences Transcription by RNA Polymerase I in Saccharomyces cerevisiae. *PLoS ONE*, *8*(2), e56793. https://doi.org/10.1371/journal.pone.0056793

Zhu, C., Gao, Y., Guo, H., Xia, B., Song, J., Wu, X., Zeng, H., Kee, K., Tang, F., & Yi, C. (2017). Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution. *Cell Stem Cell*, *20*(5), 720-731.e5. https://doi.org/10.1016/j.stem.2017.02.013

Zhu, J.-K. (2009). Active DNA demethylation mediated by DNA glycosylases. *Annual Review of Genetics*, *43*, 143–166. https://doi.org/10.1146/annurev-genet-102108-134205

Zhu, P., Guo, H., Ren, Y., Hou, Y., Dong, J., Li, R., Lian, Y., Fan, X., Hu, B., Gao, Y., Wang, X., Wei, Y., Liu, P., Yan, J., Ren, X., Yuan, P., Yuan, Y., Yan, Z., Wen, L., … Tang, F. (2018). Single-cell DNA methylome sequencing of human preimplantation embryos. *Nature Genetics*, *50*(1), 12–19. https://doi.org/10.1038/s41588-017-0007-6

Lebenslauf