Epigenetic regulation of gene expression by RNA modifications in Drosophila

Giulia Franco

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : CSV- Chimie et Sciences du Vivant
Spécialité : Chimie Biologie
Unité de recherche : Biologie et Biotechnologie pour la Santé

Régulation épigénétique des ARNs par méthylation chez la drosophile

Epigenetic regulation of gene expression by RNA modifications in Drosophila

Présentée par :

Giulia FRANCO

Direction de thèse :

Marie-odile FAUVARQUE Ingénieur-chercheur CEA, Université Grenoble Alpes	Directrice de thèse
Ramesh PILLAI Université de Genève	Co-directeur de thèse

Rapporteurs :

Donal O' CARROLL
PROFESSEUR, University of Edinburgh
Clément CARRE
MAITRE DE CONFERENCE HDR, Université Paris 6 - Pierre et Marie Curie

Thèse soutenue publiquement le **10 février 2023**, devant le jury composé de :

Marie-Odile FAUVARQUE DIRECTRICE DE RECHERCHE, CEA centre de Grenoble	Directrice de thèse
Donal O' CARROLL PROFESSEUR, University of Edinburgh	Rapporteur
Clément CARRE MAITRE DE CONFERENCE HDR, Université Paris 6 - Pierre et Marie Curie	Rapporteur
Emilie BRASSET PROFESSEUR DES UNIVERSITES, Université Clermont-Ferrand 1 - Auvergne	Examinatrice
Winfried WEISSENHORN PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes	Président
Ramesh PILLAI PROFESSEUR, Université de Genève	Co-directeur de thèse

Invités :

Emmanuel TAILLEBOURG
DOCTEUR EN SCIENCES,

ACKNOWLEDGEMENTS

My PhD work has been developped at CEA Grenoble in Gen&Chem Lab and it was co-founded by the Université Grenoble Alpes and Université de Genève. I am thankful to the two Universities as well as to the CSV Doctoral School to have given me the opportunity to work in such an exciting and dynamic environment.

I would like to thank the jury members Dr. Donal O' Carroll, Dr. Clément Carré, and Dr. Emilie Brasset and the president Dr. Winfried Weissenhorn for having accepted to evaluate my thesis manuscript and my PhD defense.

I wish to acknowledge the CSI members Dr. André Verdel and Dr. Wojciech Galej for the scientific exchanges and precious suggestions given to me during our TAC meetings.

I am immensely grateful to Marie-Odile, my supervisor, a constant mentor and pillar in these hard three years during which we experienced the Covid pandemic. You have always been by my side and encouraged me, I won't never forget how much you have taught to me.

I wish to thank my co-supervisor, Ramesh, who has been an inspiring and stimulating teacher. I had the unique chance to be guided by two excellent scientists who complemented each other and had strongly contributed to my scientific and personal growth. Despite Geneva is two hours far from Grenoble and we could not meet so frequently, I could easily interact and exchange with Ramesh's students. Thank you all guys for your contribution to my work and for the building discussions we had. We are lucky to be guided by such motivating leaders.

Thank you Emmanuel for your patience, teaching, support and presence, I knew that I could count on you and you have always been available to give a hand. I am so grateful to you.

Julie, I met you more than three years ago and I do not remember one single sad day spent with you. I have already told you several time that I will be owing you forever. You have become a friend, a person I have counted on for everything. Thank you from the bottom of my heart.

I acknowledge each single member of the Gen&Chem Lab, past and present, because I am sure I will never find again such a compact and supportive team as you are. You welcomed me with open arms since the first day and you have included me even if my French is still *'moche'*. I have learnt a lot from any of you. Thank you.

An unconditional thanks goes to my family, they have given me the wings to fly high. There are not enough words for explaining how much I love them.

Thank you Roberto, you know more than anyone else what these three years have meant for me and you have always been my compass. *Ad maiora.*

I would like to acknowledge Martina, my italian ‘partner in crime’, you have been present any day, any moment of this experience and I have found in you a real friend.

A huge thanks is for my italian friends Annina, Chiara, Maria and many more, who are spread all over the world. There are no mountains nor oceans that can separate us. We are stronger.

Merci et au revoir Silvia, Iris and Anna, we have spent great time and adventures together. You will always occupy a special place in my heart.

ABSTRACT

A huge number of epigenetic modifications occur on eukaryotic messenger RNAs such as the addition of a 5′ cap, a poly(A) tail or of internal marks including the methylation of adenosine ring at position 6 (m^6A). These modifications tightly regulate mRNAs stability and translational activity, therefore providing crucial additional levels of gene expression regulation.

The first chapter of my thesis focuses on the study of the PCIF1 protein using *Drosophila melanogaster* as a model. PCIF1 is a RNA cap-specific m^6A methyltransferase that binds to RNA polymerase II (RNA Pol II) and deposits a methyl group on the adenosine ring if the first transcribed nucleotide is an adenosine (m^6Am). In knock-out mice, the lack of PCIF1 results in reduced body weight and deregulation of a pool of genes in several tissues, particularly in testis. In *Drosophila*, Pcif1 is conserved but a single amino acid substitution inactivates the methyltransferase activity. I used CRISPR-Cas9 to create *Pcif1 Drosophila* mutants and performed their phenotypic and molecular analysis in order to determine the non-catalytic functions of Pcif1. I showed that *Pcif1* mutants displayed reduced fertility. Moreover, RNA sequencing revealed a set of dysregulated genes in *Pcif1* mutant ovaries indicating a role in gene expression regulation. I also observed that reduced *Pcif1* dosage in heterozygous flies induced a suppression of the *white* gene silencing due to the position-effect variegation (PEV), in which the euchromatic *white* gene is translocated near the heterochromatin rendering it sensitive to slight modifications of gene expression regulation. Finally, experiments performed by our collaborators in Pr. E. Brasset laboratory revealed that Pcif1 binds to euchromatin at interphasic chromosomal bands, indicating that it is associated with actively transcribed genes. Taken together, our results suggest that despite the lack of catalytic activity, Pcif1 is involved in gene expression regulation, putatively acting as a scaffolding protein recruited in the RNA Pol II complex, a hypothesis reinforced by the fact that Pcif1 is known to bind to the C-term of Pol II.

The second chapter of my thesis is centred on the identification of alternative RNA cap structures and decapping enzymes that ensure both the degradation of mRNA and their recycling into nucleotides, in *Drosophila*. The canonical cap is defined as the N^7-methylguanine (m^7G-cap) that is attached to the first transcribed nucleotide of eukaryotic mRNAs and hydrolysed by the conserved DCP2 decapping enzyme. DCP2 belongs to a large family of evolutionary conserved hydrolases, Nudix, which consists of 22 members in mammals that all share the ability to hydrolyse different metabolite substrates. In addition, several alternative cap structures have been identified so far in various bacteria or eukaryotes such as: Nicotinamide Adenine Dinucleotide (NAD), Flavine Dinucleotide (FAD), diphospho-CoenzymeA (dpCoA) and diadenosine-tetraphosphate (Ap_4A), Uridine Diphosphate Glucose (UDP-Glc), and Uridine Diphosphate N-acetylglucosamine (UDP-GlcNAc). Since little is known about the role of the Nudix family members and the putative presence of alternative cap structures in higher eukaryotes, my goal was to perform a functional analysis of the Nudix family in *Drosophila*. By using an RNAi-based genetic screen approach, I induced the silencing of each of the 12 *Nudix* genes in a tissue specific manner in *Drosophila*. My work pointed out that the closely related Nudix proteins Nudt19A and Nudt19B are required for full fly fertility. These proteins are predicted to hydrolyse Coenzyme A but nothing is known about their putative decapping activity. We generated CRISPR/Cas9 mutants of *Nudt19A* and *Nudt19B*, in order to further investigate their specific function in fly fertility and potential role in mRNA decapping in eukaryotes.

RÉSUMÉ

Un grand nombre de modifications épigénétiques ciblent les ARN messagers eucaryotes comme l'ajout d'une coiffe 5', d'une queue polyA ou de marques internes telles que la méthylation du cycle adénosine en position 6 (m^6A). Ces modifications régulent la stabilité et l'activité traductionnelle des ARNm et constituent des niveaux essentiels de régulation de l'expression génique chez les eucaryotes.

Le premier chapitre de ma thèse porte sur l'étude de l'ARN méthyltransférase PCIF1 chez la mouche *Drosophila melanogaster*. PCIF1 se lie à l'ARN polymerase II (ARN Pol II) et dépose un groupe méthyl sur le cycle adénosine (m^6Am) si le premier nucléotide transcrit est une adénosine. Chez la souris, l'absence de PCIF1 entraîne une perte de poids et une dérégulation de nombreux gènes dans plusieurs tissus, notamment dans le testicule. Chez la drosophile, Pcif1 est conservée mais une substitution d'un seul acide aminé inactive l'activité méthyltransférase. J'ai analysé les conséquences phénotypiques et moléculaires de mutants *Pcif1*générés par la méthodologie CRISPR/Cas9 afin de déterminer les fonctions indépendantes de l'activité catalytique de Pcif1 chez la drosophile. J'ai montré que les mutants *Pcif1* mâles et femelles présentaient une fertilité réduite. De plus, le séquençage des ARNm a révélé un ensemble de gènes dérégulés dans les ovaires mutants. J'ai également observé une suppression de la bigarrure de l'œil induite par effet de position (PEV) chez les hétérozygotes *Pcif1/+* qui témoigne d'un effet sur l'expression du gène *white*. Des expériences complémentaires réalisées au sein du laboratoire du Pr. E. Brasset ont révélé que Pcif1 se lie à l'euchromatine au niveau des bandes chromosomiques interphasiques correspondant aux gènes activement transcrits. L'ensemble de nos résultats montrent que malgré l'absence d'activité catalytique, Pcif1 est impliquée dans la régulation de l'expression génique chez la drosophile. Pcif1 pourrait agir comme modulateur au sein du complexe transcriptionnel, une hypothèse renforcée par le fait que Pcif1 se lie directement à la partie C-terminale de l'ARN Pol II.

Le deuxième chapitre de ma thèse est centré sur l'identification de structures alternatives de la coiffe et d'enzymes capables de les hydrolyser, conduisant à la dégradation des ARNm et au recyclage des nucléotides. La coiffe canonique consiste en une N7-méthylguanine (m^7G-cap) qui est attachée au premier nucléotide transcrit des ARNm eucaryotes ; elle est hydrolysée par DCP2. DCP2 appartient à une grande famille d'hydrolases conservées au cours de l'évolution, la famille Nudix, qui comprend 22 membres chez les mammifères. Plusieurs structures de coiffe alternatives ont été identifiées chez des bactéries ou divers eucaryotes, telles que : le Nicotinamide Adénine Dinucléotide (NAD), la Flavine Dinucléotide (FAD), le diphospho-CoenzymeA (dpCoA) et le diadénosine-tétraphosphate (Ap4A), l'Uridine Diphosphate Glucose (UDP-Glc) et l'Uridine Diphosphate N-acétylglucosamine (UDP-GlcNAc). Peu de données existent sur le rôle des membres de la famille Nudix et l'existence ou non de coiffes alternatives chez les eucaryotes supérieurs. Dans ce contexte, mon objectif était d'effectuer une analyse fonctionnelle de la famille Nudix chez la drosophile. En utilisant une approche de criblage génétique reposant sur l'expression d'ARN interférents de façon ciblée dans différents tissus, j'ai induit l'extinction de chacun des 12 gènes Nudix chez la drosophile. Mes travaux ont mis en évidence que les deux protéines Nudt19A et Nudt19B sont nécessaires à la fertilité normale des mouches. D'après leur séquence, ces protéines, qui sont étroitement apparentées, sont des hydrolases du Coenzyme A, une activité qui reste à confirmer expérimentalement. Nous avons généré des mutants CRISPR/Cas9 des deux gènes *Nudt19A* et *Nudt19B*, afin

d'étudier plus en détail leur fonction dans la fertilité et leur rôle potentiel dans l'hydrolyse d'une coiffe alternative chez les eucarytes.

ABBREVIATIONS

abd-A,B : abdominal-A, B

AD: Alzheimer Disease

Adar : Adenosine deaminase acting on RNA

ADP : Adenosine Diphosphate

ADPR : ADP-Ribose

ADPRC : Adenosine Diphosphate Ribocyclase

ALKBH5 : ALKB Homologue 5

ALYREF : Aly-REF carrier

Am : 2'O-methyladenosine

AMP : Adenosine Monophosphate

Ap_4A : diadenosine-tetraphosphate

APEX : Ascorbate Peroxidase

APOE4 : APOlipoprotein ε4

Ars2 : Arsenic resistance protein 2

A. thaliana : *Arabidopsis thaliana*

ATP : Adenosine Triphosphate

ATP5A : ATPase subunit 5 A

Aza-IP : 5-Azacytidine-mediated RNA Immunoprecipitation

Bam : Bag of marbles

BCSC : Breast Cancer Stem Cells

bFGF : basic Fibroblast Growth Factor

Bgcn : Benign gonial cell neoplasm

BLMH : Bleomycin Hydrolase

blw : bellwether

brat : brain tumour

BRD4 : BRomodomain-containing Domain protein 4

C. elegans : Caenorabditis elegans

CAPAM : Cap-specific Adenosine Methyltransferase

CBC : Cap Binding Complex

Cbp20 : Cap binding protein subunit 20

Cbp80 : Cap binding protein subunit 80

CCR4-NOT : Carbon Catabolite Repression-Negative On TATA-less

CD : Catalytic Domain

CFIm : Cleavage Factor Im

CDK2 : Cyclin Dependent Kinase 2

circRNA : circular RNA

CMTR1,2,3 : Cap Methyltransferase 1,2, 3

CNS : Central Nervous System

COS7 : CV-1 (simian) in Origin, and carrying the SV40 genetic material cells

CPSF : Cleavage and Polyadenylation Specificity Factor

CPV : Cytoplasmic Polyhedrosis Virus

CRC : ColoRectal Cancer

CstF : Cleavage stimulating Factor

CTD : C-Terminal Domain

dac : dachshund

DCP2 : Decapping enzyme 2

Dhh1 : DExD/H box RNA Helicase 1

DIPP : Diphospho Inositol Pyrophosphatase

DNA : Deoxiribonucleic Acid

dpCoA : diphospho-CoenzymeA

dpp : decapentaplegic

dsRNA : double strand RNA

dsx : doublesex

DVL3 : Dishevelled segment polarity protein 3

DXO : Decapping Exoribonuclease

E. coli : Escherichia coli

Edc : Enhancer of mRNA decapping

EGFR : Epithelial Growth Factor Receptor

eIF4 : eukaryotic Initiation Factor 4

eIFH: eukaryotic Initiation Factor H

Elav : Embryonic lethal abnormal vision

EMS : Ethyl Methane Sulfonate

ESCs : Embryonic Stem Cells

ey : eyeless

eya : eyes absent

FAD : Flavine Dinucleotide

FCP1: RNA polymerase II subunit A C-terminal domain Phosphatase 1

Fl(2)d : Female lethal-2-d

Flacc : Fl(2)/associated complex component

FOXM1 : Forkhead box M1

fru : fruitless

FTO : Fat mass and Obesity-associated protein

FZD7 : Frizzled class receptor 7

GDP : Guanosine Diphosphate

GMP : Guanosine Monophosphate

GMR-Gal4 : Glass Multimer Reporter Gal4

Gpc4: Glypican 4

Gro : Groucho

GRO-seq : Global Run-On sequencing

GSCs : Germline Stem Cells

GTP : Guanosine Triphosphate

GWAS : Genome Wide Association Studies

HEK293 : Human Embryonic Kidney cells

HeLa : Henrietta Lacks cells

Hepa 1-6 : Hepatocellular carcinoma cells

Hh : hedgegog

HIF : Hypoxia Inducing Factor

Hint : Histidine triad nucleotide-binding protein

HIV-1 : Human Immunodeficiency Virus 1

HNRNP : Heterogeneous Nuclear RiboNucleo Proteins

HQ-domain : histidine and glutamine rich domain

HSCs : Hematopoietic Stem Cells

Hsp70 : Heat-shock protein 70

IGF2BP1-3 : Insulin-like Growth Factor 2 mRNA-Binding Proteins 1–3

IHC : ImmunoHistoChemistry

Ime4 : Inducer of meiosis 4

Imp : IGF-II mRNA-binding protein

IFN1 : Interferon Type 1

IR : Inverted Repeat

ITGB6 : Integrin β6

KH : K Homology

KO : Knock Out

LARP1 : LA-Related Protein 1

LCN2 : Lipocalin-2

lncRNA : long non-codin RNA

Lsm1p: eukaryotic Like Sm protein 1

Lz : Lozenge

m^1A : N1-methyl adenosine

m^6A : N6-methyladenosine

m^6Am : N6-dimethyladenosine

m^5C : N5-methyl cytosine

m^7G : N7-methylguanosine

MAC : m^6A-METTL Complex

MACOM : m^6A-METTL Associated COMplex

MALAT-1 : Metastasis Associated Lung Adenocarcinoma Transcript 1

MAT2A : Methionine Adenosyl-Transferase 2A

MCCs : Merkel Cancer Cells

MCF7 : Michigan Cancer Foundation 7 cells

MDD : Major Depressive Disorder

Mei-P26 : Meiotic-P26

METTL14 : Methyltransferase-Like 14

METTL16 : Methyltransferase-Like 16

METTL3 : Methyltransferase-Like 3

MFSD4 : Major Facilitator Superfamily Domain containing 4

miCLIP : methylation-individual-nucleotide-resolution Crosslinking and Immunoprecipitation

miRNA : micro RNA

MITF : Microphthalmia-associated Transcription Factor

MNK1 : MAP Kinase interacting Serine/Threonine Kinases

Mot1p: Modifier of transcription protein 1

MP : Microprocessor components

mRNA : messanger RNA

msl-2 : male-specific lethal 2

mth: MuT homologue

mTORC1 : mammalian Target Of Rapamycin Complex 1

NAD : Nicotinamide Adenine Dinucleotide

NAFLD : Non-Alcoholic Fatty Liver Disease

NCBP1, 2 : Nuclear Cap Binding Protein 1, 2

NCINs : Non-Canonical Initiating Nucleotides

NET-seq : Native Elongating Transcript sequencing

NER : Nucleotide Excision Repair

NGT-Gal4 : *nanos*-Gal4

Nito : Spenito

NLS: Nuclear Localization Signal

NMD : Nonsense Mediated Decay

NMJ : Neuro-Muscular Junction

NRD : N-terminal Regulatory Domain

Nrg : Neuroglian

NSUN2 : RNA cytosine C(5)-methyltransferase

Nthl1 : Nth-like DNA glycosidase 1

NTP : Nucleotide TriPhosphate

NUDIX : NUcleoside DIphosphate linked to another moiety X

NXF1: Nuclear RNA eXport Factor 1

PABPN1 : Poly(A) Polymerase Binding Protein Nuclear 1

Paip2 : Polyadenylate binding protein interacting protein 2

PBE : Pumilio Binding Element

PDX1 : Pancreatic and Duodenal homeoboX 1

PAP : Poly(A) Polymerase

PARP : Poly ADP-Ribose Polymerase

P-bodies : Processing bodies

PCIF1 : Phosphorylated CTD Interacting Factor 1

PDH : Pyruvate DeHydrogenase

PDK4 : Pyruvate Dehydrogenase Kinase 4

PEV : Position Effect Variegation

Pin1 : Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1

PNRC2 : Proline-rich Nuclear Receptor Coactivator 2

Pol κ : RNA Pol II subunit κ

Poly(A) : poly Adenosine

PP-InsP5 : diphospho Inositol Pentakisphosphate

PRNC: Proline-Rich Nuclear Co-regulatory protein

Prrc2A : Proline rich coiled-coil 2A

P-TEFb : Positive Transcription Elongation Factor b

PUS : Pseudouridine Synthase

RAM : RNMT-Activating Mini-protein

RBM15/15B : RNA Binding Motif protein 15/15B

RBPs : RNA Binding Proteins

RD : Retina Determination

RGG : arginine/glycine-rich

RIG-I : Retinoic Acid Inducible Gene-I

RIP : RNA ImmunoPrecipitation

RNA : Ribonucleic Acid

RNAi : RNA interference

RNA Pol II : RNA Polymerase II

RNA seq: RNA sequencing

RNGTT : RNA GuanylylTransferase and 5'-Triphosphatase

RNMT : RNA guanine-7-Methyltransferase

RN7SK : RNA Component of 7SK nuclear ribonucleoprotein

RPL19 : Ribosomal Protein L19

RpS3 : Ribosomal protein S3

ROS : Reactive Oxygen Species

RRM : RNA Recognition Motif

rRNA : ribosomal RNA

RT-qPCR : Real Time quantitative Polymerase Chain Reaction

RUNX : Runt-Related Transcription Factor

S. cerevisiae : Saccharomyces cerevisiae

S. venezuelae : Streptomyces venezuelae

SAM : S-Adenosyl Methionine

scaRNA : small cajal body RNA

shep : shepard

shRNA: short hairpin RNA

siRNA: silencing RNA

snoRNA : small nucleolar RNA

SNP: Single Nucleotide Polymorphism

snRNA : small nuclear RNA

So : Sine oculis

Sp1 : Serine protease 1

SRSF3 : serine/arginine-rich (SR) Splicing Factor 3

SRY : Sex-determining Region Y protein

ssRNA : single strand RNA

Sxl : Sex lethal

T2D : Type 2 Diabetes

TCA : TriCarboxylic Acid

TCF4 : Transcription Factor 4

TCF7L2 : Transcription Factor 7 Like 2

TFIID : Transcription Factor II D

TLC : Thin Layer Chromatography

TOP : Terminal OligoPyrimidine

tra : *transformer*

TREX : TRanscription-EXport protein

TRMT6/61 : tRNA Methyltransferase 6/61

tRNA : transfer RNA

TSS : Transcription Start Site

UAS : Upstream Activating Sequence

UDP : Uridine DiPhosphate

UDP-Glc : Uridine DiPhosphate Glucose

UDP-GlcNAc : Uridine DiPhosphate N-Acetyl Glucosamine

UMP : Uridine Monophosphate

Upf1p: Up frameshift 1 protein

UTR : Untranslated Region

Vir : Virilizer

VIRMA : VIR-like m^6A Methyltransferase Associated

wg : wingless

WT : Wild Type

WTAP : Wilms-Tumour Associated Protein

XIST : X Inactive Specific Transcript

XPO2 : eXPOrtin 2

XRN1: Exoribonuclease 1

YTHDC1,2 : YTH Domain-Containing protein 1,2

YTHDF1,2 ,3 : YTH Domain-containing Family protein 1,2,3

ZC3H13 : Zinc finger CCCH-type containing protein 13

Ψ : pseudouridine

SUMMARY

INTRODUCTION 19
Chapter 1: mRNA modifications and the role of PCIF1 at the 5′ cap 20
1. mRNA modifications and their functional relevance 20
1.1. mRNA modifications 21
1.1.1. The first identified RNA modification: Pseudouridine (Ψ) 22
1.1.2. N1-methyladenosine (m^1A) 23
1.1.3. 2'O-methylation 24
1.1.4. 5-methylcytosine (m^5C) 25
1.1.5. The most abundant mRNA modification: N^6-methyladenosine (m^6A) 26
1.1.6. N^6, 2′O-dimethyladenosine (m^6Am) 27
1.2. Writers of m^6A modification 28
1.2.1. The methyl transferase complex METTL3/14: the m^6A mRNA writer 28
1.2.2. METTL16: a writer of structured RNA substrates 30
1.2.3. CAPAM or PCIF1: the cap-specific methyltransferase protein 31
1.3. Readers of m^6A modification 32
1.3.1 The YTH domain-containing readers: the YTHDF1-3 and YTHDC1-2 proteins 33
1.3.2. A minor set of m^6A readers other than YTH family 34
1.4. Erasers of m^6A modification 35
1.4.1. The Fat mass and obesity-associated protein FTO is a m^6A eraser 35
1.4.2. ALKBH5 demethylase operates around m^6A-enriched stop codon 36
1.4.3. The conserved FTO and ALKBH5 erasers contribute to normal human metabolism and physiology 37
1.5. *Drosophila melanogaster* RNA writers, readers and erasers 38
1.5. mRNA writers in *Drosophila* 39
1.5.1. The main m^6A writer in *Drosophila* is the Ime4/Mettl14 complex 39
1.5.2. The WTAP orthologue in fly: the Female-lethal (2)d component (Fl(2)d) 40
1.5.3. The Fl(2)d/associated complex component or Xiong (Flacc/Xiong) 42
1.5.4. SPOC-protein family: the Spenito component (Nito) 42
1.5.5. The Virilizer component (Vir) assembles with the methyltransferase complex 43
1.5.6. PCIF1 orthologue in *Drosophila* (Pcif1) is catalytically inactive 44
1.6. Readers in *Drosophila* model 45
1.6.1. The YT521-B (CG12076) is the closest orthologue of the human YTHDC1 45
1.6.2. The mammalian YTHDF2 is encoded by the CG6422 gene in fly 45
1.6.3. The Benign gonial cell neoplasm (Bgcn) reader is mostly expressed in fly germ cells 46
Chapter 2: Canonical versus alternative cap structures and decapping enzymes: the Nudix family 49
2. A variety of caps and the cap binding proteins 49
2.1. The canonical 7-methylguanine cap (m^7G-cap) 51
2.2. The cap binding proteins (CBP) 52
2.2.1. The cap binding complex (CBC) is conserved in eukaryotes 53

2.2.2. The eIF4F complex binds to the 5′ cap of eukaryotic mRNA 55
2.3. The NUDIX family 57
2.3.1. The canonical Decapping enzyme 2 (DCP2) or NUDT20 57
2.3.2. The *mth1, mth2* and *mth3* homologues encode for NUDT1, NUDT15 and NUDT18 59
2.3.3. NUDT2 hydrolases the diadenosine tetraphosphate (Ap_4A) metabolite 61
2.3.4. NUDT3, NUDT4, NUDT10 and NUDT11 are DiphosphoInositol PyroPhosphatases 62
(DIPPs) 62
2.3.5. NUDT5 is involved in chromatin remodelling 64
2.3.6. NUDT7, NUDT8 and NUDT19 hydrolyse diphospho-CoenzymeA substrates 64
2.3.7. The conserved NUDT12 hydrolyses NAD-capped RNAs 68
2.3.8. NUDT16 hydrolyses NAD and FAD substrates 70
2.3.9. The NUDT6, NUDT9, NUDT13, NUDT14 and NUDT17 Nudix members 72
2.3.10. NUDT21 hydrolase is the Cleavage and Polyadenylation Specificity Factor 5 (CPSF5) that contributes to the removal of the poly(A) tail 75
OBJECTIVES OF MY THESIS 80
RESULTS 81
Chapter 1: Catalytic dead Pcif1 regulates gene expression and fertility in 82
Drosophila 82
1.1.Presentation of the two papers included in this manuscript (Pandey et al.,Cell Reports and Franco et al., accepted for publication on RNA journal) 82
1.2. The catalytic-dead Pcif1 regulates gene expression and fertility in *Drosphila*..**Erreur ! Signet non défini.**
Chapter 2: Nudt19B is required for fly fertility 120
2. A genetic screen for the identification of new decapping enzymes and alternative cap structures in *Drosophila melanogaster* 120
2.1. *Nudt20 (Dcp2)* silencing affects normal cell differentiation in fly photoreceptors 121
2.3. Nudt19A is required for normal fly locomotion 122
2.4. Nudt19B depletion affects fly fertility 124
2.5. Generation of CRISPR-Cas9 mutant to characterize the *Nudt19A* and *Nudt19B* genes in *Drosophila* 128
2.6. Cloning *Nudt19A*, *Nudt19B* and *Nudt19C* for recombinant protein expression to test 130
dpCoA-decapping activity *in vitro* and for generating transgenic flies 130
SUPPLEMENTAL FIGURES RESULTS-Chapter 2 133
DISCUSSION AND PERSPECTIVES 137
Chapter 1. Catalytic dead Pcif1 has a role in fertility and gene expression in *Drosophila* 138
1.1. Putative mechanism of action of the catalytic dead Pcif1 in gene regulation 138
1.2. Novel high-throughput technologies may unveil Pcif1 binding-enriched genes 139
1.3.Proximity labelling assay may identify Pcif1 interacting proteins in gene transcription regulation 139
Chapter 2. Functional role of Nudt19 proteins in fly physiology 142
2.1. The *Nudt20 (Dcp2)* gene in *Drosophila* is required for proper photoreceptor shape 142
2.2. The *Nudt19A* may be required in post-mitotic neurons 143
2.3. *Nudt19B* is required for full fertility in fly 143
2.4. CRISPR-Cas9 knock out mutants may validate the fertility phenotype 145

2.5. Biochemical assays to test dpCoA activity *in vitro* and *in vivo* 145
MATERIALS AND METHODS 148
Chapter 2 149
2.1. RNAi silencing of the *Nudix* family genes in *Drosophila melanogaster* 151
2.2. Phenotypic analysis of flies carrying *Nudix* gene silencing 151
2.2.1. Eye phenotype 151
2.2.2. Fertility analysis 151
2.2.3. Locomotion analysis 152
2.2.4. RT-qPCR 152
2.3. Generation of flies carrying a P element insertion in one of the *Nudt19* genes and a chromosomal deletion covering the *Nudt19* locus 153
2.4. Cloning of *Nudt19A*, *Nudt19B* and *Nudt19C* genes for protein purification 153
2.5. Cloning of *Nudt19A, Nudt19B* and *Nudt19C* sequences in pBS vector 153
2.5. Site directed mutagenesis for generating catalytically-dead Nudt19 fly proteins 154
2.6. Cloning of *Nudt19A*, *Nudt19B* and *Nudt19C* sequences in pETM11-His6-Strep-SUMO TEV vector (pETM11) 155
2.7. Cloning of *Nudt19A, Nudt19B* and *Nudt19C* sequences in p[PUAST] vector for fly transgenesis 156
2.8. Generation of *Nudt19* CRISPR-Cas9 mutant flies 157
REFERENCES 159
ANNEX 181
The Mammalian Cap-Specific m^6Am RNA Methyltransferase PCIF1 Regulates Transcript Levels in Mouse Tissues (Pandey et al., 2020 Cell Reports) 181

LIST OF FIGURES

INTRODUCTION

Figure 1. Modifications on mRNA occur at 5′ end and at the 3′ end and at internal sites of the molecule. 21
Figure 2. Pseuduridine (Ψ) is generated from isomerization of uridine. 22
Figure 3. The N1-methyladenosine (m^1A) modification. 23
Figure 4. The 2'O-methylation mostly occurs at the 5' end of mRNAs. 24
Figure 5. 5-methylcitosine modification (m^5C) preferentially occurs downstream the translation initiation sites. 25
Figure 6. N6-methyladenosine (m^6A) occurs at internal site and close to 3' UTR region. 26
Figure 7. N6-dymthyladenosine (m^6Am) occurs at the TSS on mRNAs. 27
Figure 8. The methyltransferase complex. 29
Figure 9. The mechanism of action of writers, readers and erasers of the m^6A modification. 32
Figure 10. Sxl mRNA maturation in *Drosophila*. 40
Figure 11. Sine oculis regulatory pathway in eye development in *Drosophila*. 41
Figure 12. The *Drosophila* Pcif1 protein structure is conserved with its mammalian orthologue. 44
Figure 13. Balance between germ stem cells self-renewal and differentiation in *Drosophila* ovaries. 46
Figure 14. Regulation of Tut, Bam, Bgcn, on their target Mei-P26 during germ cells differentiation. 47
Figure 15. The mammalian Nudix family. 50
Figure 16. Mechanism of Pumilio 2 repression of mRNA translation in *Xenopus* oocytes. 52
Figure 17. The CBC binds to the m^7G cap and to mRNA by the RRM motifs. 53
Figure 18. Mechanism of translation initiation mediated by eIF4F complex. 56
Figure 19. Schematic representation of the Dcp2 protein in *S. cerevisiae* and its involvement in different mRNA decay pathways. 57
Figure 20. The three MTH proteins share a common MuT box. 59
Figure 21. The insertion of 8-oxo-guanine in the DNA sequence generates a mismatch. 59
Figure 22. The diadenosine tetraphosphate (Ap_4A) metabolite. 61
Figure 23. The diphosphoinositol pentakisphosphate PP-InsP5 molecule. 62
Figure 24. NUDT7, 8 and 19 hydrolases share a common CoA box. 65
Figure 25. The mammalian NUDT7 hydrolyses the diphosphate bond within the CoA when attached to mRNA *in vitro*. 66
Figure 26. NUDT12 hydrolyses the diphosphate bond within the NAD-capped RNAs. 68
Figure 27. The FAD capped mRNA is removed by the NUDT16 hydrolyse. 71
Figure 28. The ADP-ribose is hydrolysed by NUDT9 *in vivo*. 73
Figure 29. NUDT14 hydrolases UDP-glucose. 74
Figure 30. Polyadenylation at the 3′end of mRNAs. 76
Figure 31. Fly cleavage and polyadenylation complex shares high similarity to the human one. 77
Figure 32. 105130 RNAi silencing of *Nudt20 (Dcp2)* in the photoreceptors results in ‘rough eye’ phenotype. 122
Figure 33. Fly locomotion is affected upon *Nudt19A* silencing in post-mitotic neurons. 123
Figure 34. Female fertility decreases upon *Nudt7, Nudt19A* or *Nudt19B* silencing in *Drosophila* ovaries. 124
Figure 35. The *Nudt19A*, *Nudt19B* and *Nudt19C* genes localize very close to each other. 125
Figure 36. *Nudt19B* trans-heterozygotes display reduced fertility in *Drosophila*. 127
Figure 37. Generation of CRISPR-Cas9 *Nudt19* knock out flies. 129

Figure 38. The predicted Glutamate catalytic residues are mutated in two Alanine residues. 130
Figure 39. The pETM11SS-His6-Strep-SUMO-Tev vector contains the SUMO sequence and the His6 and Streptavidin tags. 131
Figure 40. The pP[UAST] vector is used for generating transgenic flies. 132
Figure 41. Protein labelling assay to identify putative Pcif1 interactors. 140
Figure 42. The NAD-capture sequencing. 146
SUPPLEMENTAL FIGURES
Suppl. Figure 1. RNAi silencing of *Nudt20 (Dcp2)* and *Nudt21(Cpsf5)* genes in *Drosophila* neurons affect fly locomotion. 133
Suppl. Figure 2. RNAi silencing of *Nudt20 (Dcp2)* and *Nudt21 (Cpsf5)* genes in *Drosophila* ovaries affect female fertility. 134
Suppl. Figure 3. The Nudt19B protein shares the highest identity to the human NUDT19 orthologue. 135
Suppl. Figure 4. *Nudt19A* and *Nudt19B* trans-heterozygous and heterozygous mutants are viable. 135
Suppl. Figure 5. RNAi silencing of the *Nudt19C* gene in *Drosophila* ovaries does not affect female fertility. ..136
TABLES
Table 1. The 12 Nudix orthologues in *Drosophila melanogaster*. 121
Table 2. Primers for PCR 149
Table 3. Primers for the Site directed mutagenesis 150
Table 4. Reagent or resource 150

INTRODUCTION

Chapter 1: mRNA modifications and the role of PCIF1 at the 5′ cap

1. mRNA modifications and their functional relevance

The central dogma of biology provides a basic explanation of the genetic information workflow from a DNA sequence to a final cellular product, the protein. The multi-step process, known as gene expression starts with the transcription of DNA in a molecule of RNA, then translated in an amino acidic sequence. Since Francis Crick and colleagues stated the dogma for the first time, a huge number of discoveries have clarified that the RNA intermediate molecule plays a crucial role in gene expression regulation. More recently, specialists of the field have coined the word 'epitranscriptome' to describe all set of chemical modifications occurring on RNA molecules without affecting the sequence itself. Around 150 structurally distinct modifications have been detected in messenger (protein-coding) RNAs (mRNAs) but also in noncoding RNAs such as tRNAs, rRNAs, snRNAs, snoRNAs and lncRNAs[1]. A huge number of transformations include the 5′ end capping, the 3′end poly(A) tailing together with a set of various kind of modifications on internal sites. The deposition of those modifications is driven by three groups of enzymes: the writers that physically change nucleosides, the readers that recognize the tag along the sequence and the erasers that remove the modification. Most of them are conserved across all three domains of life (Archea, Bacteria and Eukarya) and their dysfunction is associated to the development of diseases and particularly of cancer[2].

The first marks were detected by chromatography and organic chemistry, but the lack of precise and sensitive instruments prevented scientists to go further. The evolution of sophisticated technologies for RNA sequencing have given the chance to identify the characterization of a multitude of modifications. The detection of RNA modifications dates back to the early 1950s, just like that of DNA modifications, when molecular biology developed. The first RNA modification discovered was pseudouridine (Ψ), which is actually the most frequent post-translational modification occurring on uridine and often considered as the fifth nucleoside. It was detected in 1951 by Cohen's group, two years after the discovery of the first DNA modification, 5-methylcytosine (m^5C). The switch from chromatography to gel electrophoresis and mass spectrometry approaches highlighted that tRNAs and mRNAs are the richest source of modifications[3]. Since those marks change the structure and the chemical properties of RNA molecules, some of them are thought to optimise translation events in the cell, whereas others seem to be neutral regarding translational but may have other functions that stay to be discovered. In particular, internal tags inserted on mRNAs and tRNAs can affect codon-anticodon recognition and their reversibility suggests a new, essentially unexplored layer of control of gene expression. Therefore, the importance of understanding the role of RNA

modifications could help in unveiling how they affect cellular metabolism in physiological and pathological scenarios[4].

I will further describe in the first chapter of the introduction to my PhD thesis work, the most relevant modifications described for mRNAs.

1.1. mRNA modifications

mRNA transfers the genetic information from DNA to a protein. Modifications occurring on mRNAs respect high fidelity to the genetic content and contribute to translation success. High-throughput technologies have shown that mRNA modifications are distributed both internal and at the 5′ and 3′ ends of the mRNA, all contributing to mRNA stability, metabolism, translation efficiency or to mRNA decay. The detected modifications on mRNAs include m^7G, m^1A, m^5C, Ψ, together with m^6A and m^6Am; they collectively form the epitranscriptome and code a new layer of information that controls protein synthesis (Figure 1).

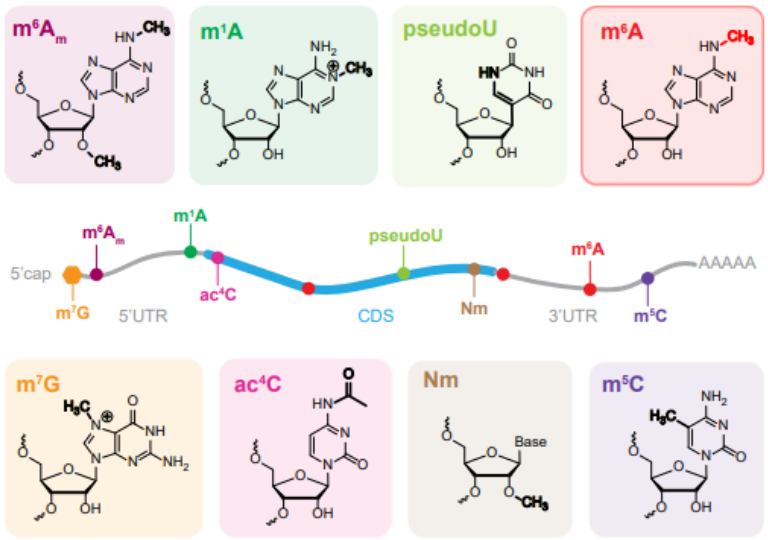


Figure 1. Modifications on mRNA occur at 5′ end and at the 3′ end and at internal sites of the molecule. Schematic representation of the most studied mRNA modifications. Methylation occurs at different positions and on various nucleosides, all along the sequence, including the 5′ and 3'UTRs. Ψ is present but less abundant. 5′-cap at the 5′UTR and the poly(A) tail at the 3′UTR are also shown in the scheme (adapted from Schubert, 2019 Produced by the Science/AAAS Custom Publishing Office)[5].

In addition to internal mRNA marks, RNA polymerase II (RNA Pol II)-transcribed RNAs are modified by the linkage of a 7-methylguanosine (m^7G) to the 5′ triphosphate found on the first transcribed nucleotide forming the 5′cap structure m7GpppN (where N is any nucleotide). This is a conserved co-transcriptional modification which ensures 5′-3′ exonuclease protection and the efficient pre-mRNA maturation, export and translation[6,7]. The mechanism of canonical and non-canonical caps will be described in details in Chapter 2.

The poly(A) tails are arrays of 100-250 adenosine added at the 3′ end of mRNAs in a co- or post-transcriptional manner and notably mediates the half-life of transcripts; however different tail lengths show specific properties. The details of poly(A) formation and removal will be presented in Chapter 2. Here, I will focus on the discussion of the main modifications distributed along mRNAs.

1.1.1. The first identified RNA modification: Pseudouridine (Ψ)

Pseudouridine (Ψ) is the first modification identified on RNAs and it is mostly present in rRNAs and tRNAs resulting in secondary structure changes. It is a conserved RNA modification across species but very few is known about Ψ modified mRNAs[8]. Ψ is generated by the isomerization of uridine by the Ψ synthase family of enzymes encoded by the Pseudouridine Synthase (PUS) genes, eight of which are well conserved from yeast to mammals (Figure 2).

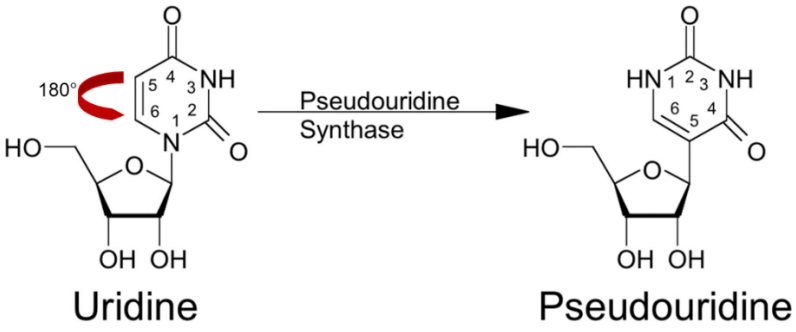


Figure 2. Pseuduridine (Ψ) is generated from isomerization of uridine.

The formation of Ψ is due to the isomerization of the uridine catalysed by the Pseudouridine synthase (PUS). The number of C atoms and N atoms is kept identical (Adapted from Addepalli and Limbach)[9].

PUS members are shown to deposit the mark on different types of RNAs according to target-specific secondary structures rather than a specific consensus sequence[10]. Moreover, defective PUS enzymes have been associated with increased bacteria sensitivity to antibiotics[11], growth failure in *S. cerevisiae*, and human disorders such as mitochondrial myopathy and sideroblastic anemia[12].

In eukaryotic mRNAs, pseudouridine modification shows dynamic changes in response to several factors including stress and cell growth state. For example, in budding yeast, the level of pseudouridylation increases during the diauxic growth (growth induced by the presence of glucose and lactose in the media) inducing mRNA stability[13]. In HeLa cells, pseudouridylation levels of ATPase subunit 5E (ATP5E), Ribosomal Protein L19 (RPL19) mRNAs targets, Metastasis Associated Lung Adenocarcinoma Transcript 1 (MALAT-1) and RNA Component of 7SK Nuclear Ribonucleoprotein (RN7SK) ncRNAs significantly decrease upon serum

starvation suggesting that environmental changes can affect RNA Ψ levels influencing the expression of genes involved in key physiological pathways in the cell[13]. Recently, Ψ modification on mRNA has been implicated in the production of mRNA vaccine for fighting the COVID-19 wide spreading[14]. The presence of Ψ sites on the synthetic mRNA reduced the immunogenicity response because of its stabilization of mRNA coding the spike protein in humans that subsequently leads to the stimulation of a controlled immune system activation[14].

1.1.2. N1-methyladenosine (m^1A)

N1-methyladenosine (m^1A) was first identified on rRNAs and tRNAs because of its low density on mRNAs. The donor molecule is S-Adenosyl Methionine (SAM), which requires the hydrolysis of 13 molecules of ATP, highlighting the significance of that modification in cellular functions[2]. m^1A immunoprecipitation coupled to mRNA sequencing analysis indicated that it usually maps close to the translation start site upstream of the first splice site specifically enriched in the 5′UTR of mRNA and promotes RNA translation[15] (Figure 3).

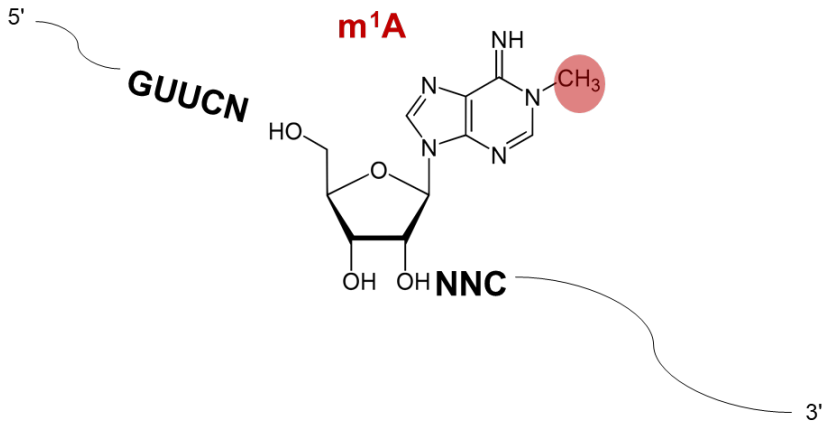


Figure 3. The N1-methyladenosine (m^1A) modification.

The m^1A modification occurs at the 5' UTR of mRNAs within the GUUCNANNC consensus sequence (in bold). The red circle indicate the methyl group. N stands for any nucleotide.

m^1A modification on mRNAs occurs at the consensus sequence GUUCNANNC (Figure 3) as well as on tRNAs and it is driven by the same set of TRMT6/61-complex writers and recognised by the same set of YTHDF1, YTHDF2, YTHDF3, and YTHDC1 readers[16,17]. ALKBH3 and ALKBH1 are the erasers of the m^1A modification on tRNAs, rRNAs and mRNAs, which are essential for balancing the m^1A turnover and the regulation of gene expression. Many evidences have reported that m^1A imbalance is related to the insurgence of various diseases, such as Hodgkin lymphoma[18], hepatocellular carcinoma[19] and oral squamous cancer[20], revealing that m^1A-related regulatory genes may be used as diagnostic or therapeutic targets for these malignancies.

1.1.3. 2'O-methylation

The 2'O-methylation modification has been identified on miRNAs and snoRNAs, but also in mRNAs and it is a conserved mark from viruses to plants and humans. The donor molecule is SAM and the enzymes involved in the deposition of the mark on the adjacent nucleotides to the cap are the so called Cap Methyltransferases (CMTRs)[21,22] (Figure 4).

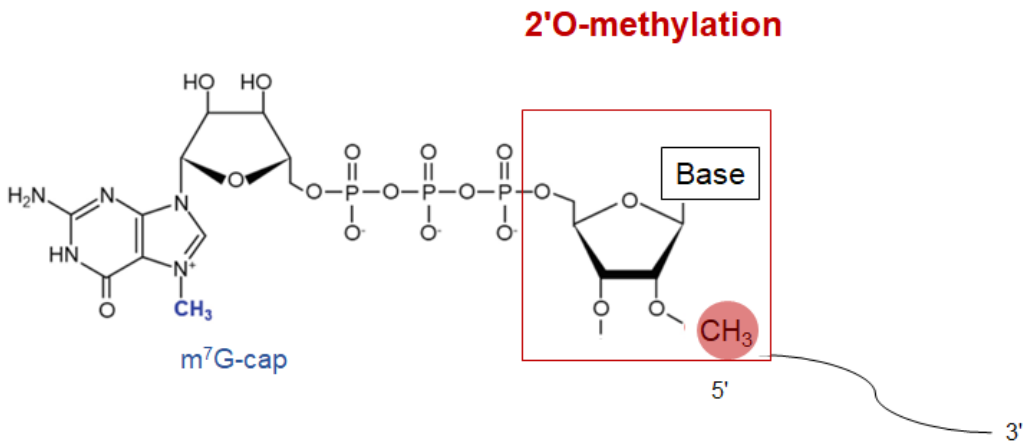


Figure 4. The 2'O-methylation mostly occurs at the 5' end of mRNAs.

The methylation on the second carbon of the ribose mostly occurs on the nucleotides adjacent to the m^7G-cap on mRNAs. The methyl group is indicated with a red circle.

Most of animals such as *Caenorhabditis elegans, Drosophila melanogaster* and mammals have two encoding genes *cmtr1* and *cmtr2*, while *Trypanosoma cruzi* has three[23]. The 2'O-methylation on the first transcribed nucleotide is catalysed by CMTR1 while CMTR2 deposits the mark on the second nucleotide. In trypanosomes, CMTR2 and CMTR3 are able to methylate the third and the fourth nucleotides close to the 5′cap, showing a redundant activity that has been identify also for CMTR2 in *Drosophila*[24,25]. Recently, it has been shown that RNA and DNA viruses use the 5′cap structure containing 2'O-methylated nucleoside to mimic the host cap-structure and have their own *cmtr* gene[26,27]. The host translation machinery may start the translational process, allowing the virus to prevent its detection as foreign RNA and escape the immune response. Indeed, it was demonstrated by Daffis and colleagues that the 2′O-methylation is evolutionary conserved and allows cellular recognition of self from non-self RNA on models of West Nile virus infection of mice fibroblasts[28]. Similarly, it was observed in coronavirus by Case et al. that when the SAM-binding region of the N7-methyltransferase of coronavirus is mutated, there is a loss of viral replication and enhanced sensitivity to immune response. These results indicate that viruses may have adapted 2′O-methyltransferase activity as a mechanism of evading host recognition systems[29].

It has also been demonstrated the presence of Nm modification on internal sites of mRNA as well as in tRNAs, rRNAs, and on Ago2 bound siRNAs and miRNAs in *Drosophila.* This modification can contribute to increase hydrophobicity, structural stability and RNA-RNA or RNA-protein interactions[30–32]. However, despite its potential source for understanding diseases, Nm functional role requires further investigation.

1.1.4. 5-methylcytosine (m^5C)

The 5-methylcytosine (m^5C) is a well-studied modification on DNA that has been detected also in rRNAs, tRNAs[32,33] and mRNAs[34]. Bisulphite treatment of RNAs purified from *Drosophila* embryos has highlighted m^5C hotspots in close proximity to untranslated regions and at the binding sites for the Argonaute component of the RISC silencing complex Ago[33,35]. Later on, the wide-spread of high-throughput technologies and next generation sequencing including m^5C-RNA immunoprecipitation (RIP)[34], 5-azacytidine-mediated RNA immunoprecipitation (Aza-IP)[36] and methylation-individual-nucleotide-resolution crosslinking and immunoprecipitation (miCLIP)[37], allowed to detect m^5C modification on mRNAs of different species and tissues. In HeLa cells and in mouse model, m^5C enrichments are present in CG rich sequences and downstream of translation initiation sites (Figure 5).

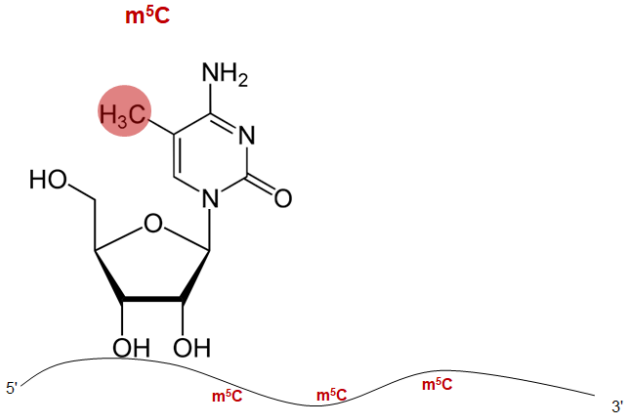


Figure 5. 5-methylcitosine modification (m^5C) preferentially occurs downstream the translation initiation sites.

The m^5C occurs downstream the translation start site in mRNA. However, internal sites can be modified as well. The methyl group is indicated with a red circle.

The tRNA methyltransferase NSUN2 was shown to methylate cytosine in mRNAs and non-coding RNAs in addition to tRNAs[38]. NSUN2 depletion is associated with mental retardation disorders in humans while in yeast, flies, and mice it impairs cellular differentiation pathways in skin, testes, and brain[37,39,40]. The m^5C seems to promote transcripts shuffling from the nucleus to the cytoplasm, by recruiting the Aly-REF carrier protein called ALYREF [38]. This protein localises at the nuclear speckles and shares amino acid residues involved in reading the m^6A

modification with the YTH family indicating that ALYREF is the main m^5C reader of mRNAs[38]. The ALYREF shuttling together with the CMTR mediated mRNA export may play a role in balancing the level of gene expression depending on the species and the tissue considered.

1.1.5. The most abundant mRNA modification: N^6-methyladenosine (m^6A)

The presence of N^6-methyladenosine (m^6A) was first detected in the 1970s when methods for investigating the prevalence of m^6A were laborious and poorly sensitive. Recently, a renewed interest in m^6A modification has emerged, showing that it is the most abundant modification in eukaryotic mRNAs and its conserved at internal sites together with 3′UTRs[41–43] (Figure 6).

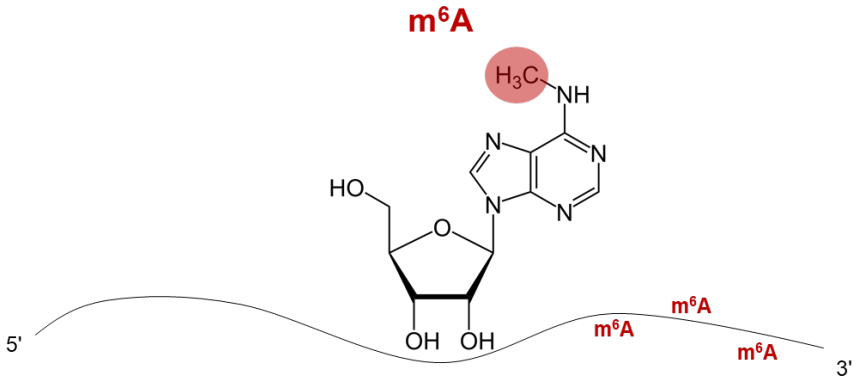


Figure 6. N6-methyladenosine (m^6A) occurs at internal site and close to 3' UTR region.
The m^6A modification is mostly distributed along mRNA internal sites and in proximity of the 3' UTR region. The methyl group is indicated with a red circle.

m^6A modification has been detected on other RNAs species such as such as tRNAs, rRNAs, circRNAs, miRNAs, and lncRNAs. The ancient techniques for identifying m^6A required incubating cells with ^{14}C-radiolabeled methionine (the precursor for the endogenous methyl donor SAM), following which the incorporation of methyl groups into RNAs could be quantified by radioactivity detection. The detection of m^6A modification on mRNAs was challenging at least for two reasons: first, it required a pre-ordained focus on a particular transcript rather than a global approach that could detect sites of adenosine methylation in all mRNAs[44–46]. Second, m^6A modification (unlike m^1A) is invisible to sequencing methods, since both methylated and unmethylated adenosines base pair with thymine or uridine and both are reverse transcribed as thymine, further endearing the study of m^6A role.
More recently, the distribution of m^6A modification has been demonstrated by NGS sequencing which showed on average three to five modifications on one third of the mammalian transcripts. However the accurate position of m^6A modifications have been reached by meRIP (methylated RIP) through the implication of antibodies directed to the mark followed by sequencing[42,47].

The m^{6}A modification is reversibly deposited on mRNAs due to the presence of enzymes involved in the deposition, recognition and removal influencing transcript maturation, stability and translation efficiency[48,49]. Moreover, it has been widely accepted that altered function of m^{6}A effectors play a role in the progression to pathological conditions that will be presented further in Chapter 1.

1.1.6. N^6, 2′O-dimethyladenosine (m^{6}Am)

A peculiar modification occurring in vertebrate mRNAs is N^6, 2′O-dimethyladenosine (m^{6}Am). It was first identified in 1975[50] at the 5' termini of mRNAs and is ten to fifteen times less abundant than m^{6}A as quantified in several recent studies in HeLa, HEK293T,3T3-L1, MEL624 cells, or brain tissues[51–53]. This modification resides one nucleotide after the 7-methylguanosine cap, if the first transcribed nucleotide is a 2′O-methyladenosine and it is co-transcriptionally inserted by the newly described cap-specific adenosine methyltransferase (CAPAM) also known as the phosphorylated C-terminal domain interacting factor 1 (PCIF1) (Figure 7).

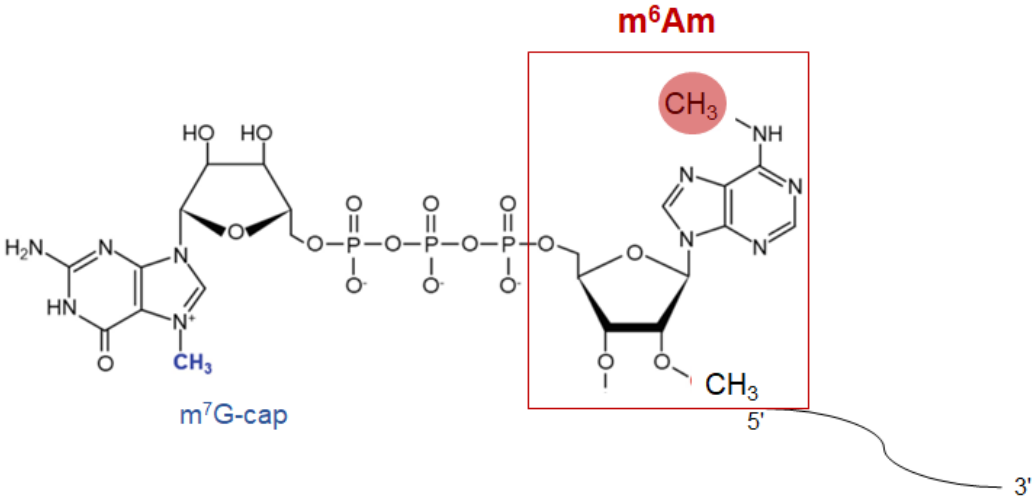


Figure 7. N6-dymthyladenosine (m^{6}Am) occurs at the TSS on mRNAs.

The first transcribed nucleotide is 2'O-methylated as soon as the m^{7}G cap is linked by an inverted 5'-5' triphosphate bond. Then, if the TSS is an adenosine, it is further methylated in position 6 of the adenosine ring, forming the m^{6}Am modification.

We will refer to it as PCIF1 all along the manuscript.

The role of m^{6}Am is still under debate. By instance, two independent works have reported opposing effect of this modification on translation efficiency. Ackichika et al. work supports the idea that the m^{6}Am modification allows capped mRNAs to be efficiently translated[54], while the work of Sendic et al., rather indicates that m^{6}Am negatively influences mRNA translation[52]. The former demonstrated that m^{6}Am modification does not stabilize mRNAs upon PCIF1 KO compared to PCIF1 WT cells. However, contradictory results point out the various possible

outcome depending on mRNAs species and highlights the necessity of additional studies to clarify m^6Am molecular role and PCIF1 physiological functions.

1.2. Writers of m^6A modification

The methyl group at the N6 position of m^6A is inserted within a RRACH motif (R represents purine and H is either A, C or U) and plays a key regulatory role in various biological events, including meiosis, cell differentiation, neuronal function, cancer proliferation, circadian rhythm, sex determination and chromosomal silencing[54]. It is mainly enriched within exons and close to the 3′UTR, peaking over the stop codon. This distribution suggests a functional coupling between modification and other RNA processing steps, such as maturation of pre-mRNAs, alternative splicing and translation events. The m^6A mark is installed co-transcriptionally on mRNAs by a complex composed of multiple subunits including the main core formed by the Methyltransferase-Like 3 (METTL3) and the Methyltransferase-Like 14 (METTL14)[55,56].

Nonetheless, the methylation process is supported by various RNA Binding Proteins (RBPs) that I will present in details below.

1.2.1. The methyl transferase complex METTL3/14: the m^6A mRNA writer

The METTL3/14 complex is formed by two subunits, the Methyltransferase-Like 3 (METTL3) and the Methyltransferase-Like 14 (METTL14), the former possesses the catalytic subunit and the latter acts as an essential component to facilitate RNA binding[55,56]. The complex is able to deposit a CH_3 group on the adenosine through the recognition of the RRACH consensus sequence, with the central A as the methyl acceptor.

Further studies highlighted the presence of the Wilms-Tumour Associated Protein (WTAP) as an additional component of the methyltransferase complex that is capable of optimising substrate recruitment and precise localization of METTL3/14 on mRNA targets[57,58]. Additional subunits are critical for the specific activities of the methyltransferase machinery. The Vir-like m^6A Methyltransferase Associated (VIRMA) is crucial for 3'UTR site recognition[59] while the Zinc finger CCCH-type containing 13 proteins (ZC3H13) guarantees nuclear localization of the writer complex[60]. The RNA Binding Motif protein 15/15B (RBM15/15B) binds U-rich regions and may facilitate the methylation of a specific set of RNAs such as the long non-coding XIST RNA[61] (Figure 8).

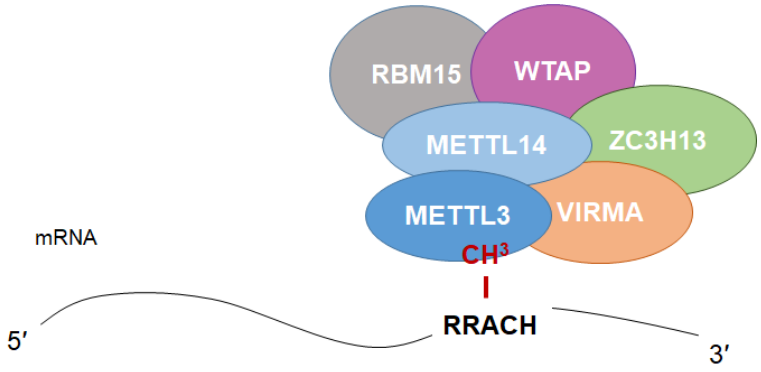


Figure 8. The methyltransferase complex.

The methyltransferase complex is composed by the METTL3 protein, which binds to the mRNA at the RRACH consensus sequence in association with METTL14. WTAP contributes to the recognition of m^{6}A sites and VIRMA component is required for 3'UTR site recognition. ZC3H13 localizes the complex in the nucleus. RBM15/15B is part of the m^{6}A writer complex and recognizes the U-rich regions.

The localization of each subunits of the multi-complex machinery changes depending on external signals and on the cell type considered. METTL3 is mostly associated with METTL14 in the nucleus, in order to elicit its methyltransferase activity[42,47].

For example, in the nucleus of mouse Embryonic Stem Cells (ESCs), METTL3/14 plays a crucial role in response to heat shock stress together with Microprocessor components (MP): MP binds to *Hsp70* gene, recruiting METTL13/14 and promoting translation of this class of chaperones[62]. Interestingly, METTL3 is also recruited in the proximity of UV light-induced DNA damage sites on chromatin in U2OS and HeLa cells. Indeed, METTL3 localizes in the nucleus 2 minutes after UV irradiation inducing transient m^{6}A modification of mRNA and subsequent fast re-initiation of transcription[63]. In this condition, METTL3 recruits also the Polκ polymerase that is involved in Nucleotide Excision Repair (NER) system to damaged sites. These observations suggest that this factor mediates Polκ recruitment to DNA damage sites through m^{6}A deposition on mRNAs coding for DNA repair proteins.

Nevertheless, METTL3/14 has also been found in cytosol of HeLa cells and lung cancer cells[64]. In lung cancer, METTL3 seems to play a role as a reader in the cytosol, where it interacts with the translation factor eukaryotic Initiation Factor H (eIFH), and the polyribosomes, then promoting oncogenes translation. Immunohistochemistry (IHC) of primary human lung adenocarcinoma samples and adjacent normal control tissue revealed that METTL3 expression is significantly increased in lung cancer, as well as Epidermal Growth Factor Receptor (EGFR) and BRomoDomain-containing protein 4 (BRD4) oncogenes[64]. It is reasonable to hypothesize that METTL3/14 is mostly nuclear, in order to transfer methyl groups on nascent mRNAs, while under stressful condition it can move to the cytosol to interact directly with the translation apparatus.

As just described above, the METTL3/14 methyltransferase complex is essential for the deposition of internal m^6A in human cells and many homologues have been identified in plants, yeasts, flies and zebrafish indicated an evolutionary conserved function among eukaryotes. In mouse embryos, this complex notably regulates the switch from pluripotent stem cells to 'primed' stem cells. By knocking out METTL3 and METTL14, pluripotent stem cells do not undergo differentiation, while knockdown of only METTL3 in primed stem cells drives them to cell cycle arrest and apoptosis suggesting that the two subunits may have the ability to function independently to each other in certain circumstances[65].

METTL3/14 role is also elicited in various processes during mouse and human neurogenesis. Indeed, METTL14 KO mouse cerebrum and human cerebral organoids showed protracted cell cycle progression in neural progenitors and extended cortical neurogenesis into postnatal stages[66]. A sever defect during embryonic hematopoietic cell development has been observed in zebrafish where METTL3 deletion causes a strong delay in haematopoietic cell differentiation[67]. In the same study, the authors showed similar phenotype in mice issuing that METTL3 role in embryonic haematopoiesis is conserved also in mammals. Thus, it is evident that METTL3/14 complex plays a key role during various processes of eukaryotic development, by methylation of transcripts directly involved in embryogenesis.

Furthermore, WTAP is responsible for localising the methyltransferase machinery to the nuclear speckles as shown by KO experiment in zebrafish : the lack of WTAP notably induces smaller head and eyes, smaller brain ventricle, curved notochord and reduced muscle size compared to wild type siblings[57]. GO analysis following RNA-seq of the WTAP- and METTL3-deficient cells demonstrated that most of these genes were related to transcription and RNA processing, further emphasizing the potential significance of WTAP and METTL3 in RNA metabolism. So, alterations in one of the components reduce the overall efficiency of the complex.

In *Drosophila melanogaster*, the orthologue of METTL3, called Ime4 (inducer of meiosis 4) and Mettl14 are conserved and will be discussed in details in the following sections.

1.2.2. METTL16: a writer of structured RNA substrates

The Methyltransferase-Like 16 (METTL16) deposits m^6A on structured RNAs carrying a specific nonamer sequence UACAGAGAA. It is known that it works on two different structure contexts: the U6 snRNAs and the vertebrate conserved hairpin 1 (hp1) in human Methionine Adenosyl-Transferase 2A (MAT2A) 3´UTR[68]. MAT2A gene encodes for the SAM synthase, which regulates cellular concentration of SAM. It is an essential evolutionary conserved donor of methyl groups regulating all the steps of gene expression among the three domains of life.

The way by which the mammals control intracellular SAM concentration is not completely understood, whereas bacteria evolved riboswitches to adapt intracellular SAM levels to new SAM synthesis[68].

In human cells, methylation of MAT2A mRNA 3′UTR hairpin structure is required for proper recognition by the YTHDC1 m^6A reader, which downregulates MAT2A mRNA in high-SAM condition. Under Met starvation, SAM concentration decreases and MAT2A mRNA is stabilized in the cytosol by a prolonged METTL16-MAT2A interaction at the 3′UTR six-hairpin region[69]. Apart to its methyltransferase activity, METTL16 is a splicing factor regulator binding to 3' exon that is retained to promote translation of MAT2A. Then the biological levels of SAM are restored by a negative feedback loop[70,71].

METTL16 methyltransferase has been demonstrated to be essential for mammalian HEK293 and HeLa cells viability as shown by the failure in obtaining knock out cells. This suggests a wide and crucial role of METTL16 in mRNA regulation[68,72]. As reported in two independent studies, METTL16 is involved in the regulation of SAM synthesis feedback loop and in the methylation of U6 snRNA, two key elements for cell survival to methionine starvation and spliceosome efficiency respectively[68,73]. In mouse, depletion of METTL16 leads to impair embryonic development with a blockade at around the 64 blastocysts state and failure of its implantation in the maternal uterus. The general transcriptome is actually heavily dysregulated, so that the early embryo is unfit to implant and prevented to continue the embryogenesis[71]. These extreme phenotypes also argue in favour of a global role in mRNA regulation.

1.2.3. CAPAM or PCIF1: the cap-specific methyltransferase protein

A new writer protein has been identified in 2018/2019 by Akichika and colleagues as well as other research groups, the cap-specific adenosine methyltransferase CAPAM, also known as phosphorylated CTD-interacting factor 1 (PCIF1). PCIF1 inserts a methyl group on 2′O-methyladenosine in position 6 of the adenine ring to give rise to m^6Am, after the 5′cap, in a SAM–dependent manner[52,54,74,75]. PCIF1 specifically recognizes the 7-methylguanosine (m^7G) cap structure and preferentially N6-methylates $m^7GpppAm$ rather than m^7GpppA. It contains an N-terminal WW domain which interacts with the Ser5-phosphorylated C-terminal domain (CTD) of the RNA pol II, suggesting that PCIF1 is recruited to the early elongation complex of RNA Pol II in nascent mRNAs[76–78].

The methyltransferase domain contains the NPPF motif that binds SAM, the donor molecule of methyl groups, whereas the helical domain helps in binding the 7-methylguanosine 5′cap[54]. PCIF1 has been identified in humans, mice, zebrafish and *Drosophila* but its functional role in flies is still not clear and even quite questionable when considering that m^6Am modification is

not present in this organism[79,80]. In mouse, the knocking out of PCIF1 encoding gene results in reduced body weight even though the animals are viable and fertile. Transcriptomic analysis showed that a pool of pseudogenes and predicted genes in mouse testis are dysregulated in PCIF1 KO mice however deeper investigation would help in unveiling putative interactors of PCIF1 that may amplify its spectrum of action.

1.3. Readers of m^{6}A modification

An expanding group of effector proteins, called 'readers' have been identified as effector proteins mediating the functional consequences of transcription (Figure 9).

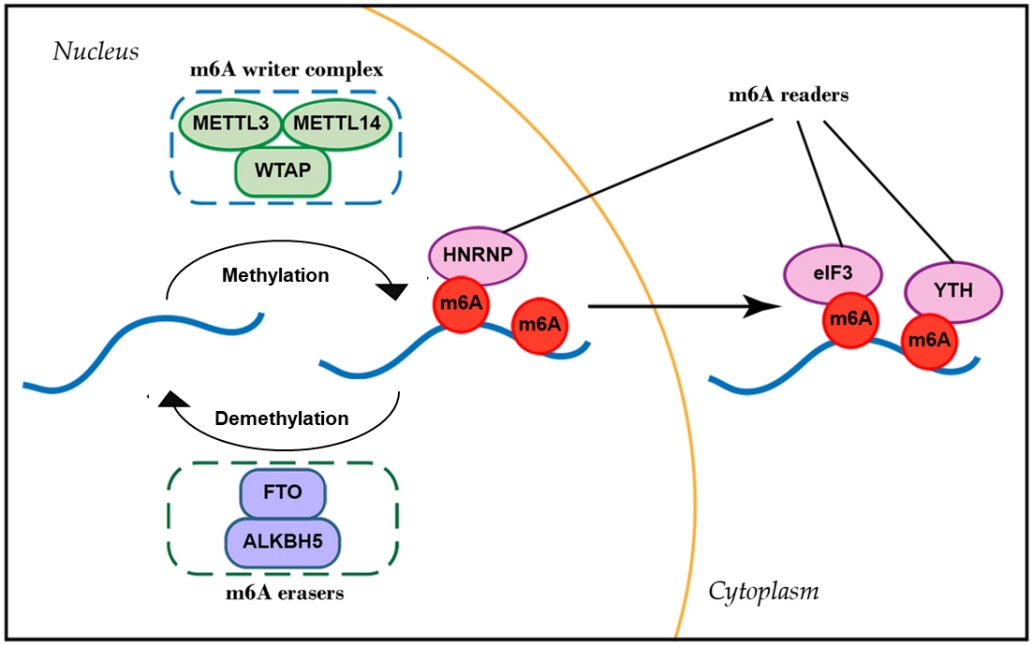


Figure 9. The mechanism of action of writers, readers and erasers of the m^{6}A modification.
The scheme shows a simplified representation of the action of the m^{6}A effectors: the 'writers' deposit the mark, the 'readers' recognize and the 'erasers' physically remove m^{6}A from mRNAs (Adapted from Chen e al.,2019)[48].

Readers are able to detect m^{6}A modifications on mRNA occurring at different sites through divergent binding modes (Figure 9). One class of m^{6}A readers is characterised by the conserved YT521-B homology (YTH) domain, including YTH domain family 1-3 (YTHDF1-3) and YTH domain containing 1-2 (YTHDC1-2) in humans. YTHDF2 is cytoplasmic and promote the degradation of its mRNA substrates by recruiting the Carbon Catabolite Repression-Negative On TATA-less (CCR-NOT) deadenylase that removes the poly(A) tail[72,81], whereas YTHDF1 and YTHDF2 promote the translation of mRNAs by interacting with the translation initiation factor eIF3[72,82]. All the YTH members are essential for interpreting m^{6}A modifications and

promote different impacts on mRNAs function or lifetime. They may act in a concerted fashion to regulate a large set of cellular physiological processes that are m^{6}A dependent[82].

1.3.1 The YTH domain-containing readers: the YTHDF1-3 and YTHDC1-2 proteins

1.3.1.1. YTHDF1-3

Early studies showed that YTHDF1 and YTHDF3 associate in a complex that binds m^{6}A sites on transcripts and recruit translation initiation factors. In this process, YTHDF3 helps in increasing the affinity of YTHDF1 towards the target mRNA and promoting the correct positioning within the ribosome[82,83]. Importantly, the YTH protein family is actively involved in stopping/blocking the reverse transcription of HIV-1 genome after infection. They indeed bind to the 3′UTR methylated sites on HIV-1 Gag transcript, leading to degradation and inhibition of protein expression[83]. In intestinal organoid experiments, it was observed that YTHDF1 ensures basal expression for regeneration and intestinal stem cell maintenance while, upon Wnt3b high stimulation YTHDF1 levels increase, resulting in hyper-expression of TCF4 transcription factor, Dishevelled Segment Polarity Protein 3 (DVL3) and Frizzled class receptor 7 (FZD7) proliferation markers and higher total volume of the organoid[84,85].

These results show that YTHDF1 reader is a strong translational regulator of the Wnt/β-catenin signalling pathway effectors. Other studies show that YTHDF1 enhances translation of eIF3 as notably observed in ovarian cancer and in Merkel Cancer Cells (MCCs)[49,86]. On the contrary, YTHDF2 has shown the ability to recognize m^{6}A modification on mRNAs and reduce their stability through the activation of the deadenylation and degradation of transcript. This role has been detected when *Ythdf2* was deleted in mouse hematopoietic stem cells (HSCs), the level of m^{6}A-transcripts increased and the most enriched genes encoded for inflammatory factors. This resulted in hematopoietic stem cell hyper proliferation and loss of HSCs[87].

YTHDF3 is a protein that shares the 65% of identity with YTHDF1 and YTHDF2 and it has been shown that it interacts with YTHDF1 and YTHDF2 in HeLa cells. In addition, RIP experiments have reported that they recognize the 60% of identical m^{6}A-transcripts expressing translation factors[82]. Taken together, YTHDF1 may contribute to recognize the m^{6}A marks and promotes translation in collaboration with YTHDF3 while YTHDF2 may contribute to m^{6}A-modified transcripts degradation in a concerted fashion. However, a recent work in HEK293 cells has reported opposite results where the three YTHDF proteins act on the same modified transcripts provoking mRNA degradation through the interaction with the CCR4-NOT deadenylation complex[88]. In conclusion, the presence of three YTHDF readers in different cell types and physiological conditions may act in promoting or degrading mRNAs with mechanisms, which require further experimentation.

1.3.1.2 YTHDC1 and YTHDC2

A further example of the YTH reader essential role in gene expression regulation is represented by the YTHDC1 nuclear reader. This protein controls the proper splicing of XIST RNA, a lncRNA involved in the XIST-dependent X-chromosome inactivation process in mammalian cells[61]. YTHDC1 binding on XIST m^{6}A sites promotes XIST- mediated gene silencing of *Glypican 4 (Gpc4)* in female mouse ESCs. Indeed, CRISPR-Cas9 homozygous knockout of YTHDC1 (*Ythdc1*$^{-/-}$) cells could not be recovered, suggesting that deletion of this gene is lethal. However, heterozygous YTHDC1 (*Ythdc1*$^{+/-}$) displayed impaired *Gpc4* silencing in response to retinoic acid in these cells. These data support the idea that YTHDC1 is required for the silencing of X-linked genes during ES cell differentiation[61].

Nuclear YTHDC1 regulates mRNA processing also by interacting with splicing factors and promoting mRNA decay[69,89]. It is able to promote exon inclusion by recruitment of serine/arginine-rich (SR) splicing factor 3 (SRSF3), a pre-mRNA splicing factor and to facilitate mRNA nuclear export by SRSF3 and Nuclear RNA export factor 1 (NXF1). A recent work has demonstrated that YTHDC2 binds preferentially to m^{6}A within the GGACU motif and plays a role in spermatogenesis[90]. However, the induced knock out of the *Ythdc2* gene in mouse pachytene spermatocytes results in microtubule-dependent telomere aggregation and a massively dysregulated gene expression that is not m^{6}A-dependent. This suggests that YTHDC2 m^{6}A-mRNA binding activity is dispensable in pachytene cell[91]. Here, the authors found that upon YTHDC2 depletion, tubulins, dyneins and kinesins were upregulated and the pachytene cells showed telomere clustering and apoptosis at late pachytene stage[91]. It is possible that YTHDC2 is required for meiotic transcriptome maintenance independently from m^{6}A modification in this particular contest.

1.3.2. A minor set of m^{6}A readers other than YTH family

A second set of m^{6}A readers are characterised by the presence of RNA Binding protein Domains (RBDs) such as the K Homology (KH) domains, the RNA Recognition Motif (RRM) domains, and/or the arginine/glycine-rich (RGG) domains. These readers are sensitive to the structural changes induced by the presence of m^{6}A. This is notably the case of the Heterogeneous Nuclear RiboNucleoProteins (HNRNPs), including HNRNPC, HNRNPG, and HNRNPA2B1, which regulate alternative splicing of target transcripts[92,93]. A third family of protein including the Insulin-like Growth Factor 2 mRNA-Binding Proteins 1–3 (IGF2BP1-3) recognizes the consensus sequence GG(m^{6}A)C, and enhances mRNA stability and translation[94]. Recently, a new reader has been identified, the Proline rich coiled-coil 2A (Prrc2A), that is able

to detect m^{6}A modifications on mRNA during oligodendrocytes myelination process. However, the mode of interaction between Prrc2A and its mRNA targets is still undiscovered[95].

All the reported evidences about reader proteins stress their essential contribution to a huge number of biological processes including infection response, stem cell differentiation and chromosome silencing.

1.4. Erasers of m^{6}A modification

The m^{6}A modification can be reversed by enzymes called RNA 'demethylases' or 'erasers' that counterbalance the activity of writers enzymes. The most representative are Fat mass and Obesity-associated protein (FTO) and ALKB Homologue 5 (ALKBH5).

1.4.1. The Fat mass and Obesity-associated protein FTO is a m^{6}A eraser

FTO belongs to the non-heme FeII/α-KG-dependent dioxygenase AlkB family proteins shown to demethylate N-methylated DNA/RNA bases through oxidation. In 2011, Jia and colleagues stated that FTO possesses a demethylase activity, able to remove the methylation mark in position N6 of adenosines[96]. Further studies have reported that it demethylates also N6, 2′O-dimethyladenosine (m^{6}Am), a modification that has an identical chemical structure in the base moiety to m^{6}A and that is found on the first base adjacent to the 5′cap (cap-m^{6}Am) of mRNAs[50,97]. Several studies on FTO functions during more than ten years have revealed its multifaceted ability of removing methylated groups from various nucleosides, such as m^{1}A in specific tRNAs, m^{6}Am in some snRNAs, internal m^{6}A and cap-m^{6}Am in ssDNA and m^{5}C in some tRNAs[53,98].

FTO knock down studies in HEK293, HeLa and 3T3-L1 cell lines showed a strong increase of m^{6}A/A ratio and m^{6}Am presence on mRNAs, confirming that it is able to remove methyl groups of most of 5′cap and internal adenosines[99]. FTO is cytoplasmic in HEK293 and HeLa cells where they target mRNAs that are more abundant in cytoplasm while the m^{6}Am is protected by 5′cap in the nucleus[96,99]. Other studies have identified FTO in nuclear speckles of HEK293, HeLa and MCF-7 cells, which are known interchromatin granule clusters enriched in pre-mRNA splicing factors[100]. FTO substrate preference is also controversial: it prefers demethylating m^{6}Am than internal m^{6}A in linear ssRNA. This preference might be due to the feature of the cap structure (m^{7}Gppp), including the positively charged m^{7}G, the negatively charged triphosphate, and the 5′ terminus that might increase the binding affinity of mRNA with FTO.

On the contrary, FTO has an increased demethylating activity towards m^{6}A compared to m^{6}Am in tertiary-structured RNAs[98]. This confirms that FTO demethylates m^{6}A and m^{6}Am, as Wei

et al. demonstrated too. Nevertheless, Wei and his group claim that FTO preference towards each of the substrates rather depends on its subcellular localization and the cell type[99].

In a previous study, it has been shown that FTO has the ability to sense the amino acid levels and activates the mammalian Target Of Rapamycin Complex 1 (mTORC1) signalling system in the cytoplasm. Indeed upon amino acid starvation, FTO mRNA levels are decreased, resulting in blockage of cell growth through the mTORC1 pathway and reduction of protein synthesis[101]. A more recent study showed that mTORC1 pathway is downregulated upon amino acids starvation but the FTO levels are not reduced. Moreover, they showed that FTO can be ubiquitinated and degraded by the proteasome, an event that can modulate its translocation in the nucleus[102]. The shuttling of FTO from the nucleus to the cytoplasm may occur due to the binding of partners as the exportin XPO2, which drives FTO towards the cytosol in COS7 cells[101]. Thus, the interaction of FTO with various partners and regulation by ubiquitination may explain its various levels in different cellular compartments where it affects mRNAs fate.

In the future, a better knowledge of FTO biochemical mode of action could serve to design inhibitors that can selectively impair its demethylating activity towards either m^6A or m^6Am in pathogenic scenarios such as obesity, cancer and neurodegenerative diseases[98].

1.4.2. ALKBH5 demethylase operates around m^6A-enriched stop codon

ALKBH5 is the second discovered RNA demethylase that catalyses the removal of the m^6A modification, when enriched around the stop codon on nuclear mRNAs as observed both *in vitro* and *in vivo*. Its demethylation activity affects nuclear mRNAs export and metabolism, supporting the broad biological roles of the reversible m^6A modification on mRNAs[98,103]. ALKBH5 has been localised at nuclear speckles where it elicits its demethylating activity on pre-mRNAs. Apparently, it exhibits a stricter substrate preference by catalysing demethylation of m^6A on ssRNAs and favouring the sequence of (Pu[G>A] m^6AC [A/C/U]) over random sequences[104].

Studies on mice have shown impaired spermatocytes maturation upon ALKBH5 knock out. This phenotype might be due to the increase of m^6As close to 3′UTR which generate longer transcripts that undergo aberrant splicing[105]. They are processed in shorter transcripts than the normal size, resulting in spermatogenesis delay and infertility. It could be interesting to speculate that difference in tissues expression between ALKBH5 (mostly expressed in ovary and testis) and FTO (ubiquitous) correlate with their specific function: ALKBH5 KO mice display spermatogenesis delay, while FTO deficiency is associated with protection from obesity, but also with growth retardation and premature death[106].

1.4.3. The conserved FTO and ALKBH5 erasers contribute to normal human metabolism and physiology

The eraser group of proteins is necessary to remove the methylation tag from transcripts, ensuring a reversible modification event. The two most studied erasers are FTO and ALKBH5, and their deficiencies lead to a wide set of altered biological processes as described above. Genome Wide Association Studies (GWAS) has allowed the identification of FTO SNPs along the first intron which are associated to predisposition to breast cancer, colorectal cancer and pancreatic cancer development[107], while FTO depletion can be implicated in the development of some neuro-pathologies[108].

In Yamaji's work, the rs8050136 variation in the first intron of *fto* increase the risk to develop colorectal cancer as well as increased leptin concentration. Leptin is the protein product of the *obese* gene and its plasma concentration is regulated by the adiponectin. These hormones are regulators of adipocytes homeostasis and accumulation of leptin is known to predispose to cancer. In this case-controls study, CRC patients showed a novel SNP rs8050136 in *fto* gene, which might impact leptin and adiponectin levels and predispose to colorectal carcinogenesis[107].

Other studies suggest that FTO plays a crucial role in neurogenesis, cell proliferation and migration pathways by removing m^6A marks on transcripts and promoting translation of a large set of mRNAs. New identified SNPs in FTO are associated to some neuropsychiatric diseases as attention deficit hyperactivity disorder in children[109], Major Depressive Disorder (MDD)[110] and Alzheimer's Disease (AD), showing that those variants might modulate the translation efficiency of target substrates in neuronal cells important for neuronal function[111]. For example, the rs9939609 single nucleotide polymorphism of FTO (FTO-AA) is proposed to be associated to AD and dementia but no causal effect has been demonstrated yet. The only confirmed risk factor at the present date for AD is APOlipoprotein ε4 (APOE4), which is hyper-translated in AD patients[111,112]. Recently, Han et al. have reported high m^6A levels of the APOE4 transcript in the hippocampus and cerebral cortex in AD mouse models, due to a significant increased METTL3 coupled to a decreased FTO expression suggesting a crucial role of FTO in the development of this disease[113].

The second characterised m^6A demethylase is ALKBH5, which contributes to the reversibility of methylation marks at the 3′UTR on RNAs. Recent studies demonstrated that Breast Cancer Cells (BCSCs) are resistant to chemotherapy because of the Hypoxia Inducing factors (HIFs) which activate the translation of factors involved in proliferation, motility and metastasis[114,115]. In BCSCs, HIFs induce elevated expression of ALKBH5 demethylase, which mediates demethylation of NANOG mRNA. It results in an increased mRNA stability and higher

NANOG protein expression which is required for pluripotency maintenance in BCSCs[116]. Studies on ALKBH5 involvement in tumorigenesis have acquired more and more relevance in the last few years, for the design of new therapeutic treatments of various cancer types.

In addition, ALKBH5 overexpression induces a consistent demethylation of the transcription factor Forkhead box M1 (FOXM1), which promotes proliferation and malignancy both in glioblastoma and lung cancer cells[117]. Furthermore, in lung cancer cells, ALKBH5 overexpression is induced by intermittent hypoxia, a common feature with breast cancer development and chemo resistance[116,118].

Interestingly, ALKBH5 seems to be crucial also in male germ cell differentiation during spermatogenesis. Mice testes expresses the highest levels of ALKBH5, and global inactivation of ALKBH5 leads to male infertility. ALKBH5 KO mice display immature testis and elongated spermatids, confirming that the lack of ALKBH5 may lead to the abnormal demethylation of transcripts involved in many steps of spermiogenesis, such as in ciliogenesis and chromatin packaging[105,119].

1.5. *Drosophila melanogaster* RNA writers, readers and erasers

The deposition of m^6A, which is the most abundant mRNA modification, is catalysed by the METTL3/14 complex in association with WTAP, Vir and RBM15 in higher eukaryotes, as mentioned before. In *Drosophila*, corresponding orthologues have been identified: Mettl3 (Ime4), Mettl14, Female lethal-2-d Fl(2)d, Virilizer (Vir), and Spenito (Nito). It is known that Ime4/Mettl14 form the m^6A-METTL complex (MAC), while Vir, Fl(2)d, Flacc and Nito constitute the m^6A-METTL Associated Complex (MACOM)[120]. The m^6A modification is reversible in mammals but scientists failed in finding orthologues of FTO and Alkbh5 demethylases in *Drosophila*. It is surprising that other members of Alkbh family are present in the cytoplasm and in the nucleus, but since their depletion does not affect the m^6A/A ratio, it is hard to believe that they could be involved in demethylating m^6A[120].

It is possible to speculate that other enzymes, classified as DNA methylases might elicit this role on *Drosophila* transcripts. If this is the case, it suggests that the evolution drove towards a further specialization of demethylating enzymes in higher eukaryotes. Lastly, orthologues of mammalian readers have been characterised in *Drosophila*: YT521-B is the closest orthologue of YTHDC1[120], Benign gonial cell neoplasm (Bgcn) is the orthologue of YTHDC2[121,122] and CG6422 is the orthologue of YTHDF1/2/3[123]. In this chapter, I will present the main players triggering m^6A modification in *Drosophila* and their functional role.

1.5. mRNA writers in *Drosophila*

1.5.1. The main m^{6}A writer in *Drosophila* is the Ime4/Mettl14 complex

In *Drosophila,* the deposition of m^{6}A modification is determined by multiprotein complex called Inducer of Meiosis 4 (Ime4), which shares high sequence and structural similarities with the mammalian METTL3. Ime4 co-localizes with Mettl14 and RNA Pol II in the nucleus and recognizes the RRACHA consensus sequence enriched in m^{6}A at internal splicing junctions and stop codons[120]. Deposition of this mark is regulated during development. Indeed, high m^{6}A levels are detected during the early stages of embryogenesis while they fall down till pupariation. During adulthood, m^{6}A is mostly detected on transcripts isolated from heads and ovaries.

To define Ime4 function in flies, Hongay et al. generated homozygous mutants by deleting the entire methyltransferase gene[124]. Homozygous mutant males and females displayed larval and pupal lethality while a small deletion at the 5′ end of *Ime4* gene resulted in a hypomorphic mutation presenting semi-lethality and subfertility. The generation of hypomorphic flies allowed to decipher further the role of Ime4 at the molecular level. Here, aberrant expression of Ime4 causes defects in germ line-soma interactions such as failure of follicle cell differentiation and germ-line cyst encapsulation, similar to defects reported for Notch signalling mutants[124]. Indeed, in *Ime4* mutants Notch activity was recorded at lower levels than observed in wild type flies, suggesting that Ime4 directly or indirectly modulates Notch expression and signalling during oogenesis. In addition, Lence et al. and Haussmann et al. demonstrated in two different studies that *Ime4* homozygous mutants carrying a deletion of the catalytic domain, are not lethal but flightless and female mutant flies showed a tendency to maleness[120,125].

In *Drosophila,* the Sex-lethal (Sxl) factor is the master regulator of sex determination. The presence of an extra exon introduces a premature stop preventing its expression in males (Figure 10).

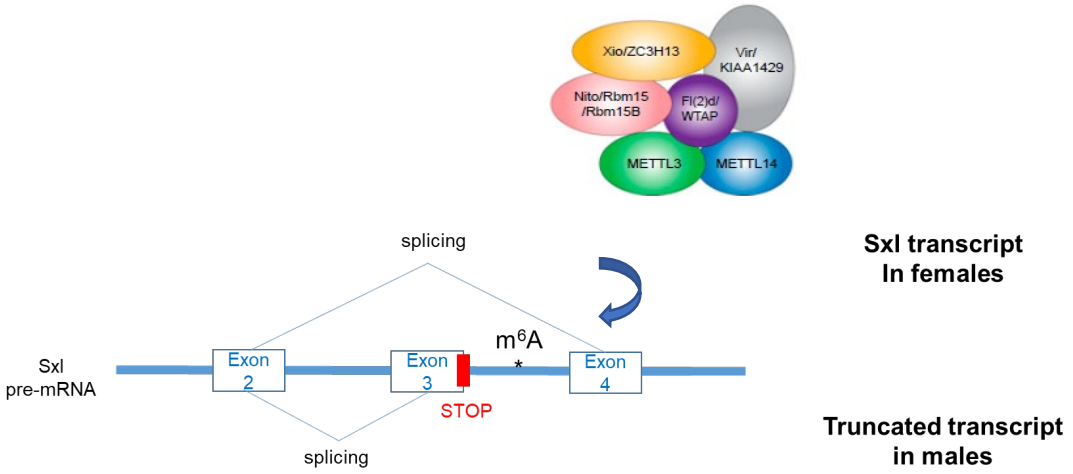


Figure 10. Sxl mRNA maturation in *Drosophila*.

In *Drosophila* Sxl mRNA maturation is controlled by the level of m^{6}A present on the transcript. During the splicing event, if the exon 3 is skipped, the Sxl factor will determine the female phenotypic characteristics in XX. If the exon 3 is maintained, the presence of a premature stop codon produces a truncated and inactive factor (Adapted from Guo *et al.* 2018)[126].

In females, Sxl regulates its own pre-mRNA alternative splicing and other sex determining transcripts fate, such as the repression of the male-specific lethal 2 (mls-2) encoding transcript. According to Haussman and collaborators, the number of hatching *Ime4* null females is lower than expected and survival females show male sexual bristle combs and tumorous ovaries[125]. Phenotypic studies clearly demonstrates the association of *Ime4* inactivation with embryonic development and sexual abnormalities due to dysregulation of target specific transcripts, but the molecular mechanisms of action of the mutated Ime4 protein are still unknown.

As in the case of the mammalian methyltransferase complex, Ime4 works in association with other proteins in *Drosophila*. Ime4 notably binds and co-localises with the protein encoded by CG7818, named Mettl14 (from its structural and sequence similarity with the human METTL14). Mettl14 acts as an adaptor of the Ime4 methyltransferase to the mRNA targets and its deletion results in similar reduced levels of m^{6}A on mRNAs, embryonic development abnormalities and female to male transformation. These observations clearly indicate a common requirement of Mett14 and Ime4 during fly development comforting the hypothesis that they act in the same protein complex[123].

1.5.2. The WTAP orthologue in fly: the Female-lethal (2)d component (Fl(2)d)

Many other methyltransferase components were identified from phenotypic variations induced by mutants. Female-Lethal (2)d is the orthologue of mammalian WTAP and localises in the nucleus together with Ime4-Mettl14, Vir, and Spenito[127]. Fl(2)d is implicated in alternative splicing regulation of genes involved in sexual determination[120,128,129], and is also required for eye development[130]. Similarly to other m^{6}A writers in *Drosophila*, Fl(2)d is highly expressed in CNS and ovaries, imaginal discs, and fat tissue during early embryogenesis. Its expression decreases during larval stage and rises again in pupal stages, which coincides with the augmentation of m^{6}A levels.

Since early 1990s, it was clear that Fl(2)d mutant display a range of phenotypes from female infertility to lethality[127,131]. This is due to the fact that Fl(2)d directly interacts with Sxl, influencing the alternative splicing of Sxl dependent targets, such as *tra* and *Sxl* pre-mRNAs. Moreover, co-immunoprecipitation experiments from ovaries extracts have shown that Fl(d)2

physically interact with early splicing factors including Snf, U170K, U2AF50, and U2AF38 in a similar fashion as Sxl, suggesting that it might be implicated in early mRNA splicing steps[129,132].

Several functional studies thus highlight that Fl(2)d plays a crucial role in sex determination and dosage compensation. Sequence analysis of Fl(2)d have shown the presence of a histidine and glutamine enriched N-terminus, namely the HQ-domain, that shares structural similarity with some transcription factors such as TFIID, Sp1 and SRY, and that is required for protein-protein interaction. This suggests that Fl(2)d may also play a role in transcription regulation through interaction with these factors[127].

As I briefly mentioned before, Fl(2)d is required for a proper fly eye development. Its role in this process depends on its association with Sine oculis (So), the first identified member of the SIX family of transcription factors[130]. So has a dual role in regulating gene expression within the retina. On one side, So factor activates the transcription of Retina Determination (RD) genes including itself, *eyeless (ey)*, *eyes absent (eya)* and *dachshund* (Figure 11).

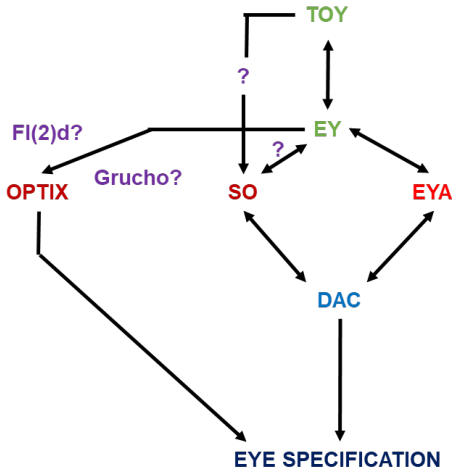


Figure 11. Sine oculis regulatory pathway in eye development in *Drosophila*.

The scheme shows that Sine oculis (So) factor activates the transcription of Retina Determination (RD) genes including itself activation, eyeless (ey), eyes absent (eya) and dachshund (dac). The regulation of So expression could be determined by Fl(2)d and Groucho factors (Adapted from Weasner *et al.* 2007)[133].

On the other side, So is required for the repression of ey expression[134–136]. Moreover, So protein modulates transcription of target genes through the interaction with factors including Eya and Groucho (Gro) and others more recently identified, such as Fl(2)d[130,137,138]. Yeast two hybrid assays and co-immunoprecipitation experiments from Kc167 cell line showed that So physically interacts with Fl(2)d and belong to the same complex in living cells. Moreover, loss of function mutation in Fl(2)d encoding gene resulted in an increased transcription of the RNA binding protein Embryonic Lethal Abnormal Vision (Elav) and the Runt-related transcription

factor (RUNX) Lozenge (Lz) encoding genes leading to overproduction of the two corresponding proteins. These two proteins are involved in the specification and maintenance of photoreceptor cells and the non-neuronal accessory cells that are comprised in each unit eye or 'ommatidium'. As a result, *Fl(2)d* mutation causes eye phenotypic defects such as photoreceptor abnormalities and a roughening of the external surface of the compound eye[130]. Altered Fl(2)d also leads to a decrease of So activity, even if the mechanism of interaction with So and other downstream transcription factors need further investigation. The involvement of Fl(2)d in eye development may shed light on abnormal eye development in humans, where mutant variants of its orthologue WTAP have been associated to retinoblastoma[130].

1.5.3. The Fl(2)d/associated complex component or Xiong (Flacc/Xiong)

Recently, two independent studies reported the existence of a new direct interactor of the methyltransferase complex in *Drosophila* called Fl(2)/associated complex component (Flacc) by Kuckles and colleagues[139] also named Xiong in Guo and collegues'work[126]. Flacc has a conserved zinc finger CCCH domain containing-protein 13 and is broadly expressed during embryogenesis. As observed for the other components of the methyltransferase complex, Flacc shows an enrichment in the brain, while a boost of expression is detected at the blastoderm stage within two hours after fertilization at a time where m^6A levels decrease, suggesting Flacc involvement in early embryonic development and m^6A deposition.

Silencing *Flacc* results in sex transformation in females, which shows male-specific sex combs in the forelimbs while external female structures are completely absent. Homozygous knock out females were lethal. On the contrary, males did not show remarkable phenotypic defects[126,139]. These genetic studies of Flacc mutants clearly demonstrate that Flacc is essential for sex determination and dosage compensation in flies, a result notably supported by the fact that its mutation induces the appearance of the male specific exon in Sxl mRNA in females. Moreover, RNA immunoprecipitation showed that Flacc mediates the interaction between Rbm15/Spenito and Fl(2)d on m^6A target sites. Guo and collaborators found out that Flacc/Xio has an additional role from further silencing experiments. Indeed, they showed that partial Flacc depletion induced abnormal wings folding and adult flightless phenotypes as for Fl(2)d and Spenito[126].

1.5.4. SPOC-protein family: the Spenito component (Nito)

Spenito (Spen and Nito complex) is an additional interactor of the methyltransferase machinery in *Drosophila* whose human orthologue is RBM15. It has been reported that Nito belongs to the SPOC-protein family, a group of large and small proteins that possess an N-terminal RRM.

In the Spenito complex, Spen is the large subunit while Nito is the small one, and both are involved in various signalling pathways. Nito, in particular, seems to be involved in wing and eye development in flies[140,141] but a new study has shown that it also plays a role in sex determination, together with Fl(2)d, Flacc and Vir[142].

During eye development, Nito has the ability to positively regulate Wingless (Wg) signalling, while its depletion results in defects in the accessory cells required for the spacing of ommatidia. Similarly, in wing and leg development, *Nito* mutation affects imaginal disc shaping due to the reduction of Senseless (Sens) expression[140,141].

Nito is thought to help the methyltransferase complex in deposing the m^6A tag to enhance translation of proliferation factors downstream Wingless (Wg) signalling. Alternatively, it might cooperate with splicing factors to control the splicing of target transcripts. Strikingly, Nito is also involved in the determination of sex in flies, together with other sex determination factors including Vir, Fl(2)d, SPF45, U1-70K, U2af38, U2af50 facilitating Sxl splicing autoregulation[122,126,142]. Silencing of *Nito* in Germline Stem Cells (GSCs) results in complete female sterility and the appearance of undifferentiated stem-like cells. A similar experiment was performed in somatic cells, resulting in the development of thickened bristle and structures resembling those of males, such as penis apparatus[142].

Since both Nito and Sxl are nuclear and expressed in both soma and germlines, Yan et al. performed RIP experiments in S2 cells to analyse whether Nito interacts with Sxl pre-mRNA. They indeed demonstrated the existence of interactions between Nito, Sxl, and Sxl pre-mRNA supporting that Nito forms a complex with Sxl and that they regulate together the alternative splicing of Sxl mRNA[139,142]. Therefore, on the one hand, Nito regulates Sxl pre-mRNA splicing in germ cells for sex determination. On the other hand, it is a key factor in embryonic development (non sex-dependent), as well as in eye and wing imaginal discs morphogenesis through the indirect regulation of Wg signalling pathway.

1.5.5. The Virilizer component (Vir) assembles with the methyltransferase complex

In *Drosophila*, Virilizer (Vir) is a nuclear component that assembles with the Ime4/Mettl14 complex. It is ubiquitously expressed during embryonic development regulating m^6A levels in males and females, and shows a particular enrichment in the nervous system and gonads[120,129].

A first study on Vir showed that female mutants show sexual conversion to male. This phenotype is due to the fact that Vir acts as a positive regulator of Sxl during sex determination in *Drosophila,* controlling the downstream activation of Tra and Msl factors[143].

Vir partial depletion severely reduces m^6A levels on transcripts of knock down cells while null *Vir* mutants are lethal at embryonic stage indicating that it is an essential gene[120,144]. Co-

immunoprecipitation analysis demonstrated that Vir interacts with Fl(2)d, Nito, Flacc and other early splicing factors ensuring proper Sxl splicing during embryonic development, but the mechanism of Vir interaction within the methyltransferase complex is still not clear[120,142]. This illustrates the difficulty to define which are the complex molecular mechanisms and interplay between multiple factors from the analysis of common phenotypic defects, such as here female sexual conversion.

1.5.6. PCIF1 orthologue in *Drosophila* (Pcif1) is catalytically inactive

Mammalian PCIF1 protein function has been investigated by different labs who demonstrated that it is a peculiar m^{6}Am writer on mRNAs. Indeed, PCIF1 inserts m^{6}Am modification on the first transcribed nucleotide if it is a methylated adenosine (Am)[52,54,74]. The *Drosophila* PCIF1 (Pcif1) shares strong structural similarities with its mammalian orthologue PCIF1. It similarly possess the WW-domain at its N-terminus, followed by a NLS domain, a helicase domain and an MTase domain at the C-terminal part (Figure 12). However, the amino acidic sequence NPPF of the catalytic domain differs from the human protein for a single amino acid, a phenylalanine to histidine (NPPH)[52].

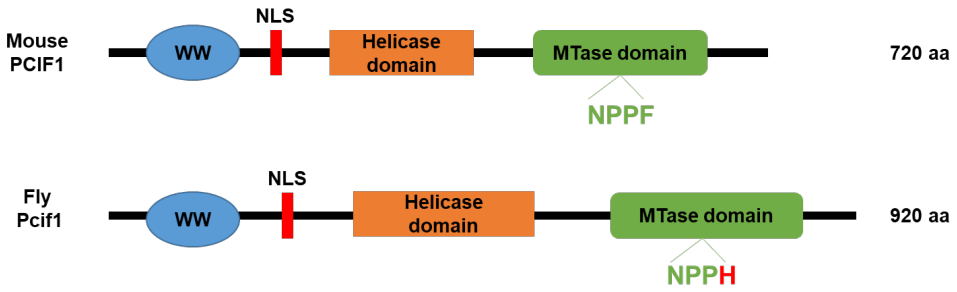


Figure 12. The *Drosophila* Pcif1 protein structure is conserved with its mammalian orthologue.
The scheme shows the structural organization of the mouse (upper) and fly (lower) PCIF1 protein. The WW domain at the N-terminal, the Nuclear Localization Signal (NLS) and the helicase domain are shown. The MTase domain is proximal to the C-term and contains the catalytic motif NPPF in mouse. The fly protein is characterised by the NPPH motif where the phenylalanine is substituted by a histidine residue.

In *Drosophila*, Pcif1 is catalytically dead which is consistent with the fact that no m^{6}Am modifications are detected in this organism[78].

Pcif1 is a nuclear protein that is able to interacts through its WW-domain with the P-Ser5 of CTD RNA Pol II, suggesting that it may influence the expression of various target genes[78]. It is reported in Flybase (http://flybase.org/) that Pcif1 is expressed mostly in fly head and ovaries. However, its physiological role in *Drosophila* is still unclear. In a previous study, Pcif1 was identified by co-immunoprecipitation experiments in a complex containing RNA Pol II and the

Cap Binding Complex (CBC) although it was just mentioned as CG number without referring to this gene as the *Drosophila* orthologue of CAPAM/PCIF1[145]. Taken together, these findings raised the question whether Pcif1 has a function in gene expression regulation independently of any methyltransferase activity that I tented to investigate in my PhD thesis work.

1.6. Readers in *Drosophila* model

1.6.1. The YT521-B (CG12076) is the closest orthologue of the human YTHDC1

YT521-B is one of the few reader proteins identified in flies: it is the closest orthologue to the human YTHDC1. In humans, this reader localises in the nucleus and binds to specific m^6A sites through its aromatic residues in the hydrophobic pocket mediating splicing events. It is known that it interacts with splicing factors such as SFR3 and SFR10 and also binds XIST, the factor controlling the X chromosome dosage compensation[146,147]. Similarly, in *Drosophila* YT521-B localises in the nucleus and recognises m^6A sites on specific target mRNAs.

In *Drosophila*, YT521-B is highly expressed in neuro-ectoderm, in heads and ovaries of adult flies, where it recognises m^6A enriched at 5′UTR transcripts. Mutated YT521-B hypomorphic mutant flies display behavioural defects, such as a decreased ability in climbing and flying, resembling the phenotypes observed in some mutants of the methyltransferase complex Ime4/Mettl14 components[122,123] which is consistent with a role in m^6A mark recognition. Moreover, RNA-immunoprecipitation studies showed that YT521-B binds with Sxl on Sxl pre-mRNA m^6A intron, modulating the alternative splicing at the third exon in males and females. Non-functional YT521-B thus causes excess of Sxl translation, leading to undifferentiated GSCs and tumorous ovarioles, however no sex-transformation was observed[123]. These results suggested that mutation of *Yt521-b* may be involved in fly sex determination together with other components and required for proper oogenesis and female fertility.

1.6.2. The mammalian YTHDF2 is encoded by the CG6422 gene in fly

The closest *Drosophila* orthologue of mammalian YTHDF2 is a predicted gene CG6422 as identified in FlyBase. YTHDF2 (encoded by CG6422) strictly localises in the cytoplasm in flies and possesses the conserved YTH domain, used to recognise specific m^6A site on transcripts[123]. It is highly expressed in CNS within two hours after fertilization while it decreases along the development, except in adult ovaries where its expression is again detected[122]. Homozygous mutants are viable and in contrast to *Yt521-b* mutants, they do not show defects in flight behaviour, suggesting that CG6422 has no role in regulating development, flight or wing shaping or that there is some redundancy with other gene(s).

As other readers, it may be involved in Sxl mRNA processing putatively directing this transcript to the translation machinery in germ cells, but its role there remain equivocal[123]. In addition, it would be interesting to study its effect on mRNAs in the central nervous system given its high expression in this structure.

1.6.3. The Benign gonial cell neoplasm (Bgcn) reader is mostly expressed in fly germ cells

Benign gonial cell neoplasm is the closest orthologue of the vertebrate m^{6}A reader YTHDC2. It is a 3′-5′ RNA helicase, highly expressed in germline and essential for ensuring a successful progression of the meiotic program of germline cells by regulating the level of m^{6}A-containing mRNAs. Indeed, YTHDC2 recognises m^{6}A tags at the 3′UTR of target mRNAs and drives them to degradation by recruiting the RNA exonuclease XRN1[148,149]. Knock out mice for Ythdc2 have defective testis and ovaries, showing an incomplete meiosis that stops at the zygotene step and causing infertility. Two of the most affected mRNA targets of Ythdc2 during spermiogenesis are the messengers of Smc3 and Cep76, two essential actors of meiosis events, which are not efficiently translated in *Ythdc2* mutants[150]. In *Drosophila,* the predicted Bgcn protein is related to the DExH-box family of RNA-dependent helicases because of its RNA binding activity but it lacks ATPase and helicase functions.

In GSCs, Bgcn cooperates with Bag of marbles (Bam) forming a repressive complex which controls the translation of proliferation-related transcripts (Figure 13).

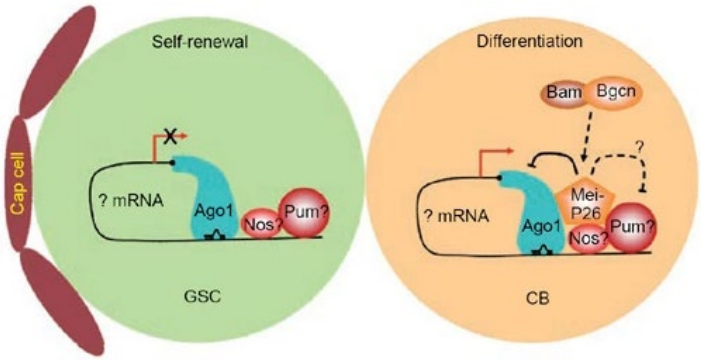


Figure 13. Balance between germ stem cells self-renewal and differentiation in *Drosophila* ovaries.

Nos/Pum maintain the stemness in germ stem cells (GSC) by repressing the expression of genes associated with cell differentiation. On the contrary, Bam-Bgcn control cell differentiation into cystoblast (CB) by activating repressors of Nos/Pum factor and leading to the expression of genes promoting cell differentiation (Adapted from Shen et al. 2008)[151].

The balance between germline stem cells self-renewal and differentiation in fly ovaries is mediated by the antagonistic relationship between the Nanos (Nos)-Pumilio translational repressor of differentiation promoting genes, and expression of Bam, a key differentiation factor (Figure 13)[152–154].

Interestingly, Bam and Bgcn have been associated to sex determination *in Drosophila,* since Bam female mutant germ cells display male inappropriate transcripts expression, similar to Sxl female mutants. It was proposed that Sxl and Bam cooperate for repressing alternative splicing and translation of Nanos mRNA in this process. Co-immunoprecipitation and chromatographic experiment indicated that Meiotic-P26 (Mei-P26), Bam-Bgcn and Sxl physically associate on Nanos mRNA in ovary germline cells, likely repressing its translation and inducing germ stem cells differentiation[153].

In *Drosophila* male germline, Sxl is not expressed while Bam-Bgcn and Mei-P26 are active and control stem cell maintenance and differentiation. Chen et al. in 2014 performed Ethyl Methane Sulfonate (EMS) largescale mutagenesis and obtained a male sterile mutant line exhibiting germ cell overgrowth. The mutation localised on the Tumorous testis (Tut) gene. They further demonstrated that the phenotype of this mutant was comparable to those of Bam, Bgcn and Mei-P26 mutants. Indeed, they belong to a common complex in which Bam recruits Tut through its N-terminal domain and Bgcn through its C-terminal portion. This complex anchors on the 3′UTR sequence of Mei-P26 mRNA to prevent its translation and ensure meiotic division at the proper time[154].

Under normal conditions, Bgcn is present in all stages of sperm cells whereas the other three proteins are not (Figure 14). Tut is very weakly expressed in GSC, or in dividing cells, thus Mei-P26 is not completely repressed. Because Mei-P26 promotes Bam expression and Bam is required for the full expression of Tut, Tut accumulates in the late amplifying cells in which Tut, Bam, and Bgcn can form a complex on the 3′UTR of Mei-P26 to repress it[154,155].

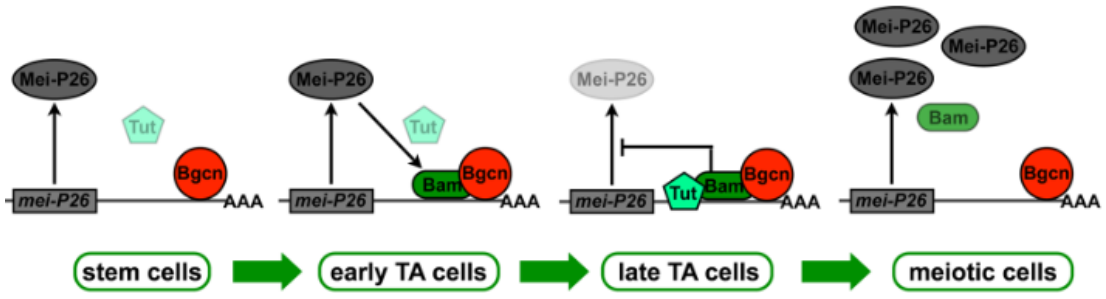


Figure 14. Regulation of Tut, Bam, Bgcn, on their target Mei-P26 during germ cells differentiation.

In normal condition Bgcn regulates Mei-P26 expression, while Tut is weakly expressed (stem cell status). Mei-P26 slight accumulation causes Bam activation, resulting in Tut increased production (early to late TA cells). Tut accumulates and together with Bam and Bgcn, represses Mei-P26. Then Tut starts to be degraded and the

repressing complex dissociates, letting the expression of Mei-P26. The cells enter meiosis (Adapted from Chen *et al*. 2014)[154].

At the end of the amplification stage, when Tut is degraded and Bam levels decrease, Mei-P26 is de-repressed and the germ cells enter the meiotic cycle. Bgcn is an unusual RNA-dependent helicase able to regulate translation through the binding of the 3'UTR of m^6A containing transcripts. The absence of the helicase and ATP activities in flies may be due to two possible reasons: it could be an ancestral RNA-dependent helicase that has lost those activities or an inactive protein that has a conserved RNA binding activity [156].

The wide set of RNA modifications that have been identified so far have shown an implication in RNA metabolism and regulation. Coding and non-coding RNAs can be modified with a variety of modification affecting RNA metabolism at different levels. One of the most abundant mRNA modifications is the m^6A and its effectors influence various aspects of mRNA from stability, splicing and translation. Their biological function is rather conserved from flies to mammals, suggesting a relevant role in cell physiology.
However, the m^6Am modification has gained importance since it has been shown to be deposited on the first transcribed nucleotide of m^7G-capped mRNAs by PCIF1 protein. Interestingly, mammalian PCIF1 is not vital however reduced body weight and gene dysregulation are observed in mutant mice[78]. On the contrary, Pcif1 is structurally conserved in fly but it is catalytically inactive, suggesting that it should have evolved a different function and may interact with other factors in regulating specific gene expression events.

Chapter 2: Canonical versus alternative cap structures and decapping enzymes: the Nudix family

2. A variety of caps and the cap binding proteins

The cap fulfils a variety of functions in all aspects of mRNA metabolism and synthesis, nucleo-cytoplasmic transport, translation, silencing, and turnover. The decapping process ensures the regulation of mRNA turnover and the regeneration of free nucleotides then available for de novo mRNA synthesis. Decapping enzyme 2 (DCP2) was discovered in yeast and subsequently identified as the first eukaryotic decapping enzyme[157]. The destiny of decapped mRNAs in eukaryotic cells is to undergo degradation in a 5′→3′ or 3′→5′ fashion, known as mRNA decay. It initiates with the polysome-associated deadenylation of the poly(A) tail, subsequent removal of the 5′ cap structure, and destruction by 5′→3′ exonucleolytic digestion or by 3′→5′ exosome-mediated decay[158,159]. Since its discovery, DCP2 has been first assumed to be the only enzyme responsible for the majority of eukaryotic mRNA decapping. However, deep investigations during the last two decades have demonstrated that other decapping enzymes exist in bacteria, yeast, plants and mammals suggesting that non-canonical cap structures can control the fate of RNA molecules[160]. Novel alternative cap structures have been notably identified *in vitro* such as Nicotinamide Adenine Dinucleotide (NAD), Flavine Dinucleotide (FAD), diphospho-CoenzymeA (dpCoA), diadenosine-tetraphosphate (Ap_4A), Uridine Diphosphate Glucose (UDP-Glc), and Uridine Diphosphate N-Acetyl Glucosamine (UDP-GlcNAc)[161]. Thus, the identification of putative alternative cap structures raised the question whether they may be hydrolysed by pyrophosphohydrolyses other than the canonical DCP2 also *in vivo*.

DCP2 actually belongs to the NUcleoside DIphosphate linked to another moiety, X (NUDIX) family, a superfamily of pyrophosphohydrolases sharing a highly conserved 23-amino acids Nudix motif (Nudix box), Gx5Ex5[UA]xREx2EExGU, where U is an aliphatic hydrophobic residue[160,162]. In mammals, there are 22 Nudix hydrolases including DCP2 and several of them are conserved in prokaryotes and eukaryotes (Figure 15).

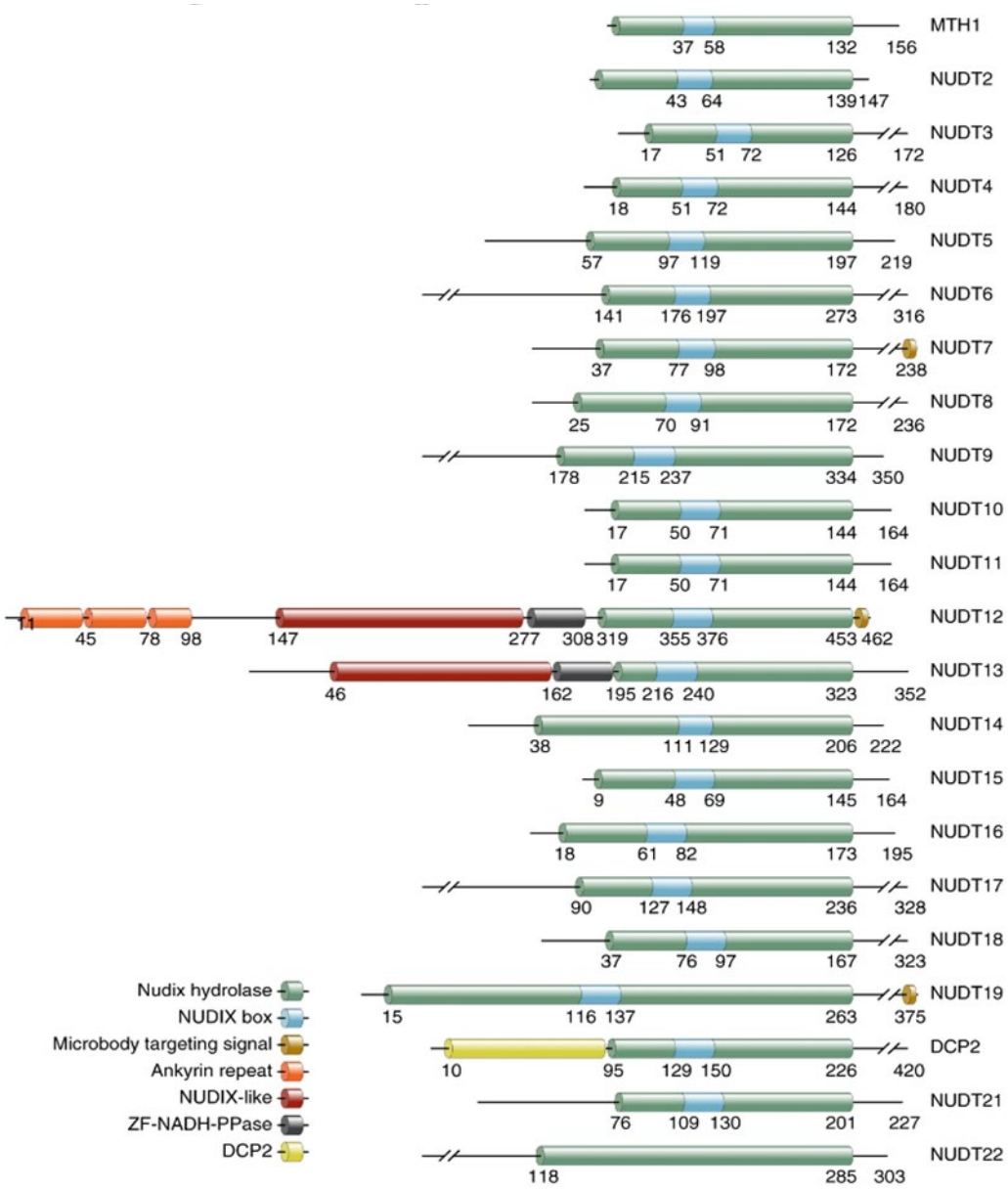


Figure 15. The mammalian Nudix family.

The Nudix family counts for a group of 22 hydrolases in mammals. They all share a common Nudix box (blue box) of 23 amino acids which possess the catalytic activity. The length of each member and the specificity varies along the family. (Adapted from Carreras-Pugivert et al. 2017)[163].

The NUDIX family of proteins hydrolyse cellular metabolites coming from various physiological processes by breaking the diphosphate bond[162,163]. Interestingly, in *A.thaliana* there are 28 Nudix genes which have been grouped in eight classes according to the substrate preferences: 8-oxo-(d)GTP (AtNUDIT1); ADP-ribose/NADH (AtNUDT2,6,7,10,19); ADP-glucose (AtNUDT14); FAD (AtNUDT23); GDP-mannose (AtNUDT9); thiamine diphosphate (AtNUDT20); and Ap_nA (AtNUDT25,26,27,13). AtDcp2 is the canonical mRNA decapping enzyme while the substrate specificity of the remaining members is still not clear[164]. Nonetheless, it is evident that the plant Nudix proteins show hydrolytic activity towards

substrates that are not always the same of the human orthologues indicating a possible specialization or adaptation during evolution.

Some of the Nudix proteins display the ability of hydrolysing the diphosphate bond of metabolites attached to RNA molecule *in vitro* and *in vivo* but the fate of the detached molecule is still unclear[162]. In this part of the introduction, I will present the canonical cap and the poly(A) tail formation and removal, including an overview on non-canonical cap structures identified so far together with the human Nudix enzymes. It is important to underline that in this work we hypothesize that the Nudix family may have finely evolved as a redundant set of proteins with a primary role of hydrolyses that can be converted in specific decapping activity in particular cell type and environmental conditions.

2.1. The canonical 7-methylguanine cap (m^7G-cap)

As stated in the first chapter of the introduction, the RNA Pol II transcribed RNAs including pre-mRNA (messenger RNA), pre-miRNA (micro RNA), pre-lncRNA (long non-coding RNA), snoRNAs (small nucleolar RNAs) and snRNA (small nuclear RNA) are characterized by the co-transcriptional addition of the cap, a 7-methylguanine (m^7G) attached at the first transcribed nucleotide by a 5′-5′ triphosphate bond[165]. The formation of the cap is a multistep process in which many partners contribute to the proper formation. Such interactions are necessary for the cap key nuclear and cytoplasmic functions including pre-mRNA processing, nuclear export, miRNA silencing, translation initiation, and mRNA decay. The 5′cap of the various RNAs species have different fates but here we will describe the destiny of the mRNA 5′cap only.

The most important function of the 5' cap is to protect the mRNA from uncontrolled degradation and to avoid that uncapped or unmethylated mRNAs trigger the innate immune response. Indeed, uncapped viral RNA is recognized as a foreign material, notably through its binding to the Retinoic Acid Inducible Gene-I (RIG-I) cytosolic receptor which ultimately induces Interferon Type 1 (IFN1) activation[28,166]. Strikingly, many viruses have developed strategies to escape the immune system by mimicking the structure of the 5′cap or exploiting the host cell machinery to form their own cap thus preventing their recognition as foreign RNA[28,167].

The first studies on the formation of the 5' cap were actually performed on viral mRNAs obtained by the cytoplasmic polyhedrosis virus (CPV) infecting the silkworm *Bombyx morii* in the 1970's[167]. It turned out that the viral 5' end of the mRNA is protected by the m^7G cap and the process starts from the triphosphate group attached to first transcribed nucleotide (PPP-NTP). Here, the RNA guanylyltransferase and 5' phosphatase (RNGTT) hydrolyses the triphosphate bond and transfer the GMP to the protruding P-NTP of the nascent transcript. Then, the RNGTT enzyme is recruited by the phosphorylated Ser5 of the C-terminal domain of the

RNA Pol II and ensures that the early emerging RNA is capped and methylated in position 7 of the guanine by the RNA guanine-7-methyltransferase (RNMT)[165]. The structure formed so far is called *cap 0*. In the case of mammalian mRNAs, *cap 0* is further modified with the 2′O-methylation of the first and the second transcribed nucleotides by the Cap Methyltransferase 1 (CMTR1) and Cap Methyltransferase 2 (CMTR2) respectively[165]. Other modifications have been identified recently as the methylation of the first transcribed nucleotide if it is an adenosine in position 6 (m^6Am), internal modifications and the 3′ binding of the poly(A) tail which all contribute to different level of mRNA turnover regulation.

2.2. The cap binding proteins (CBP)

The mRNA half-life is tightly controlled and this is due to the cross talk of a huge number of proteins that interacts and/or bind the canonical cap directly or indirectly. Cap binding complex (CBC) and eukaryotic initiating factor 4F (eIF4F) are the major cap binding proteins in higher eukaryotes where the CBC ensures mRNA splicing and nuclear export while eIF4F promotes translation[165]. However, additional binding proteins have been identified such as LA-Related Protein 1 (LARP1), Pumilio 2 and the exon-junction complex core heterodimer Y14/Magoh. LARP1 is able to bind at the 5′ cap of ribosomal protein encoding transcripts by the recognition of the Terminal OligoPyrimidine (TOP) sequence. It results in transcript stability and the blockage of translation. The phosphorylation of LARP1 induces mRNA release and translation leading towards ribosome assembly and cell growth[168].

Pumilio 2 is a 3′ UTR RNA binding factor that represses cell differentiation in eukaryotes. However, it has been demonstrated that it is able to bind to the cap through the Trp^{344} residue impairing eIF4E binding and inhibiting translation (Figure 16).

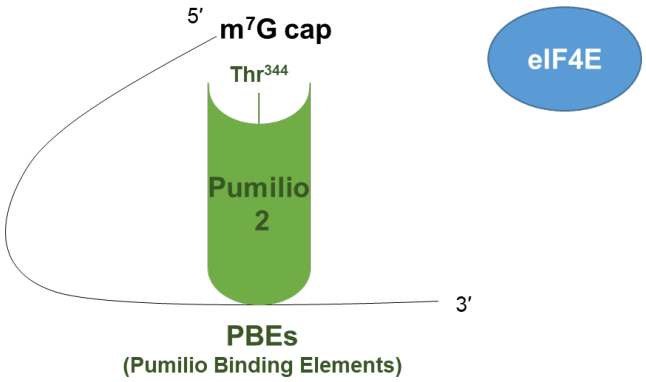


Figure 16. Mechanism of Pumilio 2 repression of mRNA translation in *Xenopus* oocytes.

Pumilio 2 binds to the Pumilio Binding Elements (PBEs) in close proximity to the 3′UTR of the target mRNA in *Xenopus* oocytes as well as to the m^7G cap through the Thr^{344} residue, displacing eIF4E and preventing translation[13].

Its ability to bind the cap is coupled with the 3′ UTR binding at the Pumilio Binding Elements (PBEs) on specific transcripts during oocyte maturation in *Xenopus,* displacing eIF4E and blocking protein translation[170]. In *Drosophila*, Pumilio is involved in the maintenance of the GSCs in the germarium by repressing the expression of differentiating genes as mentioned in the first Chapter of the introduction. Moreover, it is directly involved in the repression of the eIF4E mRNA at Neuromuscular Junctions (NMJs) during larval synaptic boutons development[171]. In mammals, Pumilio 1 and Pumilio 2 repress mRNAs translation by recruiting CCR4-NOT complexes which remove poly(A) tails and prevent translation[172]. These results point out an important role of Pumilio in modulating mRNA translation through its interaction with the PBEs and the 5′ cap.

The RNA binding protein Y14/Magoh inhibits DCP2 decapping activity by directly binding the cap and DCP2, as shown in pull down assays. Moreover, Y14/Magoh reduces the hydrolysis of m^7G cap by DCP2 *in vitro*[173]. Further studies would help in discovering additional partners that bind the canonical m^7G cap directly or indirectly. Here, I will limit the presentation to the most conserved cap binding factors.

2.2.1. The cap binding complex (CBC) is conserved in eukaryotes

The CBC is well conserved in all eukaryotes and is composed by two subunits: the Nuclear Cap Binding Protein 1 (NCBP1) and the Nuclear Cap Binding Protein 2 (NCBP2)[6]. It has been shown that none of them has the ability to bind *in vitro* to the m^7G cap alone suggesting that they act synergistically in a protein complex during the capping process. Indeed, the NCBP2 shows a pocket where the m^7G cap locates due to the presence of two tyrosine residues (Tyr^{20} and Tyr^{43}) but a conformational change driven by NCBP1 is necessary for NCBP2 binding (Figure 17).

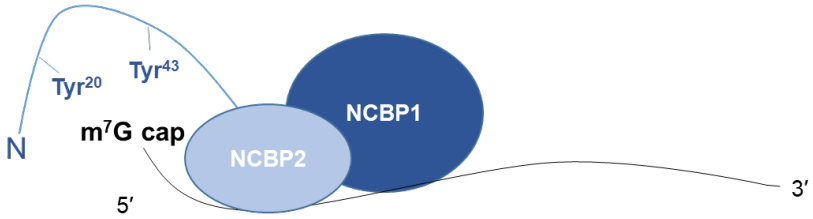


Figure 17. The CBC binds to the m^7G cap and to mRNA by the RRM motifs.

The NCBP2 N-term arm contains the Tyr^{20} and Tyr^{43} that are involved in the m^7G cap binding. The NCBP1 interacts with the NCBP2 in order to induce a conformational change of the N-term that exposes the cap binding pocket. Both NCBP1 and NCBP2 contain RNA Recognition Motif (RRM)[6].

Indeed, only through the interaction of NCBP1 with NCBP2, the N-terminal domain of NCBP2 containing the pocket for the m^7G will bind to the mRNA cap. Moreover, both subunits activity are RNA-dependent[174,175]. Surprisingly, CBC depletion is not lethal in *S. cerevisiae, A. thaliana,* or in mammalian cells but it leads to severe changes in gene expression and defective development[176-177]. In *Drosophila*, the CBC is also conserved and formed by the Cbp80 (NCBP1) and Cbp20 (NCBP2) subunits. In fly, Cbp80 has not catalytic activity but is required for the Cbp20 N-terminal anchoring to the 5' cap of the pre-mRNA. Moreover, CBC also regulate 5′ pre-miRNA processing in normal condition as well as the siRNA-mediated silencing during viral infection[178].

It is known that the CBC maps at 5′ end and within gene bodies of RNA Pol II transcribed genes as reported by ChIP experiments[177] and that it can directly and indirectly regulate transcription via the recruitment of several transcription factors to gene promoters. For example, in *S. cerevisiae*, CBC recruits Mot1p, a transcription factor that promotes or represses gene expression in a gene-specific fashion by activating or inhibiting RNA Pol II recruitment[6]. In addition, CBC interacts with Burp1 and Ctk1p kinases to phosphorylate RNA Pol II at Ser2 and recruit transcription factors triggering increasing methylation of H3K36, an epigenetic mark that enhances transcription[179]. Similarly, in mammals, CBC recruits Positive Transcription Elongation Factor b (P-TEFb) to induce RNA Pol II phosphorylation and promote transcript elongation. However, in higher eukaryotes, the transcription limiting step is not RNA Pol II recruitment but escaping promoter-proximal pause.

In fly, the CBC is able to interact with Arsenic resistance protein 2 (Ars2) regulating small RNAs processing. Indeed, CBC-Ars2 binds to pri-miRNAs through the 5′cap and Ars2 activates the Microprocessor cleavage to generate pre-mRNAs. In a similar fashion, it has been proposed that Ars2 is able to activate Dcr2 for the mediated siRNA silencing upon viral infection in *Drosophila* through the interaction of CBC which, in this context, may act as co-factor of Ars2[180].

The fact that CBC binds to m^7G is fundamental for the steps following the transcription such as splicing, nuclear export and translation. In HeLa cells[181], in *Xenopus* oocytes[182], in *A. thaliana*[183] and in *S. cerevisiae*[184], the deletion of the CBC reduces the activation of snRNPs to splicing sites, leading towards reduced splicing efficiency. Moreover, CBC guarantees efficient mRNA export to the cytoplasmic compartment through the interaction with the transport-export complex TREX, which is recruited in proximity of the transcripts in the nucleus. When the capped-mRNA is properly spliced and ready to be translated, CBC is displaced in favour of eIF4E which promotes translation initiation[185].

Despite the fact that the overall cap binding activity of the CBC is shared among eukaryotes, it is important to highlight that it is able to interact with a wide spectra of factors that can regulate different steps of mRNA turnover and that these different functions can be specific to each species.

2.2.2. The eIF4F complex binds to the 5′ cap of eukaryotic mRNA

The eukaryotic Initiation Factor 4 F (eIF4F) is a protein complex conserved in eukaryotes and consists in a helicase component eIF4A, a cap binding component called eIF4E, and a scaffolding component eIF4G. eIF4F binds through the two conserved Trp^{56} and Trp^{102} to m^7G cap with, as for CBC, an increased affinity when the first transcribed nucleotide is 2′O-methylated[186]. The eIF4F complex is able to recruit eIF3, the initiating factor that activates the tRNA and the 40S ribosome subunit for starting translation. The protein synthesis is balanced by eIF4E component regulation through the mTOR and MAP Kinase interacting Serine/Threonine Kinases (MNK1) which promote cell proliferation and cell growth according to oxygen, nutrient and growth factor availability[187,188]. Additionally, eIF4E possesses two homologues called eIF4E2 and eIF4E3 with lower affinity to the cap but interacting in particular conditions at the place of eIF4E[189,190].The eIF4E2 factor has shown the ability to mediate translation upon hypoxia when eIF4E is inhibited. In this status, a complex is formed by eIF4E2, HIF-2α (Hypoxia Inducible Factor 2 α) and RBM4 (RNA Binding Motif 4) which binds to specific RNA hypoxia response elements and induces translation[190]. Under eIF4E overexpression, eIF4E3 competes with eIF4E for cap binding reducing expression of oncogenes in some subtypes of leukemia thus suggesting eIF4E as a clinical target for inhibitors[191].

Finally, eIF4B and eIF4H are two proteins required for the helicase activity of eIF4A by modulating its affinity to ATP and to the unsolved secondary structures at the 5' end of the mRNAs[192,193]. Moreover, eIF4B is essential for the 40S subunit assembly and its depletion is associated with reduced cell proliferation and cell survival[194].

The transition from the CBC bound m^7G-capped RNAs towards the first step of translation occurs when importin α bound to NCBP1 interacts with importin β in the cytosol. Here, the CBC is released and eIF4F binds the m^7G cap. The binding of eIF4F to the m^7G cap is coordinated by the eIF4E and the eIF4G co-factor (Figure 18).

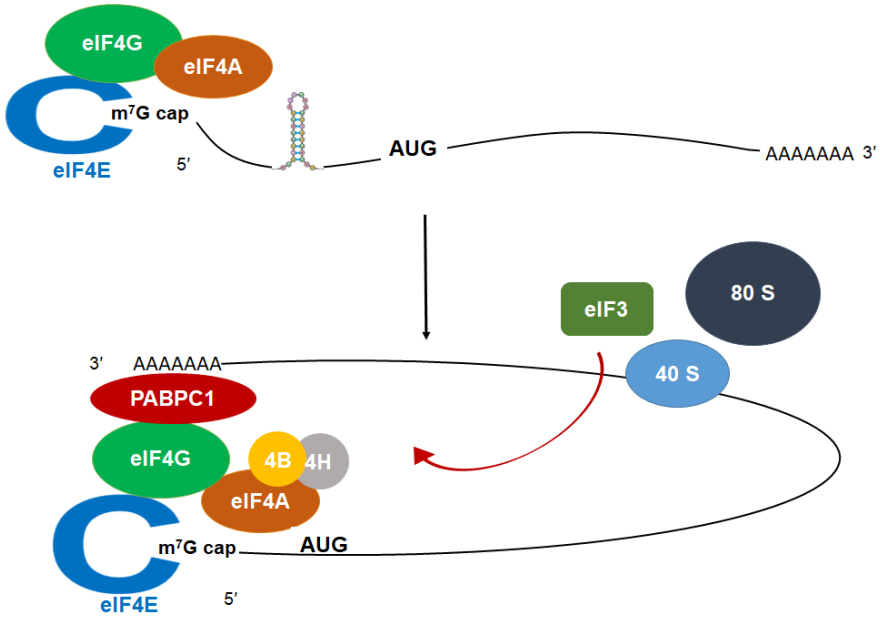


Figure 18. Mechanism of translation initiation mediated by eIF4F complex.

Once the CBC is released, the m^{7}G cap is bound by the eIF4E factor and the eIF4G scaffolding protein. eIF4A RNA-dependent helicase is bound too but it is inhibited by the interaction with 4E and 4G. The eIF4B and eIF4H are then recruited and activate the helicase activity of eIF4A by ATP hydrolysis to solve the hairpin structures at the 5′UTR region. The eIF4G recruits the PABP bound to the poly(A) tail inducing the mRNA circularization. Here, the eIF3 is actively recruited to the complex which stimulates the 40S subunit to join. The pre complex is ready for the recruitment of the 80S subunit and the beginning of translation.

The eIF4G is able to bind to the 3' end binding protein PABP and further promote the mRNA circularization, the recruitment of the 40S subunit, the joining of the 80S, and the activation of eIF4A helicase activity[195,196]. Indeed, the auto-inhibition of eIF4A is due to the eIF4E and eIF4G interaction and the support of the eIF4B and eIF4H complex induces the ATP-dependent unfolding of mRNA secondary structures[197]. Subsequently, protein translation can continue through the activation of the translation factors and the recruitment of the tRNAs.

It is commonly accepted that the eIF4F complex is overexpressed in a wide group of cancers such as lymphomas, angiosarcomes, lung carcinomas, and hepatomes in mice[198] as well as increased angiogenesis[199] and metastasis formation[200], indicating that increased protein synthesis is deleterious for normal cell physiology. For this reason, some studies on eIF4F aim to find a strategy to block dangerous protein synthesis and uncontrolled cell proliferation.

2.3. The NUDIX family

2.3.1. The canonical Decapping enzyme 2 (DCP2) or NUDT20

The decapping enzyme 2 is the major decapping enzyme, first discovered in *Saccharomyces cerevisiae* and conserved from bacteria to eukaryotes[157]. The *dcp2* gene encodes for a protein of 970 amino acids in yeast, while is shorter in humans (420 aa) which is organised in three conserved structural domains: the N-terminal Regulatory Domain (NRD), the Catalytic Domain (CD) containing the Nudix box that are linked by a short hinge involved in conformational changes of the enzyme and C-terminal disordered domain (Figure 19).

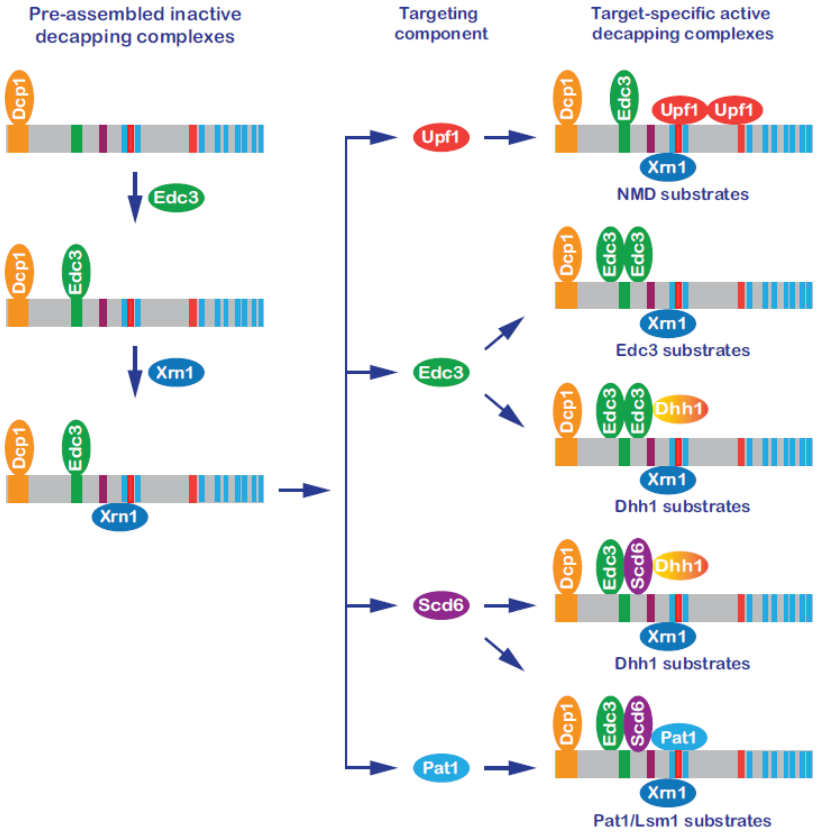


Figure 19. Schematic representation of the Dcp2 protein in *S. cerevisiae* and its involvement in different mRNA decay pathways.

The yeast Dcp2 protein is 970 amino acids long and is characterised by an N-terminal regulatory domain which binds to Dcp1 (yellow box), a Nudix box (in grey, between the yellow and the green boxes) and a C-terminal domain containing binding sites for interacting proteins as Upf1 (red boxed), Scd6 (violet box) and Pat1 (blue boxes). On the right side of the scheme, various fates of mRNA decay are indicated according to the factors bound to Dcp2 (He et al.,2022)[201].

The N-terminal regulatory domain of Dcp2 is involved in the binding with Dcp1, forming the active decapping enzyme, while the Nudix domain is able to bind the capped RNA and catalyse the hydrolysis of the 5' cap. The C-terminal disordered domain is the platform mediating the recruitment of protein partners that can regulate decapping activity in a substrate specific manner (Figure 19). The Dcp1-Dcp2 holoenzyme is involved in the canonical 5'→3' mRNA

turnover but also in the Nonsense Mediated Decay (NMD) of aberrant transcripts by the removal at the 5' end of the m^7G-cap. The result is an overhanging mono-phosphorylated 5' transcript and a 7m-GDP. Mammalian DCP2 is mostly expressed in brain and testis while it is undetectable in heart and liver, suggesting that other decapping enzymes elicits this role in other tissues. DCP1 is reported to be ubiquitously expressed (UniProt database).

DCP1 and DCP2 co-localise in the Processing bodies (P-bodies) and are both essential for the decapping process in metazoa: the deletion of any of the two enzymes results in decapping deficit[202]. Interestingly, DCP1 is exclusively detected in P-bodies while a minor amount of DCP2 is detected in the nucleus. Their activity is enhanced in yeast by the interaction with a set of protein partners, some of which are conserved from flies to humans: Enhancer of mRNA decapping protein 1 an 2 (Edc1 and Edc2) in yeast and Edc4 in flies and humans[157,201,203], the human Proline-rich Nuclear Receptor Coactivator 2 (PNRC2) and probably PNRC1[204]. These proteins have in common a proline-rich motif that interacts with Dcp1 hydrophobic side and increases RNA binding and decapping. Mugridge et al. proposed a model for the activation of Dcp1-Dcp2 in yeast where the binding of PNRC2 and Edc1 and Edc2 to the hydrophobic cleft of Dcp1 led a conformational change of Dcp2, which allows the scanning of the capped mRNA[159]. Subsequently, Dcp2 recognises the cap structures performing a further conformational change due to the interaction between the Nudix catalytic domain of Dcp2 and the activating motif YAGxxF of Edc1 protein and catalyse the decapping process[205]. Later, it was shown that, in vertebrates, Dcp1 interacts with Pat1 for the holoenzyme activation and together with DExD/H box RNA Helicase 1 (Dhh1) and eukaryotic Like Sm protein 1 (Lsm1p-7p), Dcp2 recognises deadenylated mRNA[206]. A set of activator proteins were identified in the non-sense mediated decay (NMD), including Up frameshift 1 protein (Upf1p), Upf2p, and Upf3p that are NMD-specific regulators[201].

The deletion of *dcp1* or *dcp2* orthologues in *C. elegans* induce reduced fertility, impaired locomotion and shortened lifespan while overexpression of Dcp1 in the brain increases worm life span and resistance to stress. Moreover, knocking down of *Dcp1* orthologue in *Drosophila* nervous system results in unexpanded wings during pupal stage and lack of cuticle tanning in adults[207]. In mice, the inducible deletion of *Dcp2* as well as *Dcp1* results in embryonic lethality while in adults it impairs fertility suggesting a pivotal role of the decapping enzyme in cell physiology[208, 209]. Altogether, the published results highlight important roles of Dcp1 and Dcp2 and their regulators in eukaryotes physiology.

2.3.2. The *mth1, mth2* and *mth3* homologues encode for NUDT1, NUDT15 and NUDT18

The *mth* (MuT homologue) gene is a member of the NUDT family and has three homologues in humans, called *mth1 (nudt1)*, *mth2 (nudt15)* and *mth3 (nudt18)* (Figure 20).

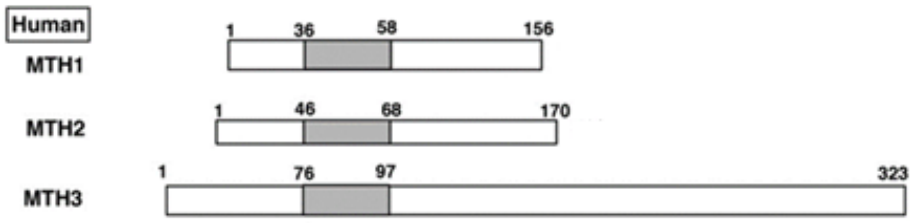


Figure 20. The three MTH proteins share a common MuT box.

The MTH1 (NUDT1), MTH2 (NUDT15) and MTH3 (NUDT18) share the MuT box close to the N-terminal part of the protein (grey box).MTH1 is 156 aa long, MTH2 is 170 aa long and MTH3 is 323 aa long (Adapted from Takagi et al. 2012)[210].

The *mth1* gene encodes a ubiquitous protein of 156 amino acids known as MTH1 (synonym NUDT1) that localises mostly in the mitochondrial matrix and in the nucleus of highly proliferating cells such as Jurkat cells[211], MEF cells[212] and HeLa cells as well as in post mitotic neurons[213]. MTH1/NUDT1 was first identified in *E. coli* and it showed the ability to selectively hydrolyse 8-oxo-GTP to 8-oxo-GMP and to inhibit oxidised nucleotide insertion on DNA[214] (Figure 21).

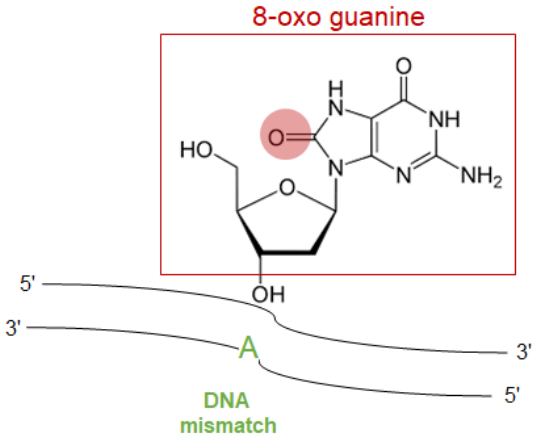


Figure 21. The insertion of 8-oxo-guanine in the DNA sequence generates a mismatch.

The mis-incorporation of an oxidised guanine (red rectangle) induces the base pairing with an adenine instead of a cytosine in green), a mutation that can be maintained and amplified during DNA replication, resulting in a C:G-> A:T transvertion.

The oxidative stress is generated by the continuous production of Reactive Oxygen species (ROS) in the cell as side products of cellular metabolism. External sources of oxidative products can further increase the amount of oxidant species in the cell, resulting in DNA

damage and/or activation of signalling pathways leading towards cell proliferation or apoptosis[215]. One of the consequences of ROS accumulation is the oxidation of deoxy-ribonucleotide that are inserted in the DNA sequence and amplified by DNA duplication. In particular, the insertion of the 8-oxo-guanine in the DNA sequence drives to C:G -> A:T transversion, causing mismatching error in the genome which have been associated with the development of breast[216], kidney[217], and lung cancer[215] as well as with neuronal degeneration[218]. The NUDT1 protein is active on the hydrolysis of the diphosphate bond within the triphosphate bond of the 8-oxo-dGTP , which allows to minimise the insertion of the oxidised nucleotide[219]. However, NUDT1 is not the only enzyme involved in the surveillance of DNA insertion of nucleotides: the OGG1 protein is actually capable of removing the 8-oxo-dG by cleaving it from the DNA sequence[220,221]. This redundant activity of the two proteins highlights that it is extremely important to restrain the genotoxic effect of the insertion of oxidised nucleotide.

In *Drosophila*, the Ribosomal Protein S3 (*RpS3*), CG42813, and Nth-like DNA glycosylase 1 (*Nthl1*) CG9272 genes encode products that have the ability to prevent incorporation of oxidised nucleotide in DNA, but only the protein encoded by CG42813 shares considerable amino acid identity to the mammalian orthologue *mth1*[222]. When it is depleted, flies show faster aging and impaired locomotion together with dopaminergic neuron loss[223]. These phenotypes were observed also in mice depleted of *mth1*, where the pool of oxidised NTP increased the DNA damage in mitochondria[212]. Similarly, in patients affected by Parkinson's disease NUDT1 concentrations are low in neurons, and increased mitochondrial DNA damage is observed[218]. These results indicate that the role of NUDT1 is essential for controlling oxidative stress dependent neurodegeneration along evolution. As in mammals, flies possess a conserved OGG1 orthologue encoded by the gene CG6779, which has a glycosidase activity mediating the removal of oxidised nucleotides from the DNA sequence. It illustrates further that evolution has selected various strategies to avoid DNA damage propagation through specialised enzymes.

The *mth2* gene encodes for a protein of 165 amino acids and localises in the cytoplasm. It was initially called as MTH2, since it was proposed to have a similar function of MTH1[216]. MTH2 possesses a Nudix box and is also known as NUDT15 (and I will further refer to this protein as NUDT15 along the manuscript). Later on, it was demonstrated that NUDT15 had actually a very low affinity for oxidised NTP while it efficiently hydrolysed active metabolites deriving from thiopurine drugs such as 6 thio-dGTP in its mononucleotide form[216,224]. Thiopurines drugs are currently used for treating various cancers and inflammatory diseases. In this context, some human NUDT15 variants have been associated with high predisposition to develop thiopurine intolerance[225] and to fail recovery from ulcerative colitis[226], inflammatory bowel disease[227] , and acute myeloid leukemia [228]. The number of case-control studies involving people carrying

alternative variants of the *nudt15* homologue are presently under way in order to contribute to the development of a personalised medicine.

Finally, the *mth3* homologue encodes for the MTH3 (or NUDT18) protein of 323 amino acids, which localises in the cytoplasm. It has low affinity for oxidised dNTPs, while it is highly active in the presence of ribonucleotides counterparts[210]. Its role is thus to contribute to 'house cleaning' of the RNA from eventual incorporation of oxidised nucleotides in a context where the RNA Pol II is able to incorporate oxidised ribonucleotide and lacks any proof reading activity[229]. In conclusion, the well-conserved MuT containing proteins guarantee an efficient DNA protection from nucleotide oxidation in prokaryotes and eukaryotes.

2.3.3. NUDT2 hydrolases the diadenosine tetraphosphate (Ap_4A) metabolite

Nudt2 gene is conserved in bacteria, yeast, plants, insects and mammals. The human gene codes for a protein of 147 amino acids that is able to hydrolyse the phosphodiester bond within the diadenosine tetraphosphate (Ap_4A) metabolite. Ap_4A is a cellular metabolite generated from the side effect of the aminoacyl-tRNA synthase which ensures the activation and the charging of the amino acid to the corresponding anticodon tRNA during mRNA translation[230,231]. Indeed, the enzyme is capable of transferring the activated amino acid onto a molecule of ATP, inducing the release of the amino acid and the formation of Ap_4A (Figure 22).

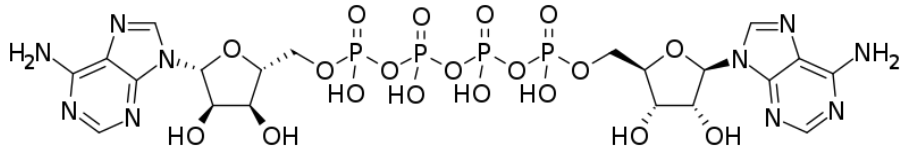


Figure 22. The diadenosine tetraphosphate (Ap_4A) metabolite.

The diadenosine tetraphosphate is the side product of the aminoacyl-tRNA synthase. It is generated by the linkage of an AMP-amino acid (that should be loaded on the tRNA acceptor molecule) with an ATP molecule (Adapted from Mattay, 2022)[232].

The physiological concentration of Ap_4A is very low in prokaryotic and eukaryotic cells[233] but it may increase upon stressful condition such as heat-shock, oxidative or ethanol stresses as shown in bacteria, therefore functioning as an 'alarmone'[234,235]. In mammals, Ap_4A metabolite acts as a second messenger in particular cell types, such as mast cells[234] and dendritic cells[236], contributing to the efficient motility of cells involved in the immune response. However, Np_4N metabolites have been detected only on bacterial RNAs as alternative caps upon oxidative stress, being the Ap_4A the most abundant[237,238]. Indeed, it has been demonstrated that the bacterial ApaH enzyme is able to hydrolyse Np_4A metabolites and keep their concentration

under control. However, ApaH activity is inhibited upon oxidative stress inducing cellular increase of Np_4A metabolite which are than loaded on RNAs[238].

In higher eukaryotes it is not clear whether Ap_4A-capped RNAs do exist but defective human *nudt2* gene has been identified as a prognostic factor in breast cancer[239,240] and some variants of the gene predispose to the development of intellectual disabilities[241,242]. In *Drosophila*, CG31713 is the orthologue of the *nudt2* gene and it is predicted to have the same hydrolysis activity as its mammalian counterpart from sequence similarity (FlyBase database). The CG31713 gene is mostly expressed in the early embryo and in adult ovaries[243] but the role of this gene is still unknown in flies. Evidently, improvement in the detection tools for poorly concentrated substrates is required for the detection of putative alternatively capped-RNAs.

2.3.4. NUDT3, NUDT4, NUDT10 and NUDT11 are DiphosphoInositol PyroPhosphatases (DIPPs)

2.3.4.1. NUDT3 is a redundant m^7G-decapping enzyme *in vitro* and *in vivo*

The human *nudt3* gene encode for a cytoplasmic protein of 172 amino acids. It is part of the DiphosphoInositol PyroPhosphatase group (DIPP) of proteins which are able to hydrolyse the diphosphoInositol Pentakisphosphate metabolite (PP-InsP5) (Figure 23) as well as Ap_6A and Ap_5A[162,244].

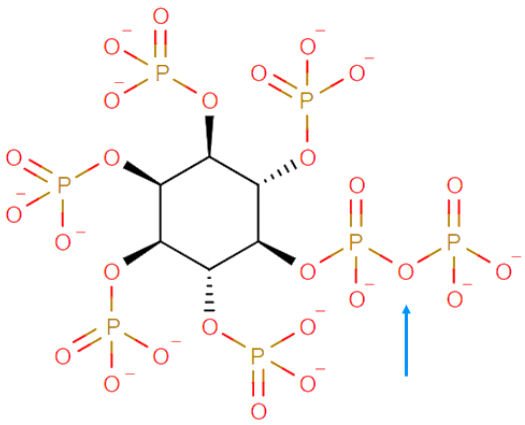


Figure 23. The diphosphoinositol pentakisphosphate PP-InsP5 molecule.

The picture shows the chemical structure of PP-InsP5. In black the inositol ring, in red the phosphate groups. The NUDT3 breaks the diphosphate bond indicated by the blue arrow (https://modelseed.org/biochem/compounds/cpd08361)[245].

PP-InsP5 is a high energy molecule present in the cell that is involved in various signalling pathways and it is tightly regulated by cAMP and cGMP molecules together with phosphatases in order to guarantee constant turnover of these metabolites[246]. PP-InsP5 plays a crucial role in the clathrin-mediated vesicle trafficking since it is tightly bound by Adaptor Protein 3 (AP3)[246]

as well as in the myelin binding due to its high affinity for the proteolipid layer[247]. Moreover, it has been found that PP-InsP5 is involved in mRNA export from the nucleus to the cytoplasm[248]. Nonetheless, it has not been clarified yet whether PP-InsP5 capped RNAs exist *in vivo* and if NUDT3 is able to decap these putative type of RNAs substrates.

Interestingly, NUDT3 would be able to remove the canonical m^7G-cap on mRNAs either *in vitro* and *in vivo* on a subset of mRNAs and its decapping activity would be moderated by PP-InsP5 itself [249–251]. Song et al. have demonstrated that upon *Nudt3* knock down in MCF7 breast cancer cell type, the transcript levels of the Integrin β6 (ITGB6) and lipocalin-2 (LCN2) encoding oncogenes were significantly increased, affecting cell morphology and motility[249]. These results proposed that NUDT3 may act as a decapping enzyme of a specific subset of mRNAs.

2.3.4.2. NUDT4 hydrolyse cleaves the diphosphate bond in PP-InsP5

NUDT4 protein belongs to the Nudix family and is a conserved hydrolyse from worm to fly and mammals. The human gene codes for two isoforms differing for one amino acid, but the isoform α is the most abundant isoform expressed as a cytoplasmic protein of 180 amino acids. NUDT4 is involved in the cleavage of the diphosphate bond within the PP-InsP5 and may share this property with NUDT3, NUDT10 and NUDT11 given their DIPP signature[162,249]. Sharma et al. have reported that NUDT4 protein does not have any decapping activity when different substrate were tested *in vitro*[160], suggesting that they hydrolyse the PP-InsP5 when this substrate is free in the cell rather than attached to RNAs.

2.3.4.3. NUDT10 and NUDT11 hydrolases encode for identical proteins which are tissue-specific

NUDT10 and NUDT11 are identical cytoplasmic proteins, which probably originated from gene duplication event on the X chromosome. During evolution in mammals, the *nudt10* and *nudt11* duplicated genes have experienced 'sub-functionalization', which means a tissue-specific expression despite maintaining the same function[176]. NUDT10 is highly expressed in liver, kidney and testis while NUDT11 is mostly present in mouse brain [252]. They are able to hydrolyse the diphosphate bond within the PP-InsP5 or the diadenosine polyphosphate removing the β-phosphate with a lower affinity for the latter substrate[252]. No decapping activity has been detected for none of the two proteins but their ability to hydrolyse cellular signalling molecules indicate that they are involved in cellular metabolic pathways. Recent studies have identified that NUDT10 expression is significantly increased in gastric cancer from a group of

Caucasian patients[253] while a variant of the *nudt11* gene is associated with high risk of developing prostate cancer [254,255].

These results indicate that they may play an important role in the physiological turnover of energetic metabolites in association with other members of the Nudix family.

2.3.5. NUDT5 is involved in chromatin remodelling

NUDT5 is a nuclear protein conserved in yeast, plants and mammals but not in fly or worm. It is known to have a hydrolytic activity in the presence of ADP-glucose, ADP-mannose and 8-oxo-GDP and 8-oxo-dGDP due to its conserved MuT domain[256]. In the presence of diphosphate it has also the ability to generate ATP molecules in the nucleus. Similarly as shown for NUDT1, NUDT5 ability to remove oxidised nucleotide is essential for the proper DNA replication and to avoid mutation insertion in the sequence, thus acting redundantly with other Nudix members to ensure DNA cleaning activity[256].

However, NUDT5 has been originally studied rather because of its involvement in the progesterone-induced increase of ATP production in the nuclei of breast cancer cells[257]. Indeed, high levels of NUDT5 expression are induced by hormone stimulation together with a significant increase in the levels of nuclear ATP. On the contrary, when *nudt5* encoding gene was silenced, the levels of nuclear ATP were reduced while not the ATP mitochondrial or cytoplasmic concentration of ATP. Moreover, the hormone-dependent increase of ATP production was associated with remodelling of the chromatin which activates the expression of pro-proliferation genes of the Atk/Cyclin D cascade[257]. In another study, it was reported that breast cancer cells from patients display increased NUDT5 expression which was associated to increased proliferation, migration and invasion[258]. In normal condition, nuclear NUDT5 may be involved either in the DNA repairing process as well as in the control of the chromatin status through the regulation of the ATP turnover. Together these evidences indicate that NUDT5 may be a valuable candidate for treating some types of cancer and few groups tested promising inhibitor of NUDT5 activity on preclinical models[259–261].

2.3.6. NUDT7, NUDT8 and NUDT19 hydrolyse diphospho-CoenzymeA substrates

2.3.6.1. NUDT7 hydrolase localises in the peroxisome in mammalian liver

The mammalian *nudt7* gene encodes a protein of 238 amino acids, which localises in the peroxisome and contains an N-terminal Nudix motif and a CoA binding motif that is common to NUDT8 and NUDT19 proteins (Figure 24).

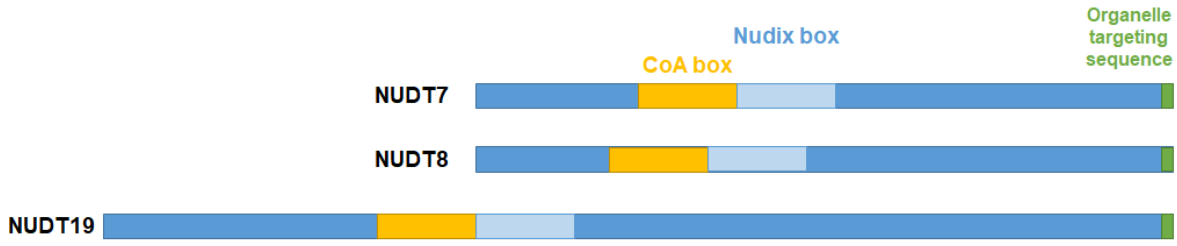


Figure 24. NUDT7, 8 and 19 hydrolases share a common CoA box.

The scheme shows the conserved CoA box (in yellow) next to the Nudix box (in light blue) of the human NUDT7, NUDT8 and NUDT19 proteins. The C-terminus contains the sequence for the organelle specific localization (in green).

The *nudt7* gene is conserved in prokaryotes and eukaryotes and is known to be involved in the hydrolysis of the diphosphate bond within fatty acyl-CoA and bile acids-CoA substrates. The CoA is the key molecule at the centre of different metabolic pathways, such as the fatty acid synthesis in the cytoplasm, the triacyl glycerol cycle (TCA) and the oxidation of medium-length chain fatty acids in mitochondria, the conjugation of bile acids and oxidation of long chain fatty acids in peroxisomes[262]. A little pool of CoA shuttles through the nuclear pores from the cytosol to the nucleus where it contributes to gene expression regulation through histone acetylation regulation[263]. The subcellular distribution of the CoA sources may be explained by the organelle-specific distribution of the three hydrolases. NUDT7 mammalian protein mostly localises in the peroxisomes; in mice, it is specific of hepatic cells, suggesting that its expression is dependent on the nutrition status. Moreover, its expression levels are inversely correlated with the CoA levels[263].

A recent study has shown that mutation of *nudt7* gene in mice induced increased levels of fatty acids in hepatic cells driving towards the onset of Non-Alcoholic Fatty Liver Disease (NAFLD). The NAFLD condition, if it is not controlled, can develop in type 2 diabetes (T2D) and liver failure[264,265]. Nonetheless, the onset of a dysregulated lipid metabolism can not be attributed only to defective peroxisome activity but also to impaired mitochondrial function due to the close interplay between these two organelles.

The mammalian NUDT7 has the ability to cleave CoA from mRNAs *in vitro* (Figure 25) together with NUDT2,8,12,15,16 and 19[160], however CoA-RNAs have yet been only detected in the bacteria *E.coli* and *S. venezuelae* [266] while not in mammals, questioning the physiological relevance of these *in vitro* studies on artificial CoA-RNAs substrates (Figure 25).

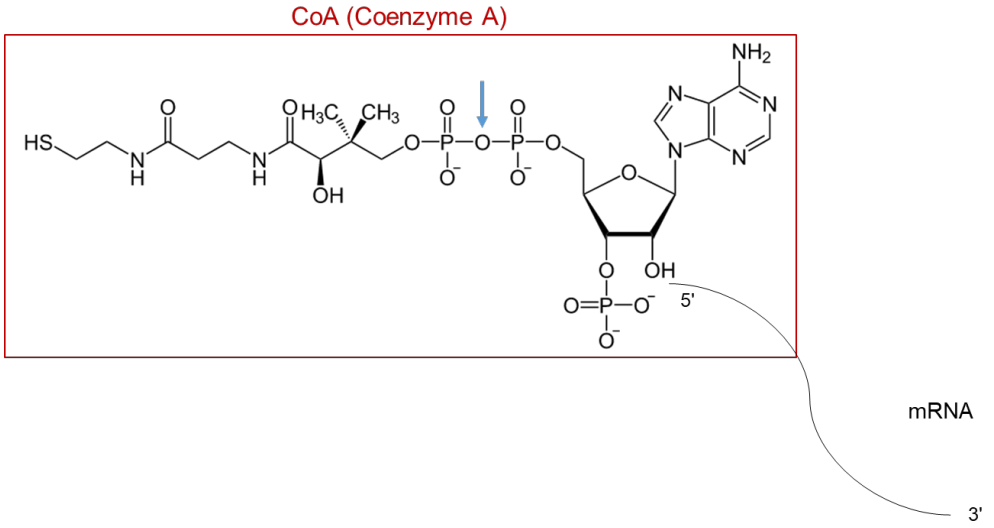


Figure 25. The mammalian NUDT7 hydrolyses the diphosphate bond within the CoA when attached to mRNA *in vitro*.

NUDT7 cleaves the diphosphate bond of the CoA metabolite when attached to mRNAs *in vitro*. The blue arrow indicate the site of cleavage of the NUDT7.

In *Drosophila*, the CG11095 gene codes for a predicted Nudt7 like-protein that is mostly expressed in the ovary. This indicates that the regulation of the CoA metabolism by the orthologue of the NUDT7 in fly may play a relevant role in ovary maintenance and/or fly fertility, a hypothesis that stays to be demonstrated by functional studies.

Interestingly, NUDT7 is widely conserved in plants but it modulates the immune response induced by phytopatogens attack, working as a phosphodiestherase on 2'3'-cyclic AMP (cAMP) or cGMP second messengers which activates cell death and defense response[267,268]. This function detected in plants indicates that the Nudix members are conserved but they have developed different substrate specificity during evolution.

2.3.6.2. NUDT8 is a novel CoA hydrolase

Mammalian NUDT8 has been recently identified as a mitochondrial protein of 236 amino acids which is mostly expressed in heart, kidneys, liver and brown adipose tissue; this enzyme is not present neither in lower eukaryotes nor in prokaryotes[269]. NUDT8 amino acid sequence shares 38% of identity with NUDT7 [269]. It is characterised by a CoA box which is located upstream of the Nudix box[269]. *In vitro* analysis showed that mammalian NUDT8 hydrolyses CoA-RNAs as well as various CoA-fatty acid species[160,269]. Differently from its close relative NUDT7, NUDT8 transcripts levels are not affected by changing nutritional status of mice. This suggests a specific regulatory mechanism controlling NUDT8 expression in the mitochondria that requires further investigation[269].

2.3.6.3. NUDT19 hydrolyse is highly expressed in mammalian kidney

The human *nudt19* gene encodes for a protein of 375 amino acids which localises in the peroxisomes and it is highly expressed in mammalian kidney[262]. It carries the Nudix box, as all the members of the Nudix family, together with the CoA box which is involved in the CoA-substrates recognition and conserved in NUDT7 and NUDT8 enzymes as stated above [160,264]. NUDT19 is able to hydrolase the diphosphate bond within the CoA-substrates and presents an enhanced activity towards medium length-CoA-fatty acids substrates[270]. Orthologues of the mammalian *nudt19* gene exist in bacteria as well as in plants, yeast, worms and in flies. Interestingly, *Drosophila melanogaster* possesses three genes closely related to the unique mammalian *nudt19* gene: CG18094, CG10194 and CG10195 that we respectively named and will referred to as *Nudt19A*, *Nudt19B* and *Nudt19C*. They encode for three proteins that share only a modest percentage of identity (around the 34%). However, no experimental work has been performed yet that would shed light on their respective role in fly development or physiology. I will present the work I have done to characterise the Nudt19 proteins functions in fly in the Results section (see Results-Chapter 2).

Homozygous knock out mice mutants for the *Nudt19* gene are viable and fertile but show enlarged heart and abnormal skin according to the International Mouse Phenotyping Consortium (IMPC database). Moreover the *Nudt19* mRNA levels do not change in mouse kidney in fed or fasted state, indicating no regulation of *nudt19* gene expression by diet[262]. Another recent study showed that NUDT19 is detected in mitochondria in mice liver, in addition to kidney, where it influences the ATP production and the respiration rates. Here, mouse model for the T2D and NAFLD were used to perform RNA sequencing in liver revealing *Nudt19* among the dysregulated genes[271]. In addition, *nudt19* knock down in hepatocellular carcinoma cell line (Hepa 1-6) presented overexpressed *pdk4* gene, which is known to induced increase ATP production and fatty acids oxidation in diabetic mouse model by inhibiting the pyruvate dehydrogenase (PDH)[272]. The *pdk4* gene is also highly expressed in diabetic patients, which is associated with insulin-dependent inhibition of the Pyruvate Dehydrogenase (PDH). These findings suggest that NUDT19 may regulate PDK4 function in hepatic mitochondria and that its activity is not only limited to peroxisomes.

Additionally, NUDT19 has shown diphosphatase activity *in vitro* on m^7G-capped and CoA-capped RNAs, however there are not evidences showing that it can act as a decapping enzyme *in vivo*[160,162,250].

In *A. thaliana*, the *nudt19* mammalian orthologue (*AtNUDT19*) is conserved and localizes in chloroplasts and peroxisomes, where it controls the levels of ROS, which are particularly

increased upon light overexposure. Indeed, it has been shown that AtNUDT19 is able to hydrolyse the NADH in chloroplast and when it is depleted, the levels of NADH strongly increased[273]. In addition, upon *AtNUDT19* mutation, a set of upregulated genes were identified which are mostly involved in the oxidative stress response and resistance suggesting that in plants AtNUDT19 contributes to the defense against oxidative stress.

2.3.7. The conserved NUDT12 hydrolyses NAD-capped RNAs

The human NUDT12 is a protein of 462 amino acid long that is conserved from bacteria to mammals; however there are not orthologues identified in fly so far. *Nudt12* gene is mostly expressed in mammalian kidney, liver and testis and has been detected either in the peroxisomes[274] or in cytoplasmic granules[275]. It is involved in the decapping of the NAD-capped RNAs through the hydrolysis of the diphosphate bond within the NAD cofactor *in vitro* and *in vivo* (Figure 26).

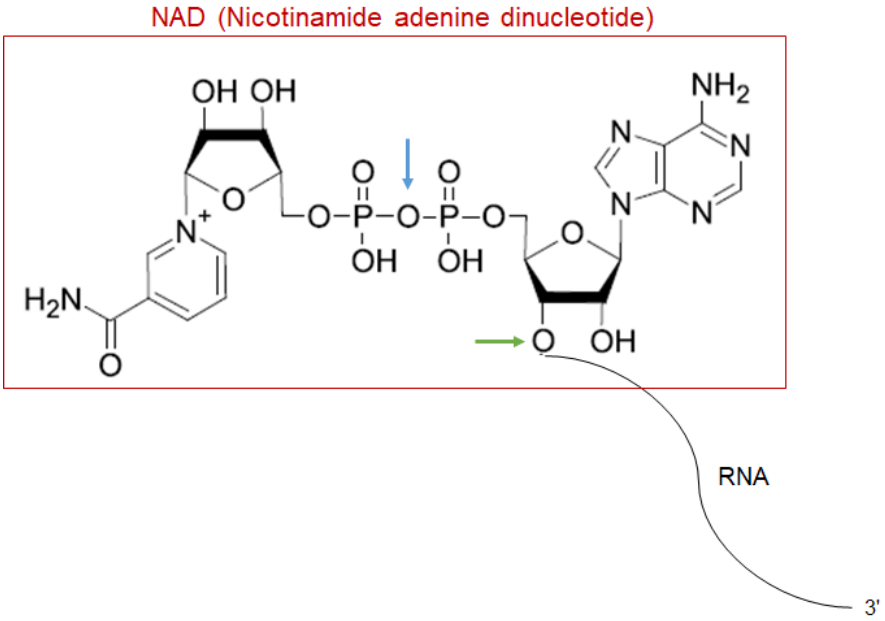


Figure 26. NUDT12 hydrolyses the diphosphate bond within the NAD-capped RNAs.

The NUDT12 enzyme breaks the diphosphate bond of the NAD-capped RNAs *in vitro* and *in vivo*. The cleavage site is indicated by the blue arrow. The cleavage site of DXO is indicated by the green arrow.

The oxidised NAD and the reduced NADH form molecules are important cofactor which regulate the oxidation/reduced status of various metabolic pathways in mitochondria, nucleus, cytoplasm and peroxisomes due to its ability of transferring electrons[276]. NAD(H) can be synthesised by three alternative ways: *de novo* synthesis through the tryptophan amino acid, by the recycling of its precursor or by its demolition due to NAD(H)-consuming enzymes[276]. Lower levels of cellular NAD(H) have been detected in several human diseases such as neurodegenerative diseases[277,278] and metabolic diseases[279,280] and the reintegration in the diet

of mice models have resulted in ameliorating effects. Thus, it is important to maintain proper concentrations of NAD(H) for the normal cellular metabolism.

NAD-capped RNAs were first identified in *E. coli* and *S. venezuelae* by RNA-Mass Spectrometry (RNA-MS) indicating that bacterial RNA Pol are able to introduce non canonical cap at the 5'end of RNAs[266,281]. Later on, different works demonstrated the presence of NAD-RNAs in yeast[282], mammals[283] and plants[284]. NAD capping may occur in bacteria as soon as the RNA Pol begins transcriptions through the recognition of a specific DNA sequence R(-2)R(-1)A(+1)S(+2), where A is the TSS, R is either A or G and S is either G or C[285]. Various promoter-specific sequences have been identified for different organisms indicating that NAD-capping occurs at particular sequences close to TSS[286]. The main difference in the fate of NAD-capped RNAs between prokaryotes and eukaryotes is that in the former RNAs are stabilised and are more resistant to RNase activity while in the latter case they are degraded[281,287]. In *E. coli*, the levels of NAD-capped RNAs are more likely to be related to the environmental and the metabolic conditions during the stationary phase or exponential phase suggesting that NAD caps integration is tightly regulated by bacterial density[286].

In human cells, the highest pool of NAD-capped RNAs derives from mitochondrial mRNAs and rRNAs where the presence of NAD is increased and the mitochondrial RNA Pols are more prone to insert NAD instead of ATP. Here, the levels of NAD attached to RNAs can be influenced by the oxidised/reduced state of the cell and external stimuli[288]. However, the levels of human NAD-capped RNAs are rather low due to the presence of the nuclear DXO and the cytoplasmic NUDT12 and NUDT16 enzymes that are able to remove NAD-caps as demonstrated *in vitro* and/or *in vivo*[160,283]. DXO is a nuclear protein not belonging to the Nudix family, which hydrolyses the 3' phosphodiester bond that links either the canonical m^7G cap or NAD to the following nucleotide, removing the entire metabolite from the 5' end of the RNA (Figure 26, green arrow). This system was first identified as a mechanism that removes defective m^7G canonical cap triggering rapid mRNA degradation[206]. Later, it was discovered that DXO is also involved in the NAD-decapping of a particular set of small RNAs: snoRNAs (small nucleolar RNAs) and the scaRNAs (small Cajal body RNAs)[283].

It is important to underline that DXO contributes only partially to the NAD-capping control in the cell. Indeed, NUDT12 plays a crucial role in removing this cap from cytoplasmic mRNAs, where it localises. In this process, NUDT12 interacts with Bleomycin Hydrolase (BLMH) through the ANK motif in cytoplasmic granules as shown in studies using HeLa cells model[275]. Here, the authors proposed that mouse BLMH may bind RNA in a concerted fashion with NUDT12; although this hypothesis is still speculative in the absence of molecular evidences demonstrating this function. In the same study, the authors reported that *nudt12* knock out mice

were viable and fertile; however a set of genes encoding for the circadian rhythm proteins was upregulated[275]. This indicates that while NUDT12 may play physiological role in circadian rhythm, it is not essential in mammals pointing towards putative redundancy by other compensating proteins, as it has been also demonstrated in bacteria[287,288]. Overall, the amount of NAD molecules is regulated by different actors depending on the subcellular localization. It is still controversial whether NAD-capped mRNAs can be translated. It has been recently reported in yeast[282], plant[284] and human[283,286] that NAD-capped mRNAs can be properly spliced and poly-adenylated, although only in plants they co-localise with ribosomes and are most probably translated[284].

2.3.8. NUDT16 hydrolyses NAD and FAD substrates

The human NUDT16 is a protein of 195 amino acids that is conserved in higher eukaryotes although neither in *Drosophila* nor in yeast is conserved. It belongs to the Nudix family of proteins and its expression is rather ubiquitous. NUDT16 has been detected mostly in the cytoplasm in human cells as well as in the nucleus suggesting a pleiotropic decapping activity[289], while in *Xenopus* it strictly localizes in the nucleus[290]. It has shown hydrolytic activity towards m^7G-capped U8 snoRNAs, small nucleolar RNAs involved in the biosynthesis of ribosomes[291,292]. In addition, NUDT16 has the ability to hydrolyse m^7G capped-mRNAs *in vitro* and *in vivo*, suggesting that it can actively regulate not only the ribosome biosynthesis but also the mRNAs turnover[160,291,292].

According to its nuclear localization, NUDT16 has also been identified as a (deoxi) Inosine phosphatase, removing mutated nucleotides that may alter the DNA or RNA sequence or inhibit DNA or RNA polymerase during DNA duplication or transcription[292]. Among them, deamination of nucleotides can generate (deoxi) Inosine tri- or di- phosphates, which may alter the normal base pairing. Silencing of *nudt16* in HeLa cells resulted in increased DNA breaks, blockage of the cells in S-phase and cell growth arrest[292]. All these results highlight that NUDT16 contributes to the DNA cleaning from potential dangerous insertion of modified nucleotides as other members of the Nudix family, such as NUDT1 and NUDT15.

More recently, *in vitro* and *in vivo* studies have shown NUDT16 decapping activity towards the canonical m^7G, NAD and Flavin Adenine Dinucleotide (FAD) capped-RNAs as well as towards, dpCoA *in vitro* (Figure 27).

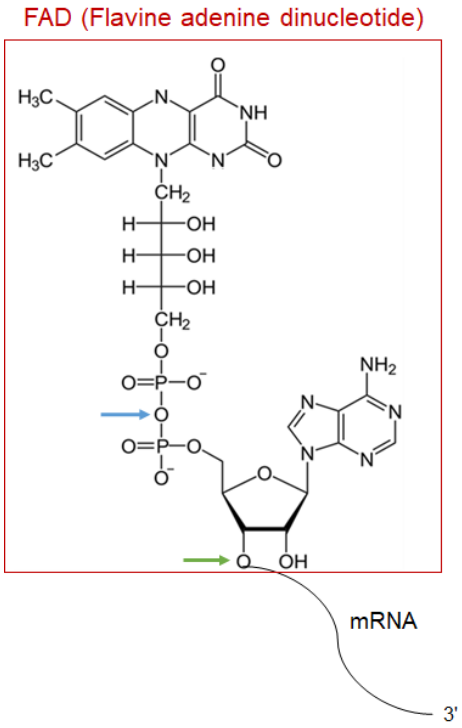


Figure 27. The FAD capped mRNA is removed by the NUDT16 hydrolyse.

The NUDT16 enzyme breaks the diphosphate bond of the FAD-capped mRNAs *in vitro* and *in vivo*. The cleavage site is indicated by the blue arrow. The cleavage site of DXO is indicated by the green arrow.

The ability of NUDT16 to remove NAD caps was actually first demonstrated *in vitro* and then confirmed *in vivo* in HEK293 cells deleted for *nudt16*[293]. In addition, NUDT16 is able to remove dpCoA-caps *in vitro*, as already shown for NUDT2, 7, 8, 12, 15 and 19[160]. Therefore, there are a surprisingly large number of Nudix proteins able to break the link between dpCoA and the 5'-RNAs in contrast to only few number of NAD-decapping enzymes[160].

In Sharma et al., the authors report that NUDT16 is a new FAD-decapping enzymes *in vitro* and in cellulo as shown in HEK293 cells when FAD is attached to short RNAs[160] (Figure 2.9, blue arrow). In a more recent work it was demonstrated that DXO human protein has also FAD-decapping activity *in vitro* and *in vivo* by removing the full FAD moiety through the breaking of the phosphodiester bond at the 5' end of the RNAs[294] (Figure 27, green arrow). The mechanism of FAD-decapping *in vivo* is, thus, controlled at different levels given that not only NUDT16 but also NUDT12 and DXO are able to hydrolyse the FAD-capped RNAs. These results suggest that FAD as well as NAD concentrations in the cell are involved in both the regulation of the metabolic state of the cell and in the mRNA turnover.

FAD, indeed, contributes to regulate the oxidised/reduced state of the cell components and its levels in mammalian cells are relatively low. It derives from the riboflavin, also known as vitamin B2, which is not synthesised in humans but need to be introduced from the diet[295]. FAD exists in the oxidised and in the reduced form (FAD/FADH) in the cell and it serves as a cofactor

for various essential metabolic pathways such as in electron transport chain in the mitochondria[296], in DNA repair[297], in nucleotide biosynthesis[298], in fatty acid oxidation, the synthesis of CoA[299,300] and in the heme group metabolism[301].

Decreased levels of FAD in humans has been associated with low intake of riboflavin which results in milder effects as sore throat, loss of hair and skin inflammation to anemia[302] and impaired nerve functions[303]. Many of these defects can be compensated by the diet uptake of riboflavin and the newest metabolic studies and diet styles are contributing to a better knowledge and consciousness of the consequences on human health due to an unbalanced diet.

2.3.9. The NUDT6, NUDT9, NUDT13, NUDT14 and NUDT17 Nudix members

2.3.9.1. NUDT6 is the antisense transcript of basic Fibroblast Growth Factor (bFGF)

Among the Nudix family, there is a small group of members for which even more little is known compared to the Nudix protein previously discussed.

NUDT6 is a human cytoplasmic protein of 316 amino acids, which is conserved in plants, flies and mammals. The protein is encoded by the antisense strand of the gene coding for the basic Fibroblast Growth Factor (bFGF), suggesting that it may interfere with the expression of the bFGF. Indeed, it has been shown that NUDT6 modulates bFGF-induced proliferation in the pituitary gland preventing tumour cell growth[304]. Another study showed that NUDT6 has a nuclear MuT like activity that can be evidenced in a context where MTH1 is depleted in rat liver but the level of oxidised nucleotides is reduced despite the lack of MTH1[305]. *In vitro* analysis has demonstrated that purified NUDT6 is not able to hydrolyse any of the alternative capped-RNAs previously presented, suggesting that NUDT6 may be implicated in cellular mechanisms other than RNA decapping.

2.3.9.2. NUDT9 hydrolyses the ADP-ribose substrates (ADPR)

The NUDT9 human hydrolyse is a mitochondrial protein of 350 amino acids that is conserved in mammals as well as in worm and flies but not in yeast or prokaryotes[306]. It is known to hydrolyse the pyrophosphate bond within ADP-ribose substrates (Figure 28), which mainly derived from the NAD catalysis or from the ADP-ribosylated proteins in mitochondria [307].

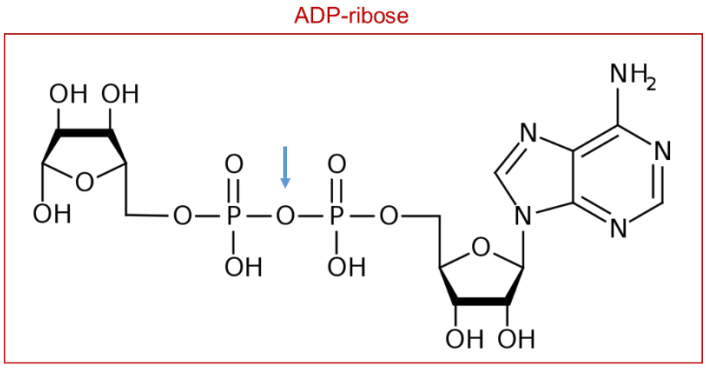


Figure 28. The ADP-ribose is hydrolysed by NUDT9 *in vivo*.

The ADP-ribose is the substrate of the NUDT9 and the cleavage site is indicated by the blue arrow.

Mono or poly ADP-ribosylation occur on amino acid carrying a free N-, S- or O- acceptor group[308]. In eukaryotes, poly ADP-ribosylation is the most common modification compared to mono- ADP-ribosylation and it is promoted by the poly ADP-ribose polymerases (PARP) which transfers the polyADP-ribose chains on specific target proteins[309]. Following DNA damage, PARPs are highly activated inducing an increased level of poly ADP-ribose and a decreased level of free ATP, which are detected as a signal for necrosis. However, the levels of the PARP are controlled by the Caspases, a class of cysteine proteases that cleave the PARP and limit poly ADP-ribose production[310]. Defective Caspase activity will led to increased poly ADP-ribose concentration and to the activation of the necrosis pathway[311]. ADP-ribosylation plays also a role in chromatin remodelling and mRNA expression regulation, since this modification can occur on histones and on mRNA binding proteins, which may result in gene silencing[312]. However, ADP-ribosylated RNAs have not been identified yet and NUDT9 hydrolyse has not shown any RNA-decapping ability *in vitro* nor *in vivo*.

2.3.9.3. NUDT13 and NUDT17 hydrolases are NAD pyrophosphatases

NUDT13 is a Nudix hydrolase of 352 amino acids which localises in the mitochondria and it is conserved in mammals and plants, but not in yeast or flies[313]. It has the ability to break the pyrophosphate bond within the NAD substrates which is a rather redundant role in the Nudix family[313]. However, NUDT13 has no decapping activity towards any of the alternative caps tested *in vitro* by Sharma et al[160]. Interestingly, it was downregulated in metastatic cells of patients affected by gastric cancer together with the Major Facilitator Superfamily Domain containing 4 (MFSD4) protein. Here, the downregulation of these factors correlated with increased motility and invasiveness[314]. Thus, the increased levels of mitochondrial NAD may contribute to tumorigenic tendencies in particular cellular environment.

NUDT17 is a Nudix hydrolase of 328 amino acids, which localises in the peroxisomes. It is conserved in mammals and in plants, but it is not expressed in yeast, flies or in prokaryotic cells. Sharma et al. have reported no decapping activity on various alternative-capped RNAs, however it has the ability to remove the canonical m^7G-cap[160,250]. As for other members of the Nudix family, additional studies will shed a light on the functional role of the poor characterised Nudix proteins in different organisms.

2.3.9.4. NUDT14 and NUDT22 proteins are UDP-glucose/galactose pyrophosphatases

NUDT14 is a cytoplasmic Nudix member of 222 amino acids, which is conserved in worms, flies, plants, and mammals[315]. It is involved in the cleavage of the pyrophosphate bond within the UDP-glucose moiety *in vitro* and *in vivo* (Figure 29).

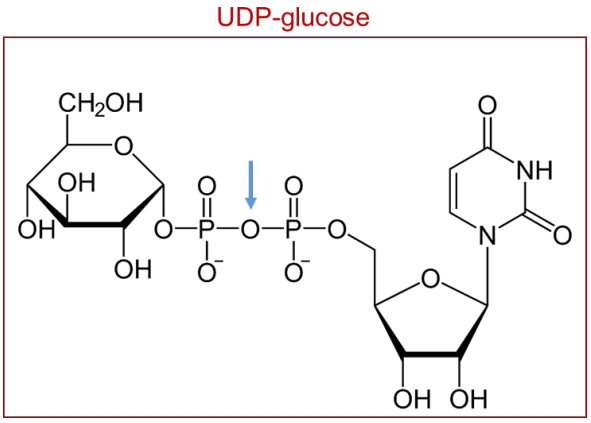


Figure 29. NUDT14 hydrolases UDP-glucose.
UDP-Glucose is hydrolysed by the NUDT14 protein within the diphosphate bond and is indicated by the blue arrow.

The UDP-glucose is the donor of glucose for the storage of glycogen in mammals and the presence of NUDT14 guarantees the release of glucose for this purpose[315]. Similarly in plants, glucose is stored as starch and depends in the activity of an orthologue of the human NUDT14 called AtASPP[164,316]. In *E. coli*, glycogen storage is assured by the orthologue ADP glucose pyrophosphatase NudF, which controls the glucose levels in the cell[317].

Interestingly, Wang et al have recently reported that Dengue virus, bacteria, yeast, mouse and human cells can load UDP-glucose caps on RNAs as alternative caps; however further analysis is necessary for understanding the fate of UDP-glucose capped RNAs and the metabolic and gene expression consequences[293].

NUDT22 is a nuclear protein of 303 amino acids, which is conserved in plants and mammals but no orthologue has been found in flies. It is a member of the Nudix family involved in the hydrolysis of the diphosphate bond within the UDP-glucose or galactose[318]. In plants, the

orthologue of NUDT22 hydrolyses UDP-mannose posing the question whether the human protein can hydrolyse different UDP-sugars[319]. NUDT22 acts and localises differently from the NUDT14, which recognises only UDP-glucose and localises in the cytoplasm. The human protein actually displays a pyrophosphatase activity not only towards UDP-glucose but also on UDP-galactose[318]. The UDP-sugars levels in the cell contribute to glycolysis, gluconeogenesis and protein glycosylation indicating that their turnover may affect important pathways.

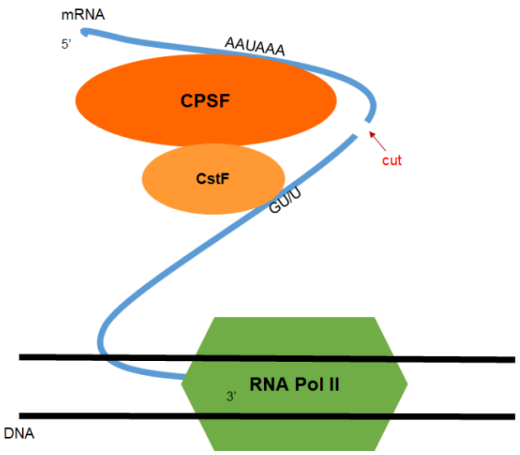
2.3.10. NUDT21 hydrolase is the Cleavage and Polyadenylation Specificity Factor 5 (CPSF5) that contributes to the removal of the poly(A) tail

NUDT21 hydrolyse is also known as the Cleavage and Polyadenylation Specificity Factor subunit 5 (CPSF5). It is a 227 amino acid protein, which is part of the Cleavage Factor Im (CFIm) complex. NUDT21 protein is a member of the Nudix family, which is highly conserved in metazoans. It takes part to the CFIm together with CSF6/CSF7 subunit which are required for the 3′end cleavage of the mRNAs and the subsequent poly(A) tail attachment.

Poly(A) tails are arrays of 100-250 adenosine added at the 3′ end of mRNAs in a co- or post-transcriptional manner. Poly(A) tails are 70-90 nucleotides long in yeast, while are longer in mammals. Tails are appended by template-independent poly(A) polymerases (PAPs) and recognised by cleavage and polyadenylation specificity factors. Despite the simplicity, they play a crucial role in multiple aspects of transcripts life[320].

As soon as the mammalian RNA Pol II transcribes the polyadenylation signal hexamer consensus sequence AAUAAA within the open reading frame, the CPSF complex binds to it and recruits the Cleavage and Stimulator Factor (CstF) (Figure 30A).

A.



B.

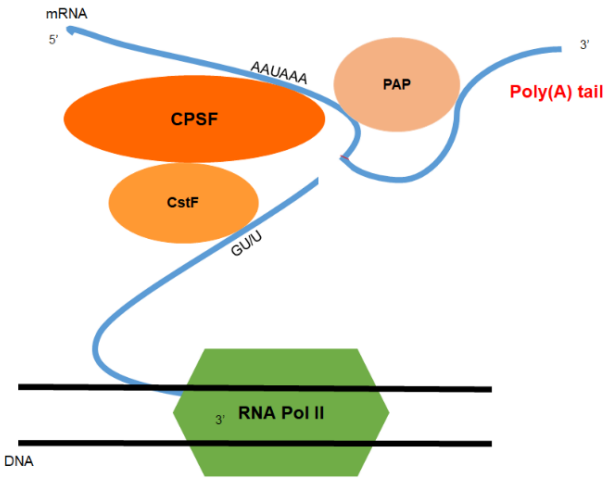


Figure 30. Polyadenylation at the 3′end of mRNAs.

A. Cleavage of pre-mRNA at 20 nt downstream of the consensus sequence AAUAAA by CPSF-CstF complex. **B.** PAP and PABPN1 bind to the nascent stretch and catalyse the addition of adenosines (Adapted from Russel PJ, 3rd edition 2014 by Steven M. Carr)[321].

The mRNA is cleaved 20 nucleotides downstream the AAUAAA sequence, leaving a free 3′OH end. The PAP protein is recruited and activated by the interaction with CSPF and CstF1. PAP transfers the first 12 AMPs starting from the 3′OH target site[322]. The poly(A) polymerase binding protein nuclear 1 (PABPN1) binds at the minimal poly(A) stretch and new molecules are recruited, forming a spherical particle on the growing poly(A) tail (Figure 30B). The spherical complex drives the interaction with the 5′cap structure and monitors the poly(A) tail length. After the synthesis of around 250 nucleotides, the addition of further adenosines is inhibited because PAP activity decreases and dissociate from the complex[322,323]. Once the poly(A) tail is attached to the mRNA, different binding proteins either drive it to the cytoplasm to start translation or induce nuclear mRNA decay.

The molecular mechanism that leads towards mRNA stabilization and export or transcript degradation depends on the length of the poly(A) tail. Nonetheless, studies on the human system have demonstrated that various conditions, such as cell type considered, cell cycle phase or environmental conditions may influence nuclear export or mRNA decay rates in addition to poly(A) tails length[324].

In particular, NUDT21 (CPSF5) binds to an upstream conserved sequence UGUA which induces the recruitment of CSPF6/CPSF7 containing the Serine/Arginine (SR) sequence, which is a C-terminal serine-arginine rich region able to bind to other SR-proteins[325]. Here, the CPSF complex is recruited by the SR motif of the NUDT21-CSPF6/CPSF7 complex and binds the conserved AAUAAA polyadenylation signal, upstream of the cleavage site[326]. Then, the CstF is activated and binds to the GU- or U-rich element downstream of the poly(A) site[327]. When the cleavage occurs, CPSF remains bound upstream to the cleavage site and recruits PAP to start the polymerization of poly(A) tail[328]. In HeLa cells, NUDT20 localises in the nucleus, in punctuate structures called paraspeckles[329]. It is still controversial whether it is involved in the nuclear export of the mRNA since only its cofactor CPSF6 has been detected in the cytoplasm and associated with the nuclear export factor 1 (NXF1) through the SR domain in HEK293 cells[330].

Various studies have shown that the suppression of NUDT21 expression in HeLa cells provokes a shift in the polyadenylation site recognition, driving towards an alternative polyadenylation[331]. This process is normally occurring in the cell and generates mRNAs with variable 3′ ends. Thus, alternative polyadenylation contributes to regulate gene expression in a tissue-specific manner as well as during development[332]. One of the best example is the human

IgM heavy chain gene which switches from a membrane protein to a secreted one according to the poly(A) site used[333]. In this case, CPSF5 is involved in regulating the recognition of the upstream poly(A) sequences by CPSF and CstF that would define the terminal exon and influence exon splicing. Indeed, CPSF5 seems to provoke the looping of the AAUAAA signal and preventing the usage of the upstream poly(A) sequence in favour of the downstream one[334,335].

Such alternative polyadenylation system has been detected in other eukaryotes, including *Drosophila melanogaster*. The cleavage and polyadenylation complex in flies contains subunits that share high similarity to the human counterparts and also influences the generation of variable 3' end transcripts (Figure 31).

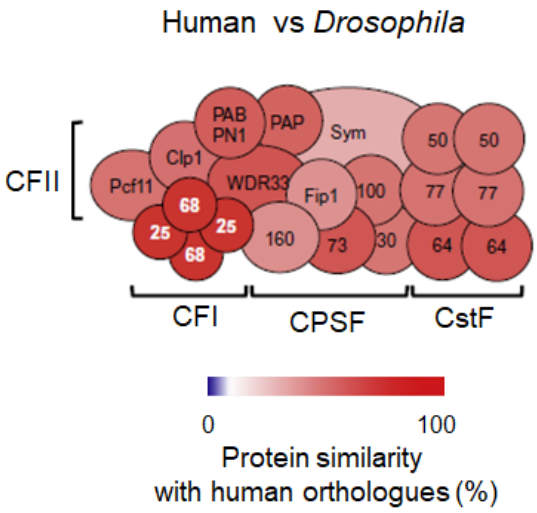


Figure 31. Fly cleavage and polyadenylation complex shares high similarity to the human one.
The cleavage and polyadenylation complex in *Drosophila* is highly conserved and shares high similarity with the human one. It is composed by the CFI (Cfi25=Cpsf5, Cfi68=Cspf6) and CFII, CSPF and CstF components. The bar below the picture indicates the percentage of similarity of the fly components with the human one going from 0 (blue) to 100 (red). CPSF5 and CPSF6 share more than 70% of similarity with the human counterparts. (Adapted from Vallejos-Bajer et al.2017[336]).

During embryogenesis, the expression pattern of the CPSF, CSFI and CstF decreases, however as soon as the central nervous system starts to develop, the two components of the CFI, that are Cpsf5 and Cpsf6, are highly expressed in the CNS[336]. When Cspf5 is depleted by RNAi silencing, genes known to have APA sites and coding for neuronal differentiation (Abdominal-A(abd-A), Abdominal-B (Abd-B), Neuroglian (Nrg), brain tumour (brat)), post transcriptional regulating proteins (Adenosine deaminase acting on RNA (Adar)) and RNA binding proteins (alan shepard (shep), IGF-II mRNA-binding protein (Imp)) showed shortened 3' ends[336]. These results are in line with mammalian studies where silencing *cpsf5* induces the shortening of the

3' ends of genes such as *cyclin D1* which is associated with cellular proliferation and tumour growth in glioblastoma[337].

In conclusion, CPF5 is a conserved factor in metazoans that is required for the regulation of the alternative polyadenylation in different cell types and during development since it can control 3' ends of transcripts, which may be sites for miRNAs or regulatory proteins binding. Further analysis may help in uncovering the regulatory system of the PAP and propose CSPF5 as a potential target for drug treatment against different types of diseases, which depend on PAP.

Given the fact that most of the Nudix proteins are conserved both in prokaryotes and in eukaryotes and that they share a common pyrophosphatase activity, hundreds of studies have tried to characterize their function applying sophisticated technologies. In the last three decades, *in vitro* and *in vivo* analysis have demonstrated that some of the members act as alternative decapping enzymes in the presence on non-canonical capped-RNAs. These results pointed out that DCP2 is not the only decapping enzyme, which have completely revolutionized the ancient idea of RNA turnover and has allowed to investigate the interplay between the metabolic conditions of the cell with gene expression regulation. Nonetheless, it still is hard to detect non-canonical capped-RNAs since their low availability *in vivo* suggesting that the sensitivity of the tools used so far need to be further implemented.

OBJECTIVES OF MY THESIS

Epigenetic modifications can occur along the mRNA sequence at internal sites as well as at the 5' and 3' ends and a wide variety of effectors are involved in the deposition, recognition and removal of these marks. The 5' cap is one of the most conserved modification on eukaryotic mRNAs and the mRNA interacting proteins contribute to modulate transcripts stability, splicing and translation despite the overall pictures is still uncomplete due to the complexity of the interactions.

As shown in my introduction, these epigenetic marks play an essential role in mRNA biology and organisms physiology. In this context, the global objective of my thesis was to contribute to a better understanding of the molecular mechanisms governing epigenetic modifications occuring at the 5' mRNA cap.

A first objective was to decipher the role of the structurally conserved methyltransferase PCIF1 in *Drosophila melanogaster*. Strikingly, this protein is catalytically dead in flies, but still shows the binding to the CTD of the RNA Pol II offering the opportunity to answer to the question whether the catalytic dead Pcif1 has a function using molecular genetics approaches in a living organism.

The second part of my thesis aimed at the genetic screening of the 12 *Nudix* orthologues conserved in flies in order to detect a potential role of this class of enzymes and further investigate their putative ability to hydrolyze non-canonical capped-RNAs *in vitro* or *in vivo*. To this end, I took advantage of tissue-directed gene silencing to impair the expression of the *Nudix* orthologues and study the resulting phenotypes in living flies.

RESULTS

Chapter 1: Catalytic dead Pcif1 regulates gene expression and fertility in *Drosophila*

1.1. Presentation of the two papers included in this manuscript (Pandey et al.,Cell Reports and Franco et al., accepted for publication on RNA journal)

The first part of my thesis project has been focused on the functional characterization of the conserved PCIF1 protein using *Drosophila melanogaster* as a genetic model.

The PCIF1 protein was first identified by Fan et al. as the Phosphorylated C-terminal domain interacting protein 1, able to bind the phosphorylated C-term of the RNA Pol II *in vitro* through its WW-domain[76]. The human PCIF1 protein is 704 amino acids long and is characterized by an N-terminal WW domain, a central helical domain able to bind to the m^7G capped mRNA and C-terminal domain, involved in the methylation process. PCIF1 has been further characterized as the methyltransferase depositing a methyl group in position 6 of the adenosine ring if the first transcribed nucleotide is an adenosine, forming the so called m^6Am modification[52,54,74]. Recent works agree on the fact that m^6Am modification contributes to stabilize mRNA but it is not clear whether it promotes or represses translation [52,54]. In mouse, the m^6Am modification is enriched at the TSS of poly(A) tailed mRNAs. Mutation of the catalytic motif does not affect mouse viability nor fertility but induces reduced body weight. Moreover, RNA deep sequencing in knock out mice showed a significant dysregulation in pseudogenes associated to the Y chromosome in mouse testis[78].

Interestingly, Pandey et al. showed that the fly orthologous protein Pcif1 is expressed in flies and structurally conserved. However, a single amino acidic substitution in the NPPF motif renders fly Pcif1 catalytically inactive. In this study, I performed the immunostaining of *Drosophila* ovaries in order to determine the subcellular localization of the fly Pcif1 protein. I showed that, as observed for its human counterpart, the fly Pcif1 protein is nuclear. Moreover, it is excluded from the nucleolus as shown through co-immunostaining of Fibrillarin, a nucleolar protein (Pandey et al., 2022 Fig 4D, Annex 1). In addition, Pandey and colleagues showed that the fly Pcif1 is able to bind to the CTD of RNA Pol II by pull down experiments using fly WW-domain. These results suggested that fly Pcif1 has conserved this feature since previous findings demonstrated that mammalian PCIF1binds the CTD of the Pol II in human cells[76,77,338].

In the continuity of this first set of results, I further unveiled the function of the Pcif1 protein in fly by generating CRISPR-Cas9 knock out mutants (Franco et al., under consideration for publication in RNA Journal, Manuscript below). In this work, we used a CRISPR/Cas9 mutagenesis strategy and obtained two independent alleles that we named *Pcif1*^{M6stop} and

Pcif1$^{F9\Delta NLS}$ (Franco et al. Fig. 1). The *Pcif1*^{M6stop} allele carries a frame-shift mutation that generates a premature stop codon. The *Pcif1*$^{F9\Delta NLS}$ mutant allele carries an in frame deletion of the predicted Nuclear Localization Signal (NLS) which restores the open reading frame. First, I analyzed Pcif1 expression in fly heads protein extracts and observed that the wild type Pcif1 is expressed as a protein of the expected sized of 130 kDa. I did not detect any signal in the mutant *Pcif1*^{M6stop} strongly suggesting that it is a null mutant which is in accordance with the genetic alteration producing a premature STOP codon. A shorter protein of 120 kDa was detected in *Pcif1*$^{F9\Delta NLS}$, which is consistent with the in-frame deletion of the NLS sequence. I then performed immunostaining and showed that, in contrast to the wild type situation where Pcif1 is nuclear, in *Pcif1*$^{F9\Delta NLS}$ mutants Pcif1 is cytoplasmic and excluded from the nucleus. This indicates that the NLS is indeed functional and strictly required for Pcif1 nuclear localization.

Phenotypic analyses were then performed on mutants outcrossed for 10 generations over control flies (*Pcif1*^{BG02557}, Franco et al., Fig. 1). The *Pcif1*^{M6stop} mutants did not show any morphological defect and were viable although around 20% of embryos failed to hatch in the progeny from *Pcif1*^{M6stop} males and females crosses. Moreover, males, but not females, showed a significantly reduced body weight. Fertility analysis demonstrated that *Pcif1*^{M6stop} mutants were less fertile than controls; this decreased fertility was particularly strong in females although mutant ovaries presented normal morphology when compared to controls. Interestingly, *Pcif1*$^{F9\Delta NLS}$ mutant flies showed similar or even stronger phenotypes compared to *Pcif1*^{M6stop} mutants. Indeed, both males and females had a significant reduced body weight. Moreover, *Pcif1*$^{F9\Delta NLS}$ mutants not only displayed a reduced fertility, but also a shorter life span, suggesting that the cytoplasmic Pcif1 has deleterious effects in the absence of the nuclear protein.

To decipher the effect of *Pcif1*mutations in gene expression regulation, I prepared fly mutant ovaries for the analysis of gene expression by RNA deep sequencing that was performed in Pr. RS Pillai's laboratory (Franco et al., Fig. 2). This analysis pointed out a set of common dysregulated genes in both mutants with a higher number of dysregulated genes in *Pcif1*$^{F9\Delta NLS}$ than in *Pcif1*^{M6stop} mutants. GO term analysis showed enrichment in genes associated with mitochondrial function in both mutants. Although we did not find any dysregulated gene that could be directly associated with the decrease in fertility of *Pcif1* mutants, defective mitochondrial function may impact gametogenesis and more particularly oogenesis[339]. In order to analyze the protein level of expression of mitochondrial markers in *Pcif1* mutants, I performed western blot using protein extracts from fly ovaries and the antibody against ATP5A mitochondrial protein (Franco et al., Suppl. Fig S2A, B). However, the ATP5A expression in

mutant ovaries was comparable to that observed in control ovaries. In addition, EM analysis of mutant ovaries performed at the ISBG platform (https://www.ibs.fr/research/research-groups/methods-and-electron-microscopy-group/schoehn-team/group-equipment/article/microscope-tecnai-12-fei) showed no remarkable defects in mitochondrial network (Franco et al., Suppl. Fig S2 B). Thus, lack of nuclear Pcif1 may slightly affect gene expression and subsequently of proteins that by acting in common networks may globally influence fly fertility and physiology.

Given the nuclear localization of the *Drosophila* Pcif1 protein and the fact that it is able both to bind the CTD of the RNA Pol II[78] and to co-immunoprecipitate in a protein complex with RNA Pol II[145], we investigated whether it associates to chromatin by performing immunostaining of the giant larval polytene chromosomes from the fly salivary glands in collaboration with Pr. E. Brasset and N. Gueguen (iGRED, Univ. of Clermont-Ferrand). They showed that Pcif1 was distributed along the euchromatic regions of the polytene chromosomes at the level of interbands known to be actively transcribed (Franco et al., Fig. 3A). In order to correlate Pcif1 association to euchromatin to its potential role in gene expression regulation, I tested the effect of *Pcif1* mutations on Position Effect Variegation (PEV) (Franco et al., Fig. 3). In this sensitized background, the wild type euchromatic *white* gene, which is responsible for the red pigmentation of the fly eye, is translocated towards the vicinity of heterochromatin, inducing a variegated eye color due to the stochastic silencing of the w^{m4h} allele that is clonally inherited in daughter cells[340–342]. I showed that reduced dosage of nuclear Pcif1 in heterozygous mutants, or hemizygous deficiencies of the *Pcif1* locus, induced a significant suppression of w^{m4h} silencing, suggesting that Pcif1contributes to heterochromatin-dependent gene silencing.

With this work, we aimed to shed a light on the evolutionary conserved PCIF1 protein apart from any methyltransferase activity using *Drosophila* as a genetic model in which this protein is catalytically inactive. We found that fly Pcif1 is nuclear and that the NLS is required for this nuclear localization. Moreover, we showed that nuclear Pcif1 was required for normal gene expression, full fertility and fly physiology. Indeed, total lack of the protein or its mis-localization in the cytoplasmic compartment resulted in gene dysregulation, decreased fertility and loss of weight. Thus, despite the lack of catalytic activity, we demonstrated that Pcif1 has a functional role in fly physiology. As stated above, we suggest that fly Pcif1 protein contribute to gene expression in interaction with the CTD of the RNA Pol II. Since we mostly observed gene derepression by either RNAseq or PEV experiments, Pcif1 may mainly contribute to RNA Pol II inactivation by directly preventing the recruitment of other important factors required for its activation and/or contributing to chromatin opening.

Investigating protein-protein interaction involving Pcif1 and generating mutants unable to interact with RNA Pol II would be very promising strategies to further understand Pcif1 activity in gene expression regulation. In addition, we could investigate Pcif1 binding to specific promoters by ChIP method in order to determine if it regulates the expression of specific set of genes in a tissue-specific manner or according to variable stress conditions.

The catalytic-dead Pcif1 regulates gene expression and fertility in *Drosophila*

Giulia Franco[1], Emmanuel Taillebourg[1], Elena Delfino[2], David Homolka[2], Nathalie Gueguen[3], Emilie Brasset[3], Radha Raman Pandey[2], Ramesh S. Pillai[2] and Marie-Odile Fauvarque[1,*].

[1]Univ. Grenoble Alpes, CEA, INSERM, BGE, F-38000 Grenoble, France. [2]Department of Molecular Biology, Science III, University of Geneva, 30 Quai Ernest-Ansermet, CH-1211 Geneva 4, Switzerland. [3] iGReD, Université Clermont Auvergne, CNRS, INSERM, Faculté de Médecine, 63000 Clermont-Ferrand, France.

Running Title: Catalytic-dead Pcif1 in gene expression

Keywords: Pcif1, m^{6}Am, m^{6}A, N^6-methyladenosine, Position-effect Variegation (PEV), Transcription

* Contact :
Marie-Odile Fauvarque
marie-odile.fauvarque@cea.fr

ABSTRACT

Eukaryotic mRNAs are modified at the 5′ end with a methylated guanosine (m^7G) that is attached to the transcription start site (TSS) nucleotide. The TSS nucleotide is 2′-*O*-methylated (Nm) by CMTR1 in organisms ranging from insects to human. In mammals, the TSS adenosine can be further N^6-methylated by RNA polymerase II Phosphorylated CTD-Interacting Factor 1 (PCIF1) to create m^6Am. Curiously, the fly orthologue of mammalian PCIF1 is demonstrated to be catalytic-dead, and its functions are not known. Here we show that *Pcif1* mutant flies display a reduced fertility which is particularly marked in females. Deep sequencing analysis of *Pcif1* mutant ovaries revealed transcriptome changes with a notable increase in expression of genes belonging to the mitochondrial ATP synthetase complex. Furthermore, the Pcif1 protein is distributed along euchromatic regions of polytene chromosomes and the *Pcif1* mutation behaved as a modifier of position-effect-variegation (PEV) suppressing the heterochromatin-dependent silencing of the *white* gene. Similar or stronger changes in the transcriptome and PEV phenotype were observed in flies that expressed a cytosolic version of Pcif1. These results point to a nuclear co-transcriptional gene regulatory role for the catalytic-dead fly Pcif1 that is probably based on its conserved ability to interact with the RNA polymerase II C-terminal domain.

INTRODUCTION

RNA polymerase II transcripts are capped at the 5′ end with a 7-methylguanosine (m^7G) moiety which is attached via an inverted 5′-5′ triphosphate bridge to the transcription start site (TSS) nucleotide (Shatkin, 1976). This cap0 structure is seen in lower eukaryotes like yeast. In organisms ranging from insects (Haussmann et al., 2022) to human (Belanger et al., 2010), the TSS nucleotide is 2′-*O*-methylated (Nm) on the ribose by CMTR1 to create the cap1 structure (Inesta-Vaquera and Cowling, 2017). In viral and mammalian RNA caps, the TSS adenosine (Am) can be further methylated at the N^6 position to form the dimethylated m^6Am (Moyer and Banerjee, 1976; Wei et al., 1975). The activity responsible for this m^6A methylation was partially purified from human HeLa cell extracts (Keith et al., 1978), and 40 years later identified to be the RNA polymerase II Phosphorylated CTD-Interacting Factor 1 (PCIF1) (Akichika et al., 2019).

PCIF1 was originally identified as a nuclear protein that uses its N-terminal WW domain to interact with the phosphorylated C-terminal domain (CTD) of RNA polymerase II (Fan et al., 2003). PCIF1 interacts with the Ser5-phosphorylated CTD of the RNA Pol II, suggesting a recruitment to the nascent RNA early during transcription (Akichika et al., 2019; Boulias et al., 2019; Sendinc et al., 2019). Presence of the m^6Am modification on the RNA cap is associated

with RNA stability in human cells (Boulias et al., 2019) and mouse testes (Pandey et al., 2020). Additionally, it is shown to also regulate cap-dependent translation (Akichika et al., 2019; Sendinc et al., 2019). We previously reported that mice lacking *Pcif1* are viable and fertile, with no dramatic phenotypes, except a slightly reduced body weight (Pandey et al., 2020). Interestingly, insect Pcif1 orthologues have mutations in the catalytic motif that are predicted to render them inactive as RNA methylases. We demonstrated that purified *Drosophila* Pcif1 is inactive as a methylase in vitro, but it retained a WW domain capable of specifically interacting with Ser5-phosphorylated CTD of RNA polymerase II (Pandey et al., 2020). This makes the fly system a unique model to investigate the non-catalytic role of Pcif1.

In this study, we generated and analysed two independent fly mutants of *Pcif1* (CG11399) to examine the *in vivo* role of *Drosophila melanogaster* Pcif1. Our results show that the protein acts in the nucleus to regulate levels of hundreds of transcripts and is required for female fertility and to a lesser extent for male fertility. We also observed that Pcif1 localised to polytene chromosome interbands known to be transcriptionally active sites and open euchromatin regions. Moreover, we observed that loss of the nuclear Pcif1 protein modified position-effect variegation (PEV) by suppressing the chromatin-dependent silencing of the *white (w)* gene in w^{m4h} mutants. We propose a model where nuclear catalytic-dead fly Pcif1 associates with RNA pol II CTD to regulate gene expression.

RESULTS

Fly *Pcif1* mutants display reduced body weight and decreased fertility

Drosophila Pcif1 is a multi-domain protein of 920 amino acids with a predicted N-terminal WW-domain and a C-terminal methyltransferase domain (Figure 1A). A central helical domain homologous to an alpha helix-rich region in the crystal structure of the human PCIF1 (Akichika et al., 2019) is indicated. A putative nuclear localization signal (NLS) is also identified between the WW and helical domains.

To investigate the physiological role of *Drosophila* Pcif1, we used the CRISPR-Cas9 genome engineering tool to generate a genomic deletion in the region encoding for the NLS in the w^{1118} background. Two mutated alleles were identified: $Pcif1^{M6stop}$ and $Pcif1^{F9\Delta NLS}$. The deletion in the $Pcif1^{M6stop}$ allele induces a frameshift ending in a premature stop codon (Figure 1B). On the other hand, the $Pcif1^{F9\Delta NLS}$ allele has an in-frame deletion of the NLS, generating a protein of 829 amino acids (Figure 1B). To clean up any potential off-target mutations, the selected alleles were backcrossed for up to ten generations with a fly line carrying a P-element inserted in the first intron of *Pcif1* ($P\{GT1\}CG11399^{BG02557}$), also referred to as the $Pcif1^{BG02557}$

line (Figure 1C). This P-element contains a *miniwhite [mw+]* gene that confers a red eye colour to the flies in a w^{1118} context, while the *Pcif1* CRISPR/Cas9 mutants generated in a w^{1118} background have white eyes, thus allowing to trace the presence of the *Pcif1* mutations along generations (see Materials and Methods). Thus, the two *Pcif1* mutant alleles used for our studies are in the w^{1118}; $Pcif1^{BG02557}$ background, with the original $Pcif1^{BG02557}$ line hereafter referred to as control. Western blot analysis of fly head lysates indicated that the control flies expressed the Pcif1 protein at similar levels as the w^{1118} strain, while the absence of detectable signal in the homozygous $Pcif1^{M6stop}$ flies confirmed these to be null mutants (Figure 1D). In contrast, flies homozygous for the $Pcif1^{F9\Delta NLS}$ allele expressed a slightly smaller protein, which is consistent with the deletion of the NLS sequence. Immunofluorescence analysis showed that the wild type protein was nuclear in the control fly ovaries, while the Pcif1$^{\Delta NLS}$ protein failed to enter the nucleus and was consequently stranded in the cytoplasm (Figure 1E). In heterozygous, $Pcif1^{\Delta NLS}$/+ ovaries, immunostaining revealed the presence of the Pcif1 protein in both the cytoplasmic and the nuclear compartments, most probably corresponding to the mutated and wild-type proteins, respectively (Figure 1E).

Adult $Pcif1^{M6stop}$ mutant flies did not show any obvious morphological defects (Figure 1F) and displayed normal longevity (Figure 1G). However, about 20% of eggs issued from crossing $Pcif1^{M6stop}$ mutant siblings did not hatch, most probably due to the lack of egg maturation or fertilization (white eggs) (Supplementary Fig. S1A). We also observed that $Pcif1^{M6stop}$ males, although not the females, had a significant reduced body weight (Figure 1H). Further investigation revealed that homozygous $Pcif1^{M6stop}$ mutants showed a significantly reduced fertility that was particularly strong in females (Figure 1I), although examination of dissected fly ovaries did not reveal any gross morphological differences (Figure 1F). Hemizygous females carrying the $Pcif1^{M6stop}$ allele over a deficiency covering the *Pcif1* locus similarly displayed a strongly decreased fertility (Supplemental Fig. 1B).

Interestingly, $Pcif1^{F9\Delta NLS}$ mutant flies displayed similar or even stronger phenotypes compared to $Pcif1^{M6stop}$ mutants. Both $Pcif1^{F9\Delta NLS}$ homozygous males and females displayed a reduced body weight and a significantly decreased fertility, which was, as observed for $Pcif1^{M6stop}$ particularly strong in females compared to a moderate effect in males (Figure 1H,I, Supplemental Fig. 1B). In contrast to $Pcif1^{M6stop}$, the $Pcif1^{F9\Delta NLS}$ mutants had a significantly reduced longevity (Figure 1G). Enhanced physiological defects observed in $Pcif1^{F9\Delta NLS}$ flies indicate that the cytoplasmic protein cannot compensate for the lack of Pcif1 nuclear protein. Moreover, the Pcif1 protein may have a deleterious effect when present in the cytoplasm in the absence of the nuclear protein.

Taken together, these results show that *Pcif1* mutants are viable but display physiological defects such as reduced body weight and decreased fertility. Female fertility decrease was notably observed in the two independent alleles outcrossed for 10 generation with the control line, as well as in trans-heterozygous mutants over a deficiency. Thus, this phenotype can be attributed to the lack of nuclear *Pcif1* rather than to a genetic background effect.

Loss of fly *Pcif1* upregulates transcripts involved in mitochondrial ATP synthesis

To understand the female fertility defect, we sequenced the mutant ovary transcriptome. We prepared total ovary RNAseq libraries from 5-day-old homozygous mutant $Pcif1^{M6stop}$, homozygous mutant $Pcif1^{F9\Delta NLS}$, and control female flies. Consistent with complete absence of the protein, there was a strong reduction in *Pcif1* transcript levels in the null mutant where the premature stop codon may induce degradation of the transcript by the Nonsense Mediated Decay (NMD) (Baker and Parker, 2004), while the levels were unchanged in the $Pcif1^{F9\Delta NLS}$ mutant (Figure 2A). Levels of several hundred transcripts were altered, with ~500 transcripts being upregulated and ~250 transcripts downregulated in the loss-of-function $Pcif1^{M6stop}$ mutant (Figure 2B, Supplementary Table S1). The altered genes are distributed across different chromosomes, with a slight overall upregulation of transcripts from the mitochondrial genome (Figure 2C). Interestingly, gene ontology analysis showed that nuclear encoded genes involved in proton transport and mitochondrial ATP synthesis were upregulated in the $Pcif1^{M6stop}$ mutant (Figure 3A, Supplementary Table S1).

Next, we used the homozygous $Pcif1^{F9\Delta NLS}$ mutant ovary transcriptome to examine if the observed transcript changes in the null mutant could be attributed to potential nuclear functions of Pcif1. The RNA sequence analysis from $Pcif1^{F9\Delta NLS}$ mutant ovaries revealed a much larger disruption of around 3000 transcripts, with an equal number being up- and down-regulated (Figure 2B, Supplementary Table S1). We attribute this larger disruption in the transcriptome to potential deleterious effects of the mis-localized $Pcif1^{\Delta NLS}$ protein in the cytoplasm (see above and Figure 1E). Significantly, a large part of the altered ovarian transcriptome in the $Pcif1^{M6stop}$ mutant overlapped with that seen in the $Pcif1^{F9\Delta NLS}$ mutant (Figure 2D). Among this shared set of altered genes are those involved in proton transport and mitochondrial ATP synthesis, and these were uniformly upregulated in both the null and the $Pcif1^{F9\Delta NLS}$ mutant fly ovaries (Figure 3B-C, Supplementary Table S1).

One of the transcript upregulated in $Pcif1^{M6stop}$ and $Pcif1^{F9\Delta NLS}$ mutant ovaries encodes the α subunit of the mitochondrial F1F0 ATP synthase of complex V (ATP5A synonym bellwether (blw) in *Drosophila*). Western blot analyses of protein extracts isolated from ovaries revealed

that the ATP5A protein was however not significantly upregulated (Supplementary Figure S2A). No obvious defects were detected on mitochondrial network of the *Pcif1*^{M6stop} null mutant by electron microscopy although we cannot exclude subtle morphological or functional differences at the level of this analysis. In the case of the *Pcif1*$^{F9\Delta NLS}$ mutant, we observed some enlarged mitochondria that may however results from the ectopic localization of the protein in the cytoplasm rather than the absence of nuclear Pcif1 protein. For this reason, a putative effect would not reflect a physiological role of Pcif1and was not further investigated (Supplementary Figure S2B).

Mammalian PCIF1 targets m^{7}G-capped RNA substrates with an adenosine as the transcription start site (TSS) nucleotide for co-transcriptional catalysis of the m^{6}Am modification, and such modified transcripts are preferentially stabilized in cell culture and mouse testes (Boulias et al., 2019; Pandey et al., 2020). However, we did not find any influence of the TSS nucleotide in regulating transcript levels in the mutant fly ovaries (Figure 3D). In conclusion, our results show that although catalytically inactive, fly Pcif1 has a functional role in regulating gene expression, and this depends on its nuclear localization. We propose that this collective regulation of a few hundred transcripts (either direct or indirect) may contribute to overall fly fertility.

Pcif1 binds to euchromatic regions along polytene chromosomes

Drosophila Pcif1 lacks the m^{6}Am RNA methylation activity found in its mammalian counterpart, but as observed for the mammalian protein, *Drosophila* Pcif1 is nuclear and has the ability to interact *in vitro* with the Ser5-phosphrylated C-terminal domain (CTD) of RNA polymerase II (Pandey et al., 2020). Our experiments above point to a role for Pcif1 in gene regulation, but the mechanism of action is unclear. To obtain further evidence for this nuclear role, we performed Pcif1 immunostaining on larval polytene chromosomes. These experiments showed that Pcif1 localised to polytene chromosome interbands which are known to contain other proteins characteristic of open euchromatin and to be transcriptionally active regions (Figure 4A). We propose that Pcif1, in association with RNA Pol II, contribute to gene expression regulation at transcriptionally active sites.

Loss of *Pcif1* suppresses position-effect variegation

The detection of Pcif1 at transcriptionally active euchromatic sites prompted us to use the sensitive position effect variegation (PEV) assay to investigate further whether loss of *Pcif1* has functional effects on gene regulation. In this experiment, a variegated eye phenotype is induced by the w^{m4h} allele, where the wild-type *white* gene (w^+) that specifies red eye colour, is

translocated from its normal location in an euchromatin region on the X chromosome to a pericentric heterochromatin zone (Reuter and Wolff, 1981; Wang et al., 2014). This induces a clonally inherited epigenetic silencing of the w^+ gene that results in a variegated eye due the different levels of expression of the gene in individual photoreceptors that constitute the compound eye in flies. This translocation renders White protein expression and subsequent eye coloration very sensitive to subtle changes in transcription of the w^+ gene.

As expected, we observed that hemizygous w^{m4h}/w^{1118} female flies displayed a variegated eye phenotype instead of the uniform red eye observed in w^+/w^+ females (Figure 4B). When placed in the heterozygous $Pcif1^{M6Stop}/+$ background, we observed a significant suppression of PEV, resulting in a uniformly dark red eye colour (Figure 4B,C). We confirmed that this is specifically due to loss of nuclear action of Pcif1, as an identical observation was made in the heterozygous $Pcif1^{F9\Delta NLS}/+$ background (Figure 4B,C). As an additional support, we showed that a similar suppression of PEV was induced in a different genetic background where flies carried a large genomic deletion covering the *Pcif1* locus at hemizygous state (w^{m4h}/+;*Df(3L)BSC563*/+) (Figure 4B,C). Moreover, two other independent genomic deletions (*Df(3L)BSC797*/+ and *Df(3L)BSC7452*/+) encompassing the *Pcif1* gene locus also displayed similar suppression of w^{m4h} silencing although with a lower penetrance in the case of *Df(3L)BSC452*/+ (Figure 4B).

These results show that, in addition to being present at transcriptionally active euchromatic sites, Pcif1 participates directly or indirectly to the epigenetic phenomenon of PEV.

DISCUSSION

Our genetic analyses revealed that fly *Pcif1* is not essential for life and normal development, while a strong and significant reduction in female fertility was observed in the two mutant alleles generated in this study. RNAseq analysis showed that absence of fly Pcif1 resulted in alteration of hundreds of transcripts. A set of upregulated transcripts in Pcif1 mutants ovaries corresponded to nuclear genes encoding mitochondrial proteins involved in the respiratory chain. Functional respiratory chain components are required in the late germarium for the proper replication of mitochondrial DNA (mtDNA) and normal mitochondrial biogenesis (Wang et al., 2019). We did not detect neither significant changes in the expression level of one of these mitochondrial proteins (ATP5A) by Western analysis > We did not detect neither significant changes in the expression level of one of these mitochondrial proteins (ATP5A) by Western analysis nor morphological defects in the mitochondrial network of $Pcif1^{M6stop}$; however in $Pcif1^{F9\Delta NLS}$ some mitochondria appeared slightly enlarged which might

be a consequence of Pcif1 mis-localization while not reflecting the physiological function of Pcif1. Nevertheless, because the observed transcript upregulation was only around 20%, this difference might be difficult to detect at the protein level by Western analysis and we may still speculate that subtle alteration of mitochondrial function may prevent proper egg maturation or development (Hill et al., 2014; Wang et al., 2019).

Analysis of another fly mutant where the nuclear localization signal (NLS) in Pcif1 is specifically deleted showed even larger number of transcript changes. We propose that this is a compounded effect of both direct loss of Pcif1-mediated nuclear functions and potential deleterious effect of an RNA-binding protein that is stranded in the cytoplasm. In fact, compared to the complete *Pcif1* null and control flies which have similar life spans, the $Pcif1^{F9\Delta NLS}$ flies have a strikingly reduced longevity. This is likely due to the potential toxicity caused by the cytosolic mis-localization of the protein.

This leads to the most important question of how the catalytic dead Pcif1 protein might directly regulate gene expression in the nucleus. Hypophosphorylated RNA pol II is recruited to gene promoters, which at initiation of transcription becomes phosphorylated on Ser5 within the consensus heptapeptide YSPTSPS repeats in the CTD of the largest subunit of the RNA pol II (Buratowski, 2009). This modification helps to recruit RNA capping enzymes that use WW domains to interact with the phosphorylated CTD (Galloway and Cowling, 2019; Ghosh et al., 2011; Haline-Vaz et al., 2008). Later during elongation, phosphorylation at Ser2 helps to recruit 3′ end processing factors (Licatalosi et al., 2002; Meinhart and Cramer, 2004) and histone methyltransferases (Hampsey and Reinberg, 2003), and this is accompanied by dephosphorylation at Ser5. Thus, the dynamic phosphorylation of the RNA pol II CTD can impact gene expression at multiple steps. Human PCIF1 (Akichika et al., 2019; Boulias et al., 2019; Pandey et al., 2020; Sendinc et al., 2019) and fly Pcif1 (Pandey et al., 2020) both interact with Ser5-phosphorylated CTD *in vitro*. Consistently, we observed that Pcif1 localised along the polytene chromosomes at the level of transcriptionally active sites. Human PCIF1 overexpression is shown to regulate transactivation of gene reporters and this is dependent on its ability to interact with the CTD (Hirose et al., 2008). Interestingly, the interaction of recombinant WW domain of human PCIF1 -that lack any catalytic activity- inhibits the dephosphorylation of Ser5 in the RNA pol II CTD *in vitro* (Hirose et al., 2008), suggesting a potential mechanism by which the fly protein may influence gene expression by similarly interfering with RNA pol II phosphorylation status. Moreover, our study demonstrated that loss of Pcif1 or mis-localization of Pcif1 to the cytoplasm influenced the epigenetic phenomenon of position effect variegation (PEV). We therefore propose that the nuclear effects of fly Pcif1 is dependent on its WW domain and its ability to interact with Ser5-phosphorylated CTD of RNA

pol II at transcriptionally active sites. In fact, fly Pcif1 was found in immunoprecipitated nuclear complexes consisting of the RNA pol II subunits and the cap-binding complex (CBP) that binds nascent RNA (Kachaev et al., 2019). Evidently, such a gene regulatory role would be co-transcriptional. Since we did not find any enrichment of genes with a TSS adenosine in these altered gene set, other mechanisms might specify such targets.

Multifunctionality of RNA modification writer enzymes is a topic that is gaining increasing attention. The focus so far has been on the catalytic activity and the impact these modifications have on gene regulation (Roignant and Soller, 2017). Nevertheless, several studies point to non-catalytic roles for some of the major RNA methyltransferases active in human cells. The nuclear m^6A writer METTL3 is shown to have a cytosolic role in enhancing translation by promoting formation of the mRNA closed loop structure (Choe et al., 2018). Another nuclear mammalian m^6A writer METTL16 methylates nuclear U6 snRNA (Pendleton et al., 2018), but METTL16 is also shown to have a cytosolic function (Nance et al., 2020) as a translation factor (Su et al., 2022). While mammalian PCIF1 is known as an m^6Am writer (Akichika et al., 2019), in this study we use the naturally catalytic-dead fly Pcif1 to describe its non-catalytic role in nuclear gene regulation to ensure female fertility and as a modulator of epigenetic processes.

MATERIALS AND METHODS

Generation of *Pcif1* mutant flies

The *Drosophila melanogaster Pcif1* orthologue (*Pcif1*, CG11399) was targeted with two gRNAs designed to delete the putative nuclear localization signal (NLS) at the N-terminus of the protein (Figure 1A). The sgRNAs were identified with the "CRISPR Optimal Target Finder" website (http://tools.flycrispr.molbio.wisc.edu/targetFinder/). They are located within the CDS, at the following positions:

chr3L:20783975 (AGCATCATCACCTGAAGCGG, sgRNA1)

chr3L:20783692 (CCTTGCGTTCCATCAGCCAG, sgRNA2)

The two sgRNAs were cloned separately into the vector pU6-BbsI-chiRNA using oligonucleotides RRoligo1161 + RRoligo1162 for sgRNA1 and RRoligo1163 + RRoligo1164 for sgRNA2 (Table S1). The positive clones were PCR identified and sequenced to confirm the presence of correct sgRNA sequences.

The resulting plasmids were injected into ~200 embryos of a *nos-Cas9* transgenic strain (genotype: *yw;attP40[nos-Cas9]/CyO, Bloomington Drosophila Stock Center #54594*) by BestGene Inc. (Chino Hills, CA, USA).

The adult founder males (G0) were crossed with virgin *w/w; MKRS/TM6* females. Ten G1 males emerging from each G0 cross were then individually crossed with five virgin females *w/w; MKRS/TM6* for three days at 25°C and each male was collected for PCR genotyping using primers CG11399_PCR2_F and CG11399_PCR2_R that flank the targeted region (Table S1). This gave PCR products indicative of alleles with either deletion (~300 bp) or the wild type allele (582 bp). The deleted PCR products were further characterized by Sanger sequencing which leads to the identification of two independent deletions of the targeted sequence (alleles $Pcif1^{F9\Delta NLS}$, $Pcif1^{M6stop}$) (Figure 1B).

In order to remove putative mutations induced by the CRISPR-Cas9 mutagenesis from the genome of *Pcif1* mutants the *Pcif1* alleles were outcrossed using the Dmel\P{GT1}CG11399^{BG02557} ($Pcif1^{BG02557}$, here referred to as the control) fly line. This fly line carries a P element containing a *miniwhite* [mw^+] gene inserted in the first intron of the *Pcif1* gene (Figure 1C) in a w^{1118} context, and showed no phenotypic defects. It was used to backcross the *Pcif1* mutant alleles for five (longevity assay) to ten generations (weight monitoring, fertility analysis and PEV assay). Therefore, except if otherwise indicated (PEV assay with w^{mh4}, see below), the phenotypic analysis of $Pcif1^{F9\Delta NLS}$, $Pcif1^{M6stop}$ alleles was performed in the w^{1118} background of the P{GT1}CG11399^{BG02557} line.

Other fly lines used in this study

Three deficiency fly lines carrying three different chromosomal deletions encompassing *Pcif1* locus on the third chromosome were used in this study to observe the variegation eye phenotype. These lines were provided by the Bloomington Drosophila Stock Center (BDSC):

25721 *w[1118]; Df(3L)BSC563/TM6C, cu[1] Sb[1]* (Df(3L)BSC563);

27369 *w[1118]; Df(3L)BSC797/TM6C, Sb[1] cu[1]* (Df(3L)BSC797);

24956 *w[1118]; Df(3L)BSC452/TM6C, Sb[1] cu[1]* (Df(3L)BSC452).

The w^{m4h} fly line was provided by Dr. C. Grimaud (Montpellier).

Hatching rate analysis

50 adult females were mated with males for 24h at 25°C on media allowing eggs collection (335 mL H_2O, 15 g Agar, 165 mL grape jus and 8 g sucrose). 100 eggs issued from each cross were transferred on fresh media and incubated at 25°C for 24h. The number of unhatched eggs were counted. Three independent replicates were performed for each cross.

Fertility analysis

50 males or virgin females were collected and crossed individually with three wild type w^{1118} flies of the opposite sex at 25°C. After a mating period of 2 days, they were transferred to a new vial for 5 days and discarded. The emerging flies were counted and the statistical analysis was performed using the one-way ANOVA test in GraphPad Prism9.

Longevity analysis

A total of 100 to 150 males of the same age were collected and raised at 25°C per group of 10 flies per vial. Flies were flipped into fresh vials every 2-3 days for the duration of the experiment. The number of dead flies was counted every 2 days. The statistical analysis was performed using the Log-rank (Mantel-Cox) test in GraphPad Prism 9.

Weight analysis

The larval density was here controlled by transferring 50 L1 larvae in new vial. The resulting males and females were weighted by group of ten flies 2-3 days after emergence on a Mettler Toledo Classic precision balance. Ten replicates were performed for each sex and genotype. The statistical analysis was performed using the one-way ANOVA test in GraphPad Prism9.

Immunofluorescence analysis

5-10 ovary pairs were dissected from control, $Pcif1^{F9\Delta NLS}$, $Pcif1^{M6stop}$ or $Pcif1^{F9\Delta NLS}/+$ females in PBS. They were immediately fixed in 4% PFA for 30 min at room temperature (RT). Fixed tissues were then washed (3 x 5 min) with 1xPBS, permealized for 5 min with PBT (1xPBS containing 0.1% Triton X-100) at RT and blocked in blocking buffer (1xPBS, 0.1% Triton X-100, 5% horse serum) for 1 h at RT. They were incubated with rabbit anti-Pcif1 (1:100)(Pandey et al. 2020) and mouse anti-fibrillarin (1:100; AbCam ab4566) antibodies in blocking buffer overnight at 4°C. They were then washed (3 x 15 min) in PBT at RT and incubated for 4 h with anti-rabbit conjugated to Alexa 488 (1:500; Invitrogen, Cat. No. A11029), anti-mouse conjugated to Cyanine 3 (1:500; Jackson ImmunoResearch, Cat. No. 111165144) antibodies and Hoechst (1:1000; ThermoFisher, Cat. No. 33342) in blocking buffer at RT. Ovaries or fat bodies were then washed in PBT (3 x 10 min) at RT and stored in PBS before mounting in Dako (Cat. No. S3022). Images were captured with The Nikon A1RSi confocal microscope (CEA, Grenoble).

Western Blot

For analysing Pcif1 expression, 10 heads from five days old males were dissected for western blot analysis. To check the expression of ATP5A mitochondrial marker, 10 pairs of ovaries from female adult flies were dissected. Samples were lysed in 100 µl of 1x Laemmli buffer (BioRad Cat. No.1610747) containing 1x β-mercaptoethanol (Sigma Cat. No.07917KE). The lysates were boiled at 95°C for 5 min and centrifuged at 12000 rpm at RT for 5 min. 30 µl of each sample were loaded on 7.5% Mini-PROTEAN® TGX (BioRad Cat. No. 456-8025). Gel electrophoresis was performed at 70V for 200 min. After separation, the proteins were transferred on Immobilon-P 0.45 µm nitrocellulose membrane (Merk Millipore Ltd. Cat. No. IPHV00010) overnight at 5V at RT using Trans-Blot SD Semi-Dry Transfer Cell system (Bio-Rad, Cat. No. 1703940). After the transfer, the membrane was washed in Tris-buffered saline (TBS, 20 mM Tris, 150 mM NaCl, pH 7.6) and blocked for 1 h at room temperature with 5% fat free milk in TBS with 0.05% Tween20 (TTBS) (Sigma, Cat. No. P7949). The blocked membranes were incubated with with rabbit anti-Pcif1 (1:1000) (Pandey et al., 2020) for 3 h at room temperature in 5% milk with TTBS, or with anti-mouse ATP5A (1:1000, AbCam ab14748) overnight at 4°C. Then, membranes were washed 3 times for 10 min with TTBS and incubated with HRP-conjugated secondary antibody at either 1:10000 dilution with anti-rabbit IgG HRP-linked or at 1:2000 dilution with anti-mouse IgG HRP-linked for 1 h at RT in 5% milk in TTBS. The membranes were washed 3 times for 10 min with TTBS and incubated with SuperSignal West Femto Maximum Sensitivity Substrate (ThermoFisher, Cat. No. 34095) for 5 min at RT or with Immobilon Forte Western HRP Substrate (Merk Millipore Ltd., Cat. No.WBLUF0500) for 15 min at RT. The signal was detected using Molecular Imager ChemiDoc™ XRS+ (BioRad Cat. No.252559) with ImageLab software (v 5.1).

Transcriptome analyses

Ovaries were dissected from 5-days old homozygous *Pcif1*^{M6stop}, homozygous *Pcif1*$^{F9\Delta NLS}$, and control female flies. Three pairs of ovaries were collected per line for RNA extraction with Trizol reagent. Ribozero-depleted deep sequencing libraries were prepared and sequenced as previously described (Pandey et al., 2020). Specifically, the NEB Next ultra-kit (NEB: E7760L kit) was used to prepare the pair-end libraries which were sequenced in high-mode on Next-seq 500 at GeneCore, EMBL.

Reads were sorted into individual libraries based on the barcodes and mapped to the fly transcriptome (Ensembl release 95) using salmon v1.3.0 (salmon quant with options -l A --validateMappings --gcBias) (Patro et al., 2017). Further analysis was performed using R version 3.6.3 and Bioconductor (Huber et al., 2015). The DESeq function of DESeq2_1.25.10

bioconductor package (Love et al., 2014) was used to obtain log2 fold changes of gene expression between control and mutant samples and the adjusted p-values. Adjusted p-value 0.1 was used as a threshold for statistical significance. In MA plots the genes with significantly different expression between the mutant and control samples were highlighted in red (Figure 2B). The sets of genes upregulated and dowregulated in the mutant ovaries were compared using Venn diagrams (VennDiagram 1.6.20) (Figure 2E). The Volcano plots were plotted using EnhancedVolcano function from EnhancedVolcano 1.3.5 package (https://github.com/kevinblighe/EnhancedVolcano). To see whether whole chromosome expression might be affected, we produced boxplots of log2 fold expression changes for genes lying on individual chromosomes with the y-axes limited to <-1,1> to better see whether the median fold change deviates from zero (Figure 2C). The wilcox.test was used to test the symmetric distribution of log2 fold changes for each chromosome. To see whether the transcript starting nucleotide affects the overall gene expression changes, we filtered only genes which produce transcripts with specific starting nucleotide and divided them into the groups based on that nucleotide. Using stats:::ecdf we then computed empirical cumulative distribution function on log2 fold changes between the mutant and control samples and plotted the distribution (Figure 2H). The group of genes found to be significantly up- or downregulated in the mutant were searched for enriched Gene Ontology terms in the Biological Process ontology using ENRICHR (Chen et al., 2013; Kuleshov et al., 2016) and top ten enriched categories were shown (Figure 2D). The proton transport genes (associated with the term GO: 0015992) were found to be enriched among upregulated genes in the mutant samples. To visualize their expression across all samples, we calculated the z-scores of log2 expression (reads per million) for each of those genes and plotted the boxplots for individual samples (Figure 2F). Similarly, the heatmap of z-scores was plotted using pheatmap 1.0.12 (Figure 2G).

Deep sequencing data generated in this study are deposited with Gene Expression Omnibus (GEO) under accession number GEO: GSE218059.

Electron microscopy imaging of fly ovaries

For transmission electron microscopy (TEM) imaging, dissected ovaries from 5 days old fertilized females were fixed for 30 min at room temperature by addition in the culture medium of a 2x fixative solution (4% paraformaldehyde, 0.4% glutaraldehyde, 0.2 M PHEM). They were then fixed for another 30 min at room temperature in 2% paraformaldehyde, 0.2% glutaraldehyde in 0.1 M PHEM (30 mM PIPES, 12.5 mM HEPES, 5 mM MgCl2, 1 mM EGTA, pH 7). Ovaries were rinsed three times in 0.1 M PHEM and post-fixed in 1% osmium tetroxide (OsO4), 1.5% potassium ferrocyanide in 0.1 M PHEM buffer for 1h at room temperature. After

3 washes in 0.1 M PHEM, they were post-stained using 0.5% uranyl acetate in 30% ethanol for 30 min at room temperature, in the dark. Ovaries were then dehydrated in graded ethanol series (50%, 70%, 90%, 5 min each, then 100% for 5 min, 3 times), and embedded in Epon resin. After polymerisation during 48 h at 65 °C, the blocs were cut on a Leica UC7 ultra-microtome using a Diatome 35° diamond knife. Ultrathin sections (70 nm) were collected on Formvar carbon coated copper grids and post-stained with 2% uranyl acetate and 2% lead citrate. Images were recorded on a Tecnai G2 Spirit BioTwin (FEI) transmission electron microscope (TEM), operating at 120 kV, using an ORIUS SC1000 CCD camera (Gatan).

Larval chromosome staining from dissected salivary glands

Chromosome squashes were performed with the acid-free technique as described in Johansen et al. 2009. *Pcif1*^{BG02557}or *Pcif1*^{M6stop} third-instar larvae were rinsed in PBS and dissected in a formaldehyde dissection buffer (0.15M PIPES, 3mM MgSO3, 1.5mM EGTA, 1.5% NP40, 2% formaldehyde). Then the salivary glands were incubated 5 min in glycerol 50% and squashed on a frosted microscope slides. After squashing, the slides were incubated for 30 min in 0.1% Triton X-100 in PBS (PBST) and blocked for 1h in PBST-BSA 5%. The slides were incubated in a humid chamber with the rabbit anti-Pcif1 (1:100) overnight. The slides were washed 3 times in PBST before being incubated with anti-rabbit conjugated to Alexa 488 in a humid chamber for 2h at RT. Two additional washes in PBS were performed and the samples stained with DAPI. The slides were immediately observed with a microscope confocal LSM800.

PEV observation and pigment quantification

5 virgin w^{m4h}/w^{m4h};+/+ females were crossed with three males of the following genotype: w^{1118};*Pcif1*^{M6stop}, w^{1118};*Pcif1*$^{F9\Delta NLS}$, w^{1118};*Df(3L)BSC797/TM3*, w^{1118};*Df(3L)BSC563/TM6* or w^{1118};*Df(3L)BSC452/TM6.* In the G1 progeny, variegated eyes were checked in transheterozygous females. Pictures of three days old adults were taken using a Leica MZFIII microscope (Leica Microsystem CHS).

For pigmentation quantification, five heads from six days-old females of the genotypes of interest were dissected and snap-frozen in liquid nitrogen. Heads were lysed in 500 ml of AEA lysis buffer (30% ethanol, 0.2% HCl) and homogenized with an electric pestler and the mix was stored at 25°C for 48h in the dark. Each sample was then centrifuged at 10000 rpm for 5 min at room temperature and the supernatant was transferred in a new tube and centrifuged again at 10000 rpm for 5 min at RT. 150 µl of the supernatant were loaded on a 96 well plate and the absorbance of the extracted pigments were measured at 480 nm. Ten replicates were analysed

for each genotype. The statistical analysis was performed using the one-way ANOVA test in GraphPad Prism 9.

ACKNOWLEDGEMENTS

We thank Dr A. Boivin, Dr WP Galej and Dr. A. Verdel for scientific discussions. We thank Dr C. Moriscot for electron microscopy analysis. We thank the iGE3 Genomics Platform at University of Geneva, and the Genomics Core at EMBL Heidelberg. This work was supported by the GRAL LabEX (ANR-10-LABX-49-01) financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003) to MOF and by the "Région Auvergne-Rhône-Alpes" (PAI 2021) to ET. G.F is supported by the IDEX PhD Fellowship from University of Grenoble, and by funding from the Swiss National Science Foundation (ERC Transfer Grant CRETP3_166923) to R.S.P. EB is supported by grants from the Agence Nationale pour la Recherche (ANR-20-CE12-0005 CHApiTRE and ANR-21-CE12-0022 BiopiC), and the French government Emergence-IDEX-ISITE initiative 16-IDEX-0001 (CAP20-25). This work used the EM facilities of the Grenoble Instruct-ERIC center (ISBG ; UAR 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-0005-02) and LabEx GRAL. The IBS Electron Microscope facility is supported by the Auvergne Rhône-Alpes Region, the Fonds Feder, the Fondation pour la Recherche Médicale and GIS-IBiSA. Stocks obtained from the Bloomington Drosophila Stock Center (NIH P40OD018537) were used in this study.

AUTHOR CONTRIBUTIONS

G.F performed most of the experiments for this study; G.F. and E.T. generated fly mutants and performed fly crosses; E.D and R.R.P prepared sequencing libraries; D.H. performed computational analyses; N.G. and E.B conducted immunostaining on polytene chromosomes; E.T. M-O.F. and R.S.P. designed and supervised the study. Manuscript preparation and writing were performed by G.F., E.T., R.S.P. and M-O.F. with input from everyone.

REFERENCES

Akichika, S., Hirano, S., Shichino, Y., Suzuki, T., Nishimasu, H., Ishitani, R., Sugita, A., Hirose, Y., Iwasaki, S., Nureki, O*., et al.* (2019). Cap-specific terminal N (6)-methylation of RNA by an RNA polymerase II-associated methyltransferase. Science *363,*(6423):eaav0080.

Baker, K. E., Parker, R. (2004). Nonsense-mediated mRNA decay: terminating erroneous gene expression. Curr Opin Cell Biol *16*,293-9.

Belanger, F., Stepinski, J., Darzynkiewicz, E., and Pelletier, J. (2010). Characterization of hMTr1, a human Cap1 2'-O-ribose methyltransferase. J Biol Chem *285*, 33037-33044.

Boulias, K., Toczydlowska-Socha, D., Hawley, B.R., Liberman, N., Takashima, K., Zaccara, S., Guez, T., Vasseur, J.J., Debart, F., Aravind, L*., et al.* (2019). Identification of the m(6)Am Methyltransferase PCIF1 Reveals the Location and Functions of m(6)Am in the Transcriptome. Mol Cell *75*, 631-643 e638.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Molecular cell *36*, 541-546.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Choe, J., Lin, S., Zhang, W., Liu, Q., Wang, L., Ramirez-Moya, J., Du, P., Kim, W., Tang, S., Sliz, P*., et al.* (2018). mRNA circularization by METTL3-eIF3h enhances translation and promotes oncogenesis. Nature *561*, 556-560.

Fan, H., Sakuraba, K., Komuro, A., Kato, S., Harada, F., and Hirose, Y. (2003). PCIF1, a novel human WW domain-containing protein, interacts with the phosphorylated RNA polymerase II. Biochem Biophys Res Commun *301*, 378-385.

Galloway, A., and Cowling, V.H. (2019). mRNA cap regulation in mammalian cell function and fate. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms *1862*, 270-279.

Ghosh, A., Shuman, S., and Lima, C.D. (2011). Structural insights to how mammalian capping enzyme reads the CTD code. Molecular cell *43*, 299-310.

Haline-Vaz, T., Silva, T.C., and Zanchin, N.I. (2008). The human interferon-regulated ISG95 protein interacts with RNA polymerase II and shows methyltransferase activity. Biochem Biophys Res Commun *372*, 719-724.

Hampsey, M., and Reinberg, D. (2003). Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. Cell *113*, 429-432.

Haussmann, I.U., Wu, Y., Nallasivan, M.P., Archer, N., Bodi, Z., Hebenstreit, D., Waddell, S., Fray, R., and Soller, M. (2022). CMTr cap-adjacent 2’-O-ribose mRNA methyltransferases are required for reward learning and mRNA localization to synapses. Nature Communications *13*, 1-13.

Hill, J.H., Chen, Z., Xu, H. (2014). Selective propagation of functional mitochondrial DNA during oogenesis restricts the transmission of a deleterious mitochondrial variant. Nature Genetics *46*,389–392.

Hirose, Y., Iwamoto, Y., Sakuraba, K., Yunokuchi, I., Harada, F., and Ohkuma, Y. (2008). Human phosphorylated CTD-interacting protein, PCIF1, negatively modulates gene expression by RNA polymerase II. Biochem Biophys Res Commun *369*, 449-455.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T.*, et al.* (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods *12*, 115-121.

Inesta-Vaquera, F., and Cowling, V.H. (2017). Regulation and function of CMTR1-dependent mRNA cap methylation. Wiley Interdiscip Rev RNA *8,* (6):e1450.

Johansen, K. M., Cai, W., Deng, H., Bao, X., Zhang, W., Girton, J., Johansen J. (2009). Polytene chromosome squash methods for studying transcription and epigenetic chromatin modification in Drosophila using antibodies. Methods *483*,87-97.

Kachaev, Z.M., Lebedeva, L.A., Shaposhnikov, A.V., Moresco, J.J., Yates III, J.R., Schedl, P., and Shidlovskii, Y.V. (2019). Paip2 cooperates with Cbp80 at an active promoter and participates in RNA Polymerase II phosphorylation in Drosophila. FEBS letters *593*, 1102-1112.

Keith, J.M., Ensinger, M.J., and Moss, B. (1978). HeLa cell RNA (2'-O-methyladenosine-N6-)-methyltransferase specific for the capped 5'-end of messenger RNA. J Biol Chem *253*, 5033-5039.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A.*, et al.* (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res *44*, W90-97.

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. (2002). Functional interaction of yeast pre-mRNA 3′ end processing factors with RNA polymerase II. Molecular cell *9*, 1101-1111.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. Nature *430,* 223-226.

Moyer, S.A., and Banerjee, A.K. (1976). In vivo methylation of vesicular stomatitis virus and its host-cell messenger RNA species. Virology *70*, 339-351.

Nance, D.J., Satterwhite, E.R., Bhaskar, B., Misra, S., Carraway, K.R., and Mansfield, K.D. (2020). Characterization of METTL16 as a cytoplasmic RNA binding protein. PLoS One *15*, e0227647.

Pandey, R.R., Delfino, E., Homolka, D., Roithova, A., Chen, K.-M., Li, L., Franco, G., Vågbø, C.B., Taillebourg, E., and Fauvarque, M.-O. (2020). The Mammalian Cap-Specific m6Am RNA Methyltransferase PCIF1 Regulates Transcript Levels in Mouse Tissues. Cell reports *32*, 108038.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods *14*, 417-419.

Pendleton, K.E., Park, S.-K., Hunter, O.V., Bresson, S.M., and Conrad, N.K. (2018). Balance between MAT2A intron detention and splicing is determined cotranscriptionally. RNA *24*, 778-786.

Reuter, G., Wolff, I. (1981). Isolation of dominant suppressor mutations for position-effect variegation in Drosophila melanogaster. Mol Gen Genet *182*, 516-9.

Roignant, J.Y., and Soller, M. (2017). m(6)A in mRNA: An Ancient Mechanism for Fine-Tuning Gene Expression. Trends Genet *33*, 380-390.

Sendinc, E., Valle-Garcia, D., Dhall, A., Chen, H., Henriques, T., Navarrete-Perea, J., Sheng, W., Gygi, S.P., Adelman, K., and Shi, Y. (2019). PCIF1 Catalyzes m6Am mRNA Methylation to Regulate Gene Expression. Mol Cell *75*, 620-630 e629.

Shatkin, A.J. (1976). Capping of eucaryotic mRNAs. Cell *9*, 645-653.

Srivastava, R., Kumar, D., Waskar, M.N., Sharma, M., Katoch, V.M., and Srivastava, B.S. (2006). Identification of a repetitive sequence belonging to a PPE gene of Mycobacterium tuberculosis and its use in diagnosis of tuberculosis. J Med Microbiol *55*, 1071-1077.

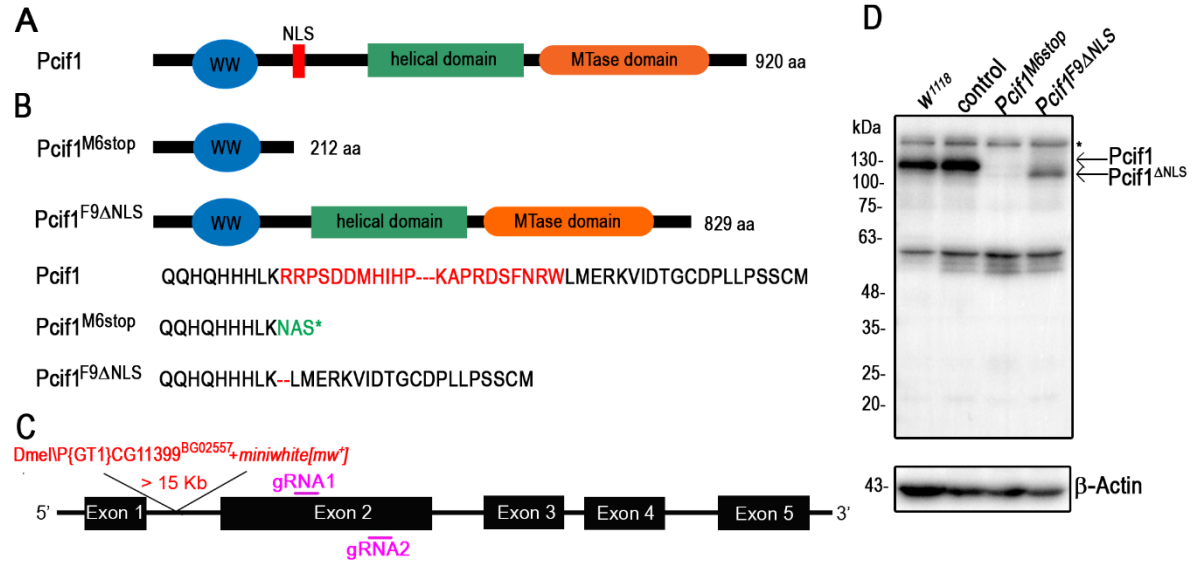
Su, R., Dong, L., Li, Y., Gao, M., He, P.C., Liu, W., Wei, J., Zhao, Z., Gao, L., and Han, L. (2022). METTL16 exerts an m6A-independent function to facilitate translation and tumorigenesis. Nature cell biology, 1-12.

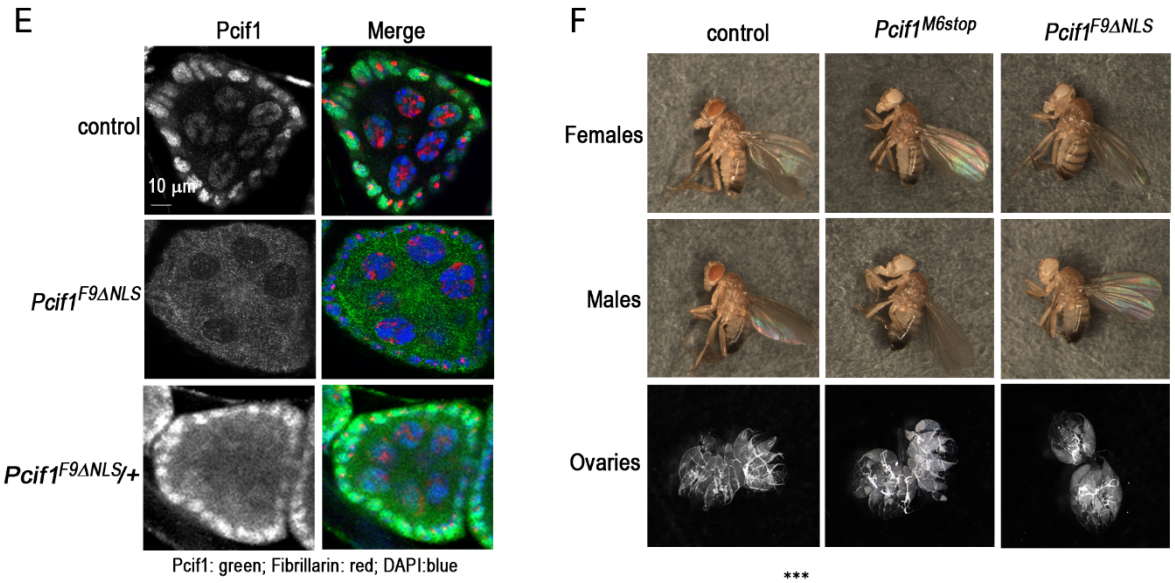
Wang, J., Lawry, S.T., Cohen, A.L., Jia, S. (2014) Chromosome boundary elements and regulation of heterochromatin spreading. Cell Mol Life Sci *71*, 4841-52.

Wang, Z.H., Liu, Y., Chaitankar, V., Pirooznia, M., Xu, H. (2019) Electron transport chain biogenesis activated by a JNK-insulin-Myc relay primes mitochondrial inheritance in Drosophila. Elife *15*;8:e49309.

Wei, C., Gershowitz, A., and Moss, B. (1975). N6, O2'-dimethyladenosine a novel methylated ribonucleoside next to the 5' terminal of animal cell and virus mRNAs. Nature *257*, 251-253.

FIGURES


A
Pcif1
NLS
WW
helical domain
MTase domain
920 aa
B
Pcif1M6stop
WW
212 aa
Pcif1F9ΔNLS
WW
helical domain
MTase domain
829 aa
Pcif1 QQHQHHHLKRRPSDDMHIHP---KAPRDSFNRWLMERKVIDTGCDPLLPSSCM
Pcif1M6stop QQHQHHHLKNAS*
Pcif1F9ΔNLS QQHQHHHLK--LMERKVIDTGCDPLLPSSCM
C
DmelP{GT1}CG11399BG02557+miniwhite[mw+]
> 15 Kb
gRNA1
5'
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
3'
gRNA2
D
w1118
control
Pcif1M6stop
Pcif1F9ΔNLS
kDa
130-
100-
75-
63-
48-
35-
25-
20-
Pcif1
Pcif1ΔNLS
43-
β-Actin


E
Pcif1
Merge
control
10 µm
Pcif1F9ΔNLS
Pcif1F9ΔNLS/+
Pcif1: green; Fibrillarin: red; DAPI:blue
F
control
Pcif1M6stop
Pcif1F9ΔNLS
Females
Males
Ovaries

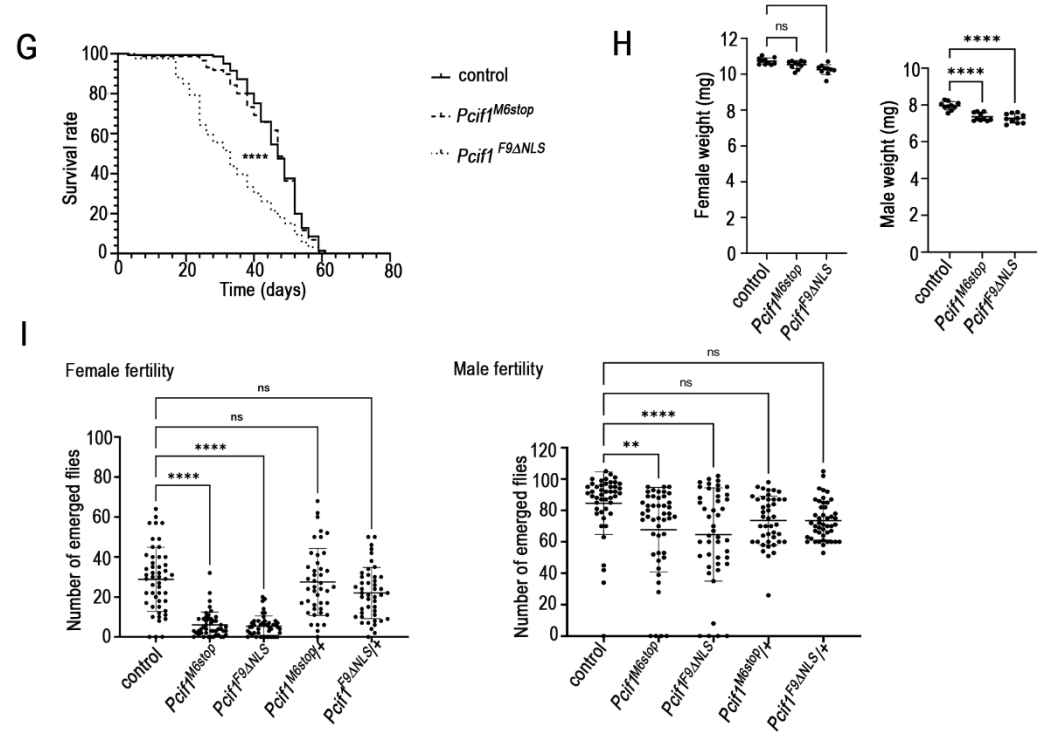

G
Survival rate
Time (days)
control
Pcif1M6stop
Pcif1 F9ΔNLS
****
H
Female weight (mg)
Male weight (mg)
***
ns
****
****
I
Female fertility
Male fertility
Number of emerged flies
ns
****
**
control
Pcif1M6stop
Pcif1F9ΔNLS
Pcif1M6stop/+
Pcif1F9ΔNLS/+

FIGURE 1. Molecular characterization of Pcif1 in *Drosophila melanogaster.* **A.** Structural organization of the wild type Pcif1 protein. The N-terminal WW domain, the Nuclear Localization Signal (NLS), the helical domain and the C-terminal Methyl-Transferase (MTase) domain are shown. **B.** The CRISPR-Cas9 mutagenesis generated the $Pcif1^{M6stop}$ allele which causes a frameshift ending in a premature stop codon generating a truncated protein and the $Pcif1^{F9\Delta NLS}$ allele which has an in-frame deletion of the NLS. **C.** Illustration of the Pcif1 locus with the indication of the *P* element insertion site in the $Pcif1^{BG02557}$ line (Dmel\PCG11399^{BG02557}), here used as control, and of the position of the sgRNAs, gRNA1 and gRNA2 designed for the CRISPR-Cas9 mutagenesis (purple lines). **D.** Western blot analysis of head protein extracts from $Pcif1^{BG02557}$ control, or $Pcif1^{M6stop}$ and $Pcif1^{F9\Delta NLS}$ mutants as indicated. Wild type Pcif1 and Pcif1$^{\Delta NLS}$ protein size are indicated by the black arrows. β-Actin was used as loading control. The asterisks indicate unspecific bands. **E.** Immunostaining of ovaries dissected from $Pcif1^{BG02557}$ controls, or $Pcif1^{F9\Delta NLS}$ homozygous or $Pcif1^{F9\Delta NLS}$/+ heterozygous females. The egg chamber, formed by an external layer of follicle cells (somatic cells) which surround the nurse cells (germ cells), is shown. Pcif1: green, DNA: blue, Fibrillarin (nucleolar marker): red **F.** Pictures of 5-days old females, males or dissected ovaries of either $Pcif1^{BG02557}$ (control) or, $Pcif1^{M6stop}$ and $Pcif1^{F9\Delta NLS}$ mutants. G**.** Longevity analysis of $Pcif1^{BG02557}$ control (n=150) or $Pcif1^{M6stop}$ (n=150) and $Pcif1^{F9\Delta NLS}$ (n=130) mutants. H. Weight analysis of $Pcif1^{BG02557}$ control or $Pcif1^{M6stop}$ and $Pcif1^{F9\Delta NLS}$ mutants: each dot represents the weight of a group of ten adult flies per genotype for either males or females as indicated. Ten replicates were performed. I**.** Fertility analysis of females or males expressed as the number of emerging flies per cross of one individual of the indicated genotype with three wild type flies of the opposite sex. Log-rank (longevity) and one-way ANOVA (weight and fertility) tests were performed using GraphPad Prism9. p value : ***P< 0.001, **** P<0.0001. ns: non significant.

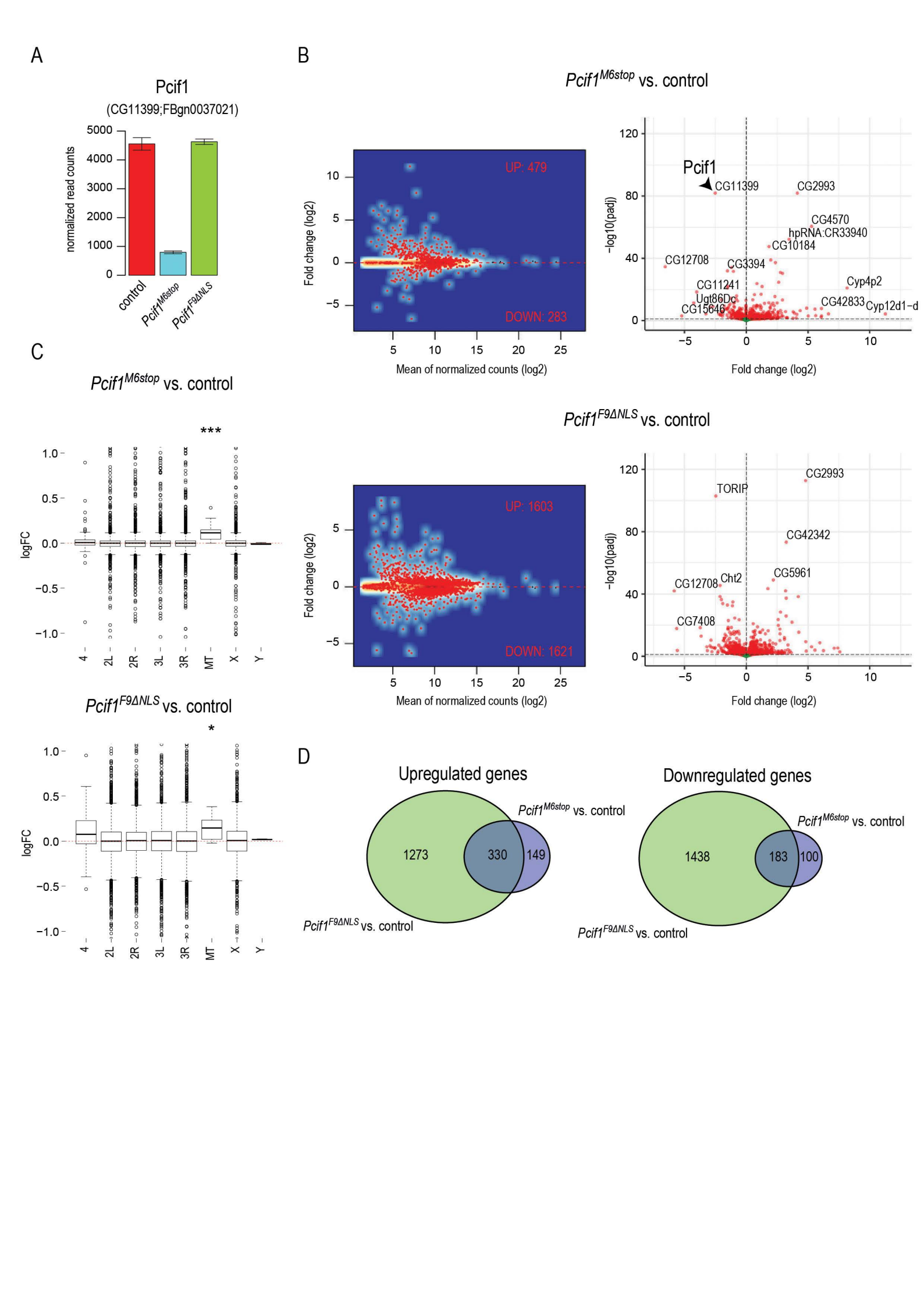

A
Pcif1
(CG11399;FBgn0037021)
normalized read counts
5000
4000
3000
2000
1000
0
control
Pcif1M6stop
Pcif1F9ΔNLS
B
Pcif1M6stop vs. control
UP: 479
DOWN: 283
Fold change (log2)
Mean of normalized counts (log2)
Pcif1
CG11399
CG2993
CG4570
hpRNA:CR33940
CG10184
CG12708
CG3394
CG11241
Ugt86Dc
CG15646
Cyp4p2
CG42833
Cyp12d1-d
−log10(padj)
Fold change (log2)
C
Pcif1M6stop vs. control
***
logFC
4
2L
2R
3L
3R
MT
X
Y
Pcif1F9ΔNLS vs. control
*
Pcif1F9ΔNLS vs. control
UP: 1603
DOWN: 1621
Fold change (log2)
Mean of normalized counts (log2)
TORIP
CG2993
CG42342
CG5961
CG12708
Cht2
CG7408
−log10(padj)
Fold change (log2)
D
Upregulated genes
Pcif1M6stop vs. control
1273
330
149
Pcif1F9ΔNLS vs. control
Downregulated genes
Pcif1M6stop vs. control
1438
183
100
Pcif1F9ΔNLS vs. control

FIGURE 2. Transcriptomic analyses in *Pcif1* mutants **A.** Barplot showing reduced **t**ranscript levels of *Pcif1* in $Pcif1^{M6stop}$ while the transcript levels in $Pcif1^{F9\Delta NLS}$ are not affected. Data shown as mean ± standard deviation from three biological replicas. **B.** MA plots and volcano plots showing gene expression changes in $Pcif1^{M6stop}$ and in $Pcif1^{F9\Delta NLS}$. Significantly dysregulated genes (padj <= 0.1) are shown in red. **C.** Boxplots showing gene expression changes from individual chromosomes. The view is limited to only changes below 2-fold to better see the median differences (p value : * P< 0.05 , *** P< 0.001; wilcox.text of symmetric distribution). **D.** Venn diagrams showing overlap in sets of upregulated and downregulated genes in $Pcif1^{M6stop}$ and $Pcif1^{F9\Delta NLS}$.

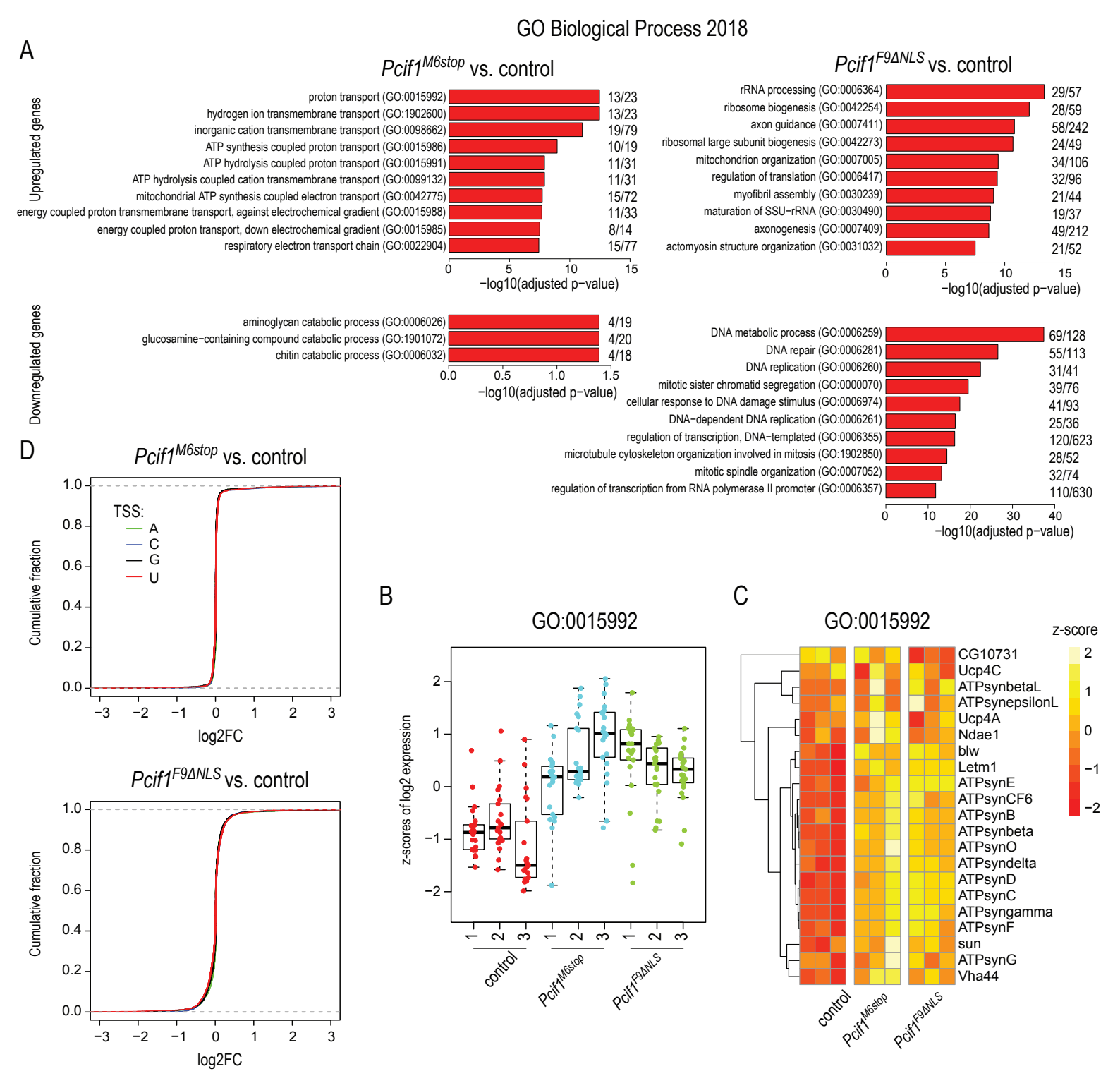

GO Biological Process 2018
A
Pcif1M6stop vs. control
Pcif1F9ΔNLS vs. control
Upregulated genes
proton transport (GO:0015992) 13/23
hydrogen ion transmembrane transport (GO:1902600) 13/23
inorganic cation transmembrane transport (GO:0098662) 19/79
ATP synthesis coupled proton transport (GO:0015986) 10/19
ATP hydrolysis coupled proton transport (GO:0015991) 11/31
ATP hydrolysis coupled cation transmembrane transport (GO:0099132) 11/31
mitochondrial ATP synthesis coupled electron transport (GO:0042775) 15/72
energy coupled proton transmembrane transport, against electrochemical gradient (GO:0015988) 11/33
energy coupled proton transport, down electrochemical gradient (GO:0015985) 8/14
respiratory electron transport chain (GO:0022904) 15/77
0 5 10 15
−log10(adjusted p−value)
rRNA processing (GO:0006364) 29/57
ribosome biogenesis (GO:0042254) 28/59
axon guidance (GO:0007411) 58/242
ribosomal large subunit biogenesis (GO:0042273) 24/49
mitochondrion organization (GO:0007005) 34/106
regulation of translation (GO:0006417) 32/96
myofibril assembly (GO:0030239) 21/44
maturation of SSU−rRNA (GO:0030490) 19/37
axonogenesis (GO:0007409) 49/212
actomyosin structure organization (GO:0031032) 21/52
0 5 10 15
−log10(adjusted p−value)
Downregulated genes
aminoglycan catabolic process (GO:0006026) 4/19
glucosamine−containing compound catabolic process (GO:1901072) 4/20
chitin catabolic process (GO:0006032) 4/18
0.0 0.5 1.0 1.5
−log10(adjusted p−value)
DNA metabolic process (GO:0006259) 69/128
DNA repair (GO:0006281) 55/113
DNA replication (GO:0006260) 31/41
mitotic sister chromatid segregation (GO:0000070) 39/76
cellular response to DNA damage stimulus (GO:0006974) 41/93
DNA−dependent DNA replication (GO:0006261) 25/36
regulation of transcription, DNA−templated (GO:0006355) 120/623
microtubule cytoskeleton organization involved in mitosis (GO:1902850) 28/52
mitotic spindle organization (GO:0007052) 32/74
regulation of transcription from RNA polymerase II promoter (GO:0006357) 110/630
0 10 20 30 40
−log10(adjusted p−value)
D
Pcif1M6stop vs. control
Cumulative fraction
TSS:
A
C
G
U
log2FC
Pcif1F9ΔNLS vs. control
Cumulative fraction
log2FC
B
GO:0015992
z-scores of log2 expression
control
PcifM6stop
PcifF9ΔNLS
C
GO:0015992
z-score
CG10731
Ucp4C
ATPsynbetaL
ATPsynepsilonL
Ucp4A
Ndae1
blw
Letm1
ATPsynE
ATPsynCF6
ATPsynB
ATPsynbeta
ATPsynO
ATPsyndelta
ATPsynD
ATPsynC
ATPsyngamma
ATPsynF
sun
ATPsynG
Vha44
control
Pcif1M6stop
Pcif1F9ΔNLS

FIGURE 3. Pcif1 regulates gene expression of proton transport genes **A.** Gene ontology analysis for differentially expressed genes. Only top 10 significantly enriched terms are shown. **B.** Boxplot showing the z-scores of expression across individual samples for proton transport genes (GO:0015992). **C.** Heatmap showing the z-scores of expression for proton transport genes (GO:0015992). **D.** Plots showing the cumulative distributions of gene expression changes in the $Pcif1^{M6stop}$ and in $Pcif1^{F9\Delta NLS}$ when compared to the control. Separate curves are plotted based on the transcription start site nucleotide (TSS).

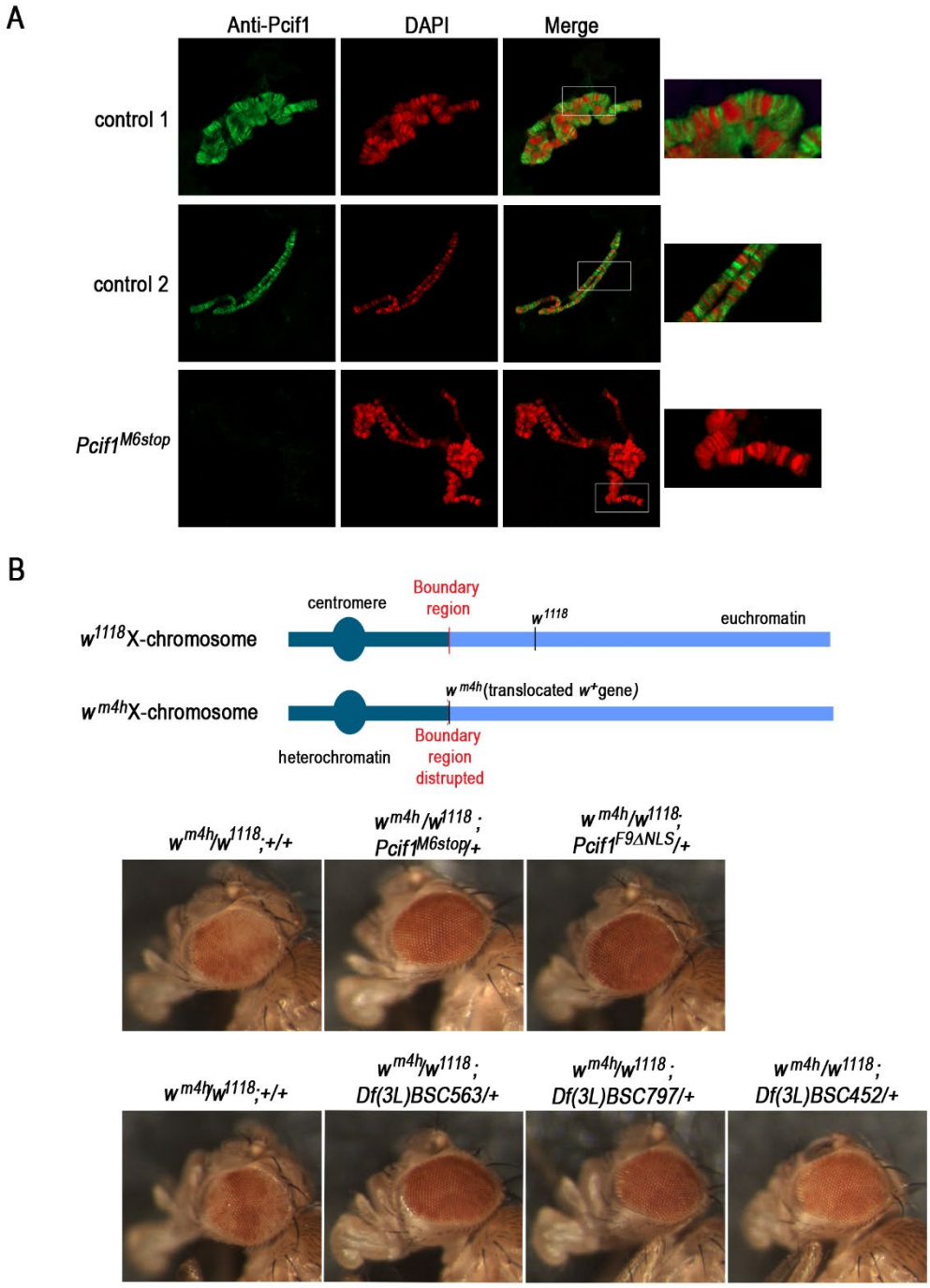
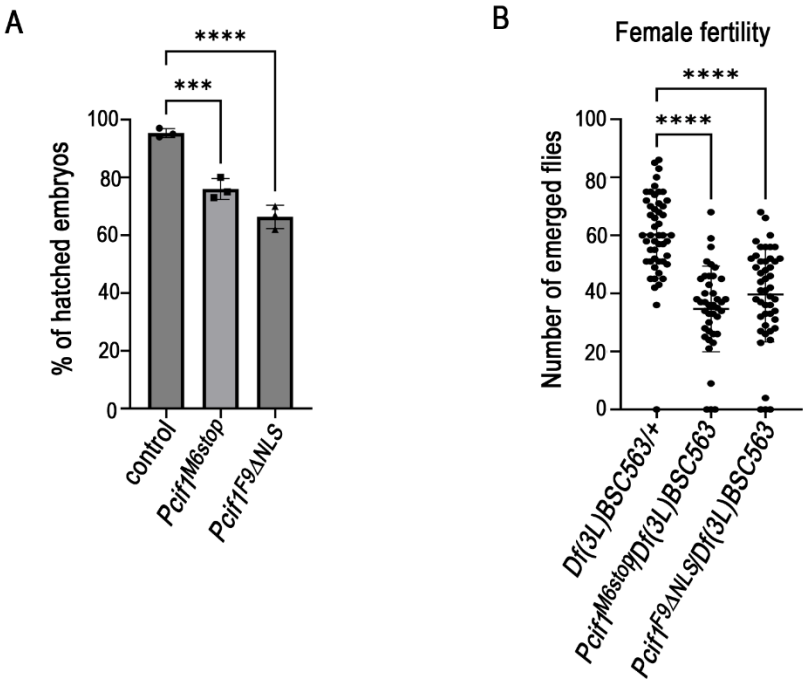
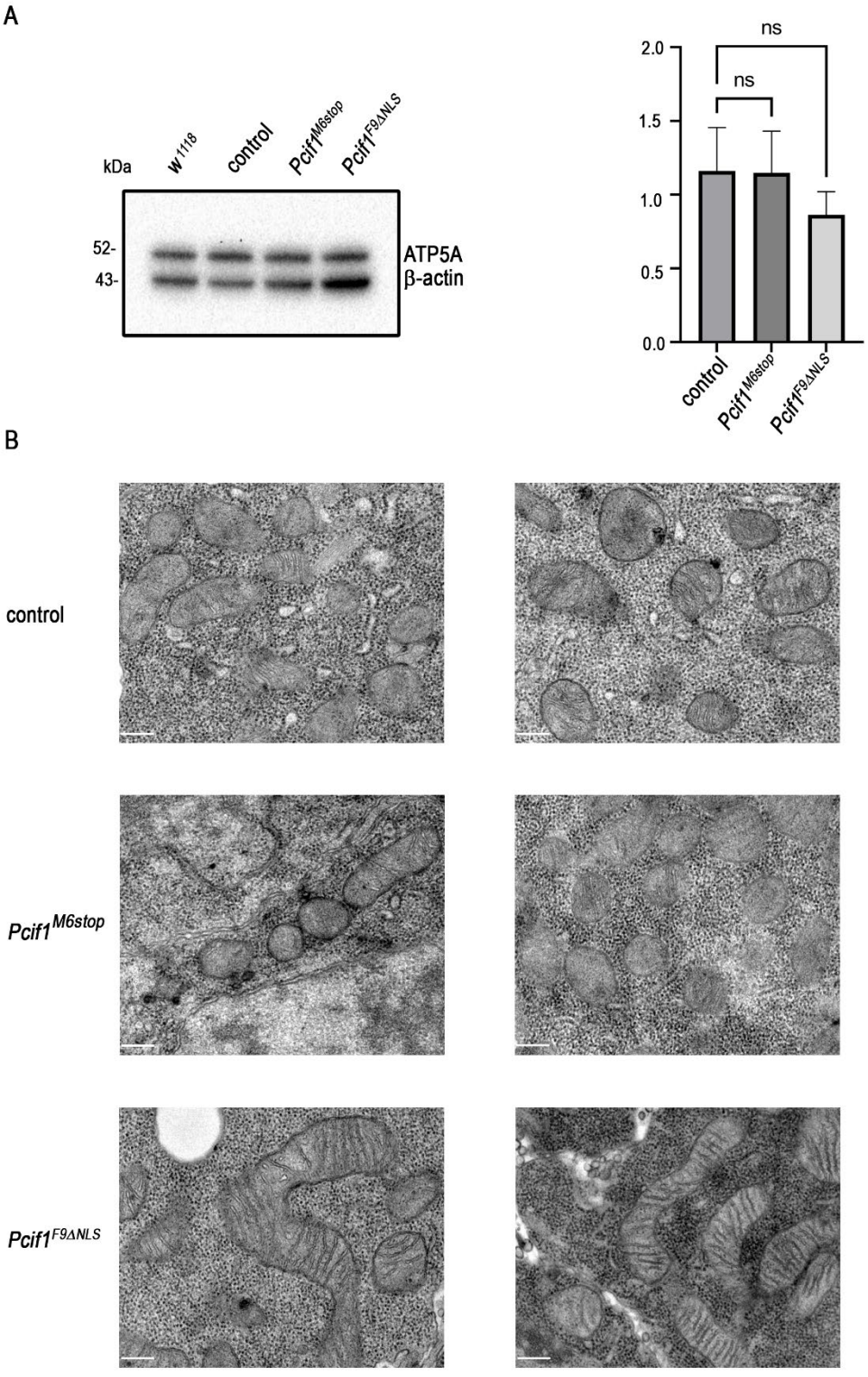

A
Anti-Pcif1
DAPI
Merge
control 1
control 2
Pcif1M6stop
B
centromere
Boundary region
w1118
euchromatin
w1118X-chromosome
w m4h(translocated w+gene)
w m4hX-chromosome
heterochromatin
Boundary region distrupted
w m4h/w1118;+/+
w m4h/w1118; Pcif1M6stop/+
w m4h/w1118; Pcif1F9ΔNLS/+
w m4h/w1118;+/+
w m4h/w1118; Df(3L)BSC563/+
w m4h/w1118; Df(3L)BSC797/+
w m4h/w1118; Df(3L)BSC452/+
C
****
****
****
Pigment absorbance
0.25
0.20
0.15
0.10
0.05
0.00
+/+
Pcif1M6stop/+
Pcif1F9ΔNLS/+
Df(3L)BSC563/+
w m4h/w1118

FIGURE 4. Pcif1 binds to euchromatic regions of polytene chromosomes and behaves as a suppressor of position-effect variegation. **A.** Larval polytene chromosome immunostaining was performed for *Pcif1*^{BG02557} controls (2 samples) and *Pcif1*^{M6stop} mutants as indicated. **B.** PEV assay. Schematic illustration of the position-effect variegation affecting the *white* gene in the w^{m4h} allele. Representative eye phenotype out of 20 to 50 females of the following genotypes: w^{m4h}/w^{1118}; *Pcif1*^{M6stop}/+ or w^{m4h}/w^{1118}; *Pcif1*$^{F9\Delta NLS}$/+ compared to control w^{m4h}/w^{1118}; +/+ flies (first row) and w^{m4h}/w^{1118}; *D(3L)*/+ females flies carrying a chromosomal deletion encompassing the *Pcif1* gene (*Df(3L)BSC563*, *Df(3L)BSC797* or *Df(3L)BSC452* as indicated on figure) compared to w^{m4h}/w^{1118}; +/+ flies (second row). **C.** Quantification of pigment absorbance was performed for five heads per genotype at a wavelength of 480 nm. Ten replicates were performed. One representative experiment out of two is shown. One-way ANOVA test was performed using GraphPad Prism9. P value: **** P< 0.0001.

Supplementary Figures



Supplemental Fig S1. A. Histograms shows the hatching rate analysis of control, *Pcif1*$^{F9\Delta NLS}$ and *Pcif1*^{M6stop} mutants expressed as percentage of hatched embryos. The mean of three independent experiments are shown. A total of 100 embryos were tested. **B.** Fertility analysis expressed as the number of emerging flies per cross of one mutant female (genotypes indicated on figure) with three wild type males. Statistical analysis was performed using GraphPad Prism9, p value: ****P<0.0001.


A
kDa
w1118
control
Pcif1M6stop
Pcif1F9ΔNLS
52-
43-
ATP5A
β-actin
ns
ns
2.0
1.5
1.0
0.5
0.0
control
Pcif1M6stop
Pcif1F9ΔNLS
B
control
Pcif1M6stop
Pcif1F9ΔNLS

Supplemental Fig S2. A. Western blot analysis of protein extracts from 10 pairs of ovaries from w^{1118}, w^{1118};$Pcif1^{BG02557}$, w^{1118};$Pcif1^{M6stop}$, w^{1118};$Pcif1^{F9\Delta NLS}$ flies. β-Actin was used as loading control. The histogram indicates the quantification of ATP5A expression over four independent replicates normalized to protein loading control. Statistical analysis was performed using GraphPad Prism9. ns: non significant. **B.** Electron microscopy pictures of ovaries of control, $Pcif1^{M6stop}$ and $Pcif1^{F9\Delta NLS}$ flies (two pictures per genotype). Scale bar correspond to 0.2 µm.

proton transport (GO:0015992) genes:

															logFC (vs. control)		padj (vs. control)	
gene	gene_biotype	chr	external_gene_name	description	RR1044	RR1045	RR1046	RR1047	RR1048	RR1049	RR1050	RR1051	RR1052	baseMean	$Pcif1^{M6stop}$	$Pcif1^{F9\Delta NLS}$	$Pcif1^{M6stop}$	$Pcif1^{F9\Delta NLS}$
FBgn0010217	protein_coding	4	ATPsynbeta	ATP synthase, beta subunit	28843,4	28424,8	25709,5	34264,7	33979,5	38330,4	34238,7	33940,4	33496,3	32358,63	0,34	0,28	0,00	0,00
FBgn0010612	protein_coding	2L	ATPsynG	ATP synthase	2469,9	2540,3	2257,3	2444,4	2603,1	2921,0	2669,0	2421,5	2579,0	2545,05	0,09	0,07	0,50	0,58
FBgn0011211	protein_coding	2R	blw	bellwether	24277,9	23992,1	20817,4	29947,7	26832,3	28081,5	30412,1	28490,3	28122,4	26774,86	0,27	0,31	0,00	0,00
FBgn0014391	protein_coding	X	sun	stunted	2259,6	2213,5	2444,1	2515,5	2442,4	2946,1	2564,3	2589,0	2459,4	2492,65	0,14	0,12	0,24	0,26
FBgn0016119	protein_coding	3R	ATPsynCF6	ATP synthase, coupling factor 6	2532,9	2545,8	2395,1	2951,5	2890,4	3355,8	3306,6	2868,5	2979,6	2869,60	0,26	0,27	0,01	0,00
FBgn0016120	protein_coding	3R	ATPsynD	ATP synthase, subunit D	3126,4	3281,3	3097,1	3794,0	3894,7	4387,2	4420,6	4195,9	4116,4	3812,61	0,32	0,40	0,00	0,00
FBgn0016691	protein_coding	3R	ATPsynO	ATP synthase, oligomycin sensitivity conferri	3673,6	3635,8	3415,9	3950,0	3953,3	4403,5	4028,1	3977,7	3893,0	3881,21	0,18	0,14	0,03	0,06
FBgn0019644	protein_coding	3L	ATPsynB	ATP synthase, subunit B	3755,2	4137,4	3514,5	4397,8	4460,3	5069,0	4627,5	4447,2	4396,3	4311,69	0,25	0,22	0,02	0,02
FBgn0020235	protein_coding	3R	ATPsyngamma	ATP synthase, gamma subunit	5172,6	5073,8	4697,2	5883,3	5749,2	6283,7	6270,5	6244,0	5716,8	5676,80	0,24	0,27	0,00	0,00
FBgn0028342	protein_coding	X	ATPsyndelta	ATP synthase, delta subunit	3156,4	2836,0	2871,7	3434,1	3338,0	3781,9	3502,2	3518,1	3342,9	3309,02	0,22	0,21	0,01	0,01
FBgn0030872	protein_coding	X	Ucp4A		664,6	727,8	729,4	752,1	892,4	791,2	645,7	748,5	785,4	748,58	0,11	0,03	0,42	0,86
FBgn0031757	protein_coding	2L	Ucp4C		3,3	3,2	5,2	0,9	6,4	3,4	5,0	3,1	1,7	3,58	0,00	-0,01	0,98	0,90
FBgn0031758	protein_coding	2L	Ucp4B	Uncoupling protein 4B	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,00	NA	NA	NA	NA
FBgn0034081	protein_coding	2R	CG10731		200,3	212,1	179,4	215,4	181,5	196,3	135,8	158,6	158,4	181,99	0,00	-0,30	0,99	0,07
FBgn0035032	protein_coding	2R	ATPsynF	ATP synthase, subunit F	1845,1	2008,2	1773,2	2378,9	2290,6	2610,3	2575,9	2498,3	2224,6	2245,02	0,33	0,35	0,00	0,00
FBgn0036568	protein_coding	3L	ATPsynbetaL	ATP synthase, beta subunit-like	0,0	0,0	0,0	0,0	1,3	0,0	0,8	0,0	0,8	0,33	0,00	0,01	NA	NA
FBgn0038224	protein_coding	3R	ATPsynE	ATP synthase, subunit E	2441,0	2527,3	2295,5	2550,9	2800,5	3014,5	3068,1	3009,8	3021,3	2747,67	0,16	0,32	0,10	0,00
FBgn0039830	protein_coding	3R	ATPsynC	ATP synthase, subunit C	26798,2	26891,8	26291,5	31745,0	31516,4	34832,4	35786,6	34329,6	32940,0	31236,83	0,28	0,36	0,00	0,00
FBgn0051477	protein_coding	3R	ATPsynepsilonL	ATP synthase, epsilon subunit-like	0,0	0,0	1,2	0,0	2,3	0,0	4,2	0,0	1,5	1,01	0,00	0,01	NA	NA
FBgn0259111	protein_coding	2L	Ndae1	Na[+]-driven anion exchanger 1	193,6	533,8	244,7	311,8	911,7	684,7	324,5	454,7	531,2	465,64	0,06	0,10	0,24	0,54
FBgn0262511	protein_coding	2R	Vha44	Vacuolar H[+] ATPase 44kD subunit	5115,8	5355,5	4743,5	5447,7	6755,8	6699,9	5726,1	6322,6	5568,5	5748,37	0,25	0,19	0,04	0,11
FBgn0284252	protein_coding	2R	Letm1	Leucine zipper and EF-hand containing transr	2948,4	2882,8	2669,3	3323,7	3622,7	3307,0	3513,4	3411,6	3484,8	3240,43	0,25	0,28	0,00	0,00

Gene ontology analysis of GO:0015992 genes :

	Overlap*	Adj.P.value	Significantly upregulated genes
upregulated in $Pcif1^{M6stop}$	13/23	3,45E-13	ATPsynbeta;Letm1;ATPsyndelta;Vha44;ATPsynCF6;ATPsyngamma;blw;ATPsynO;ATPsynD;ATPsynC;ATPsynF;ATPsynE;ATPsynB
upregulated $Pcif1^{F9\Delta NLS}$	12/23	2,88E-06	ATPsynD;ATPsynbeta;Letm1;ATPsyndelta;ATPsynC;ATPsynF;ATPsynE;ATPsynCF6;ATPsyngamma;blw;ATPsynO;ATPsynB

* upregulated genes/all genes

Supplementary table S1.

List of genes dysregulated in *Pcif1*$^{F9\Delta NLS}$ and *Pcif1*^{M6stop} mutants together with expression of proton transport genes (GO:0015992).

Chapter 2: Nudt19B is required for fly fertility

2. A genetic screen for the identification of new decapping enzymes and alternative cap structures in *Drosophila melanogaster*

The second part of my PhD project consisted in the realization of a genetic screen of the *Nudix* gene family in *Drosophila* for the discovery and study of putative new decapping enzymes that may remove alternative cap structures. The interest for this topic arose from recent findings that demonstrated that some enzymes of the Nudix family, other than the canonical DCP2, are able to hydrolyze alternative cap structures attached at the 5′ end of RNAs in prokaryotes and eukaryotes[282–284]. One of the main limitation for the observation of alternative cap structures and determination of their function in eukaryotes is their scarcity due to the low cellular concentration of non-canonical metabolites putatively involved in the formation of these alternative caps[160]. In this context, we hypothesized that silencing each of member of the Nudix family in *Drosophila* may both reveal their potential functions and provoke the accumulation of alternative cap structures that would become detectable by Mass Spectrometry. By performing BLAST searches, I aligned each of the 22 human Nudix proteins to the *Drosophila melanogaster* database and I found 12 orthologues (Table 1). Interestingly, three NUDT19 orthologues are present in the fly genome which are the result of a recent triplication common to the *Diptera* order.

In order to bring light to Nudix proteins possessing a physiological function in *Drosophila*, I induced the silencing of each of the 12 *Nudix* genes in a tissue-specific manner using the UAS-Gal4 system. To do so, I used three different existing driver lines expressing the yeast *Gal4* gene under the control of a tissue-specific promoter: the Glass Multimer Reporter Gal4 (GMR-Gal4) line that expresses the GAL4 protein in the photoreceptors, the *nanos*-Gal4 (NGT-Gal4) line that expresses the GAL4 protein in the ovaries and the Embryonic lethal abnormal vision Gal4 (Elav-Gal4) line that expresses the GAL4 protein in post-mitotic neurons. Virgin females carrying one of these GAL4 expressing transgenes were crossed with males carrying a *Nudix* silencing transgene expressed under the control of a UAS (Upstream activating sequence) promoter activated by the GAL4 factor. Here, silencing transgenes (RNAi) express inverted repeated sequences (IR) generating either a long double strand RNA (dsRNA) of about 200 nucleotides that is further processed into 21mers by the Dicer machinery, or a short interfering RNA (siRNA). Two different RNAi lines per each Nudix orthologues were tested (Table 1.) with each of the three GAL4 driver lines in order to increase the chance to detect a specific

phenotype associated with the silencing of a *Nudix* gene. Indeed, in some cases, RNAi lines produce off-target effects or inefficient silencing of the target gene.

Drosophila melanogaster	***Homo sapiens***	**Predicted Function**	**RNAi fly line**
CG12567	NUDT 1	8-oxo-dGDP phosphatase activity, NUDIX hydrolase-like domain superfamily	BDSC ID: 57241 VDRC ID: 47420
CG31713 (Datp/Apf)	NUDT 2	Bis(5'-nucleosyl)-tetraphosphatase, NUDIX hydrolase-like domain superfamily	VDRC ID: 39150, 102412
CG6391 (Aps)	NUDT 4	diphosphoinositol-polyphosphate diphosphatase activity, NUDIX hydrolase-like domain superfamily	VDRC ID: 109431, 110607
CG8128	NUDT 6	ADP-Ribose diphosphatase, NUDIX hydrolase-like domain superfamily	VDRC ID: 47740, 107574
CG11095	NUDT 7	COA Pyrophosphatases, NUDIX hydrolase-like domain superfamily	BDSC ID: 67793 VDRC ID: 37911
CG4098	NUDT 9	ADP-ribose diphosphatase activity, NUDIX hydrolase-like domain superfamily	VDRC ID:105277, 110042
CG42813	NUDT 14	UDP-sugar diphosphatase,	BDSC ID: 56045, 57708
CG18094 (Nudt19A)	NUDT 19	COA Pyrophosphatases, NUDIX hydrolase-like domain superfamily	BDSC ID: 62908 VDRC ID: 100138
CG10194 (Nudt19B)	NUDT 19	COA Pyrophosphatases, NUDIX hydrolase-like domain superfamily	VDRC ID: 41170, 107721
CG10195 (Nudt19C)	NUDT 19	COA Pyrophosphatases, NUDIX hydrolase-like domain superfamily	VDRC ID: 26771, 109468
CG6169 (Dcp2)	NUDT 20	m7G(5')pppN diphosphatase activity, NUDIX hydrolase-like domain superfamily	VDRC ID: 22272, 105130
CG3689 (Cpsf5)	NUDT 21	hydrolase activity, NUDIX hydrolase-like domain superfamily	VDRC ID: 45278, 105499

Table 1. The 12 Nudix orthologues in *Drosophila melanogaster*.

The table contains the CG numbers of the 12 orthologues of the Nudix family in *Drosophila* (first column), the corresponding Nudix proteins in human (second column), the predicted function of the fly protein according to sequence similarity to the human orthologue (third column) and the identification code of the RNAi lines provided by the Vienna Drosophila RNAi stock Center (VDRC) or Bloomington Drosophila Stock Canter (BDSC) that were used in this work (forth column).

2.1. *Nudt20 (Dcp2)* silencing affects normal cell differentiation in fly photoreceptors

Silencing *Nudix* genes in the photoreceptors did not produce any marked eye phenotype except in the case of one out of the two *Nudt20* silencing transgenes (that correspond to *dcp2* orthologue in *Drosophila*). *Nudt20* silencing in the photoreceptors indeed produced a 'rough eye' phenotype, which is commonly attributable to apoptotic cells or mis-differentiation of photoreceptor cells (Figure 32). The eye phenotype observed upon *Nudt20* silencing show that

impairing *Dcp2* normal expression affects cell differentiation and/or survival as reported in another study[207].

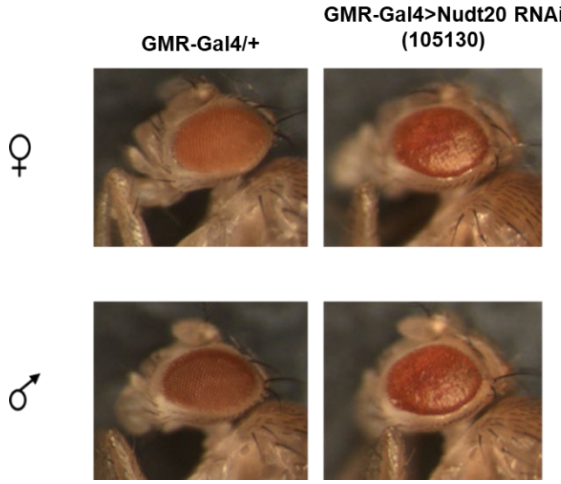


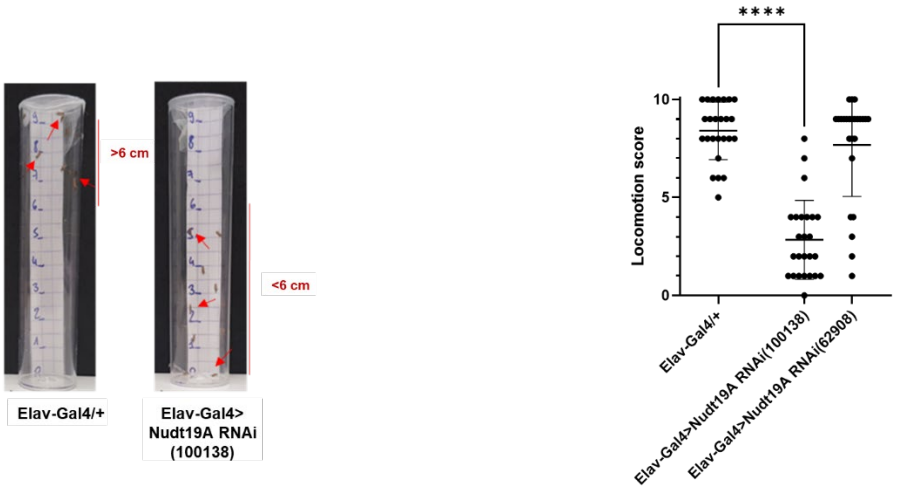
Figure 32. 105130 RNAi silencing of *Nudt20 (Dcp2)* in the photoreceptors results in 'rough eye' phenotype.
The pictures show the 'rough eye' phenotype observed in females (first row on the right) and males (second row on the right) upon 105130 RNAi silencing in the photoreceptors. Flies were grown at 30°C to increase GAL4 mediated expression of the RNAi transgenes.
Genotypes: *GMR-Gal4/+* is the abbreviation of *GMR-Gal4/+;+/+* and *GMR-Gal4>Nudt20 RNAi(105130)* is the abbreviation of *GMR-Gal4/+; UAS Nudt20 RNAi(105130)/+*.

Interestingly, only one of the two RNAi silencing transgenes targeting *Nudt20 (Dcp2)* showed the rough eye phenotype suggesting that probably the other RNAi transgene was poorly or not efficient.

2.3. Nudt19A is required for normal fly locomotion

Then, we investigated the effect of the silencing of each of the 12 *Nudix* genes in post-mitotic neurons by performing a locomotion test as described in Nichols et al.[343]. Ten males were placed in an empty vial and, after having gently hit the vial to ensure that the ten flies were at the bottom, a locomotion score ranging from 0 to 10 was calculated as the number of flies reaching or overcoming a distance of 6 cm from the bottom of the vial in 5 seconds (Figure 33). This measurement was repeated five times on the same group of ten flies with a 30 seconds rest between two subsequent runs and five independent groups were tested per genotype (see Materials and Methods). This screen revealed that silencing *Nudt20* (*Dcp2*) or *Nudt21* (*Cpsf5*) in post-mitotic neurons significantly impaired fly locomotion (Suppl. Figure 1). Furthermore, we identified a defective locomotion phenotype in flies carrying one of the two *Nudt19A* silencing transgenes (Figure 33A)

A.



B.

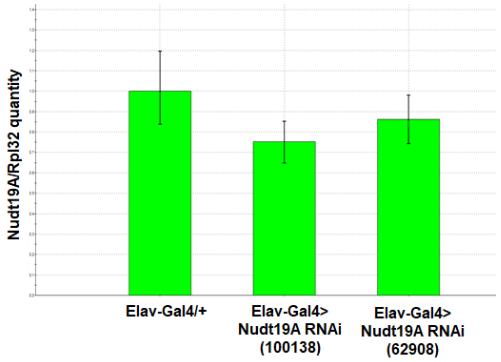


Figure 33. Fly locomotion is affected upon *Nudt19A* silencing in post-mitotic neurons.

A. The scattered plot indicates the locomotion score (5 test per group x 5 independent replicates= 25 dots per genotype). Statistical analysis was performed using GraphPad Prism. **** = p<0.0001. **B.** The histogram shows the RT-qPCR analysis of mRNA levels in male head lysates to check the level of expression of *Nudt19A* upon silencing by the Elav-Gal4 driver.

Genotypes: *Elav-Gal4/+* is the abbreviation of *Elav-Gal4/+;+/+* and *Elav-Gal4>Nudt19A RNAi(100138* or *62908)* is the abbreviation of *Elav-Gal4/+; UAS Nudt19A RNAi (100138 or 62908)/+*.

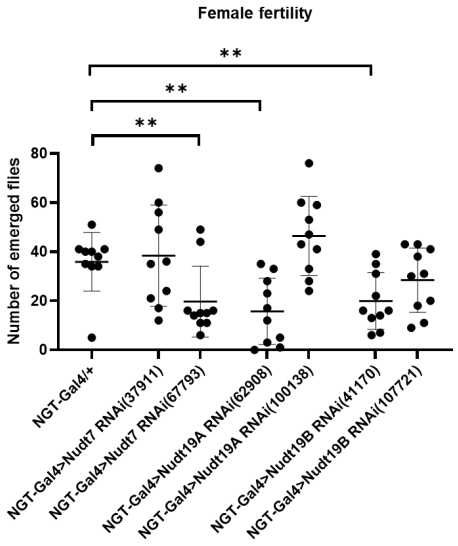
The phenotype observed correlated with a reduced expression levels of *Nudt19A* gene transcripts in the fly head in the 1000138 RNAi line (Figure 33B). The observed modest reduction of *Nudt19A* transcripts is attributable to the fact that the silencing was induced in neurons in the brain while I performed the RT-qPCR in the whole head.

The second RNAi line did not produce any locomotion defects and displayed no reduction in *Nudt19A* expression in the head suggesting that it was not or poorly efficient in inducing *Nudt19A* gene silencing in this particular setup.

2.4. Nudt19B depletion affects fly fertility

In a third set of experiments, we tested whether the silencing of the *Nudix* family genes may result in defective fertility. To this end, virgin females carrying both the NGT-Gal4 driver and each of the *Nudix* RNAi transgenes (*NGT-Gal4/Nudt RNAi)* were crossed with wild type males and the number of progeny was counted from each cross. I observed that silencing either *Nudt20 (Dcp2)* or *Nudt21 (Cpsf5)* in the ovaries significantly decreased female fertility confirming their known essential function in fly development and physiology[344,345] (Suppl. Figure 2). Moreover, I observed that at least one RNAi line out of two induced significant decrease of female fertility in the case of *Nudt7*, *Nudt19A* or *Nudt19B*, three genes with unknown function in *Drosophila* (Figure 34).

A.



B.

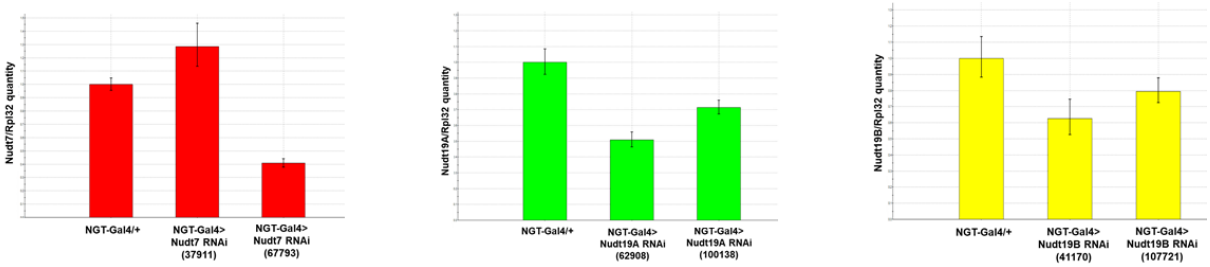


Figure 34. Female fertility decreases upon *Nudt7, Nudt19A* or *Nudt19B* silencing in *Drosophila* ovaries.

A. The scattered plots show the female fertility analysis expressed as the number of emerging flies in the progeny of the flies carrying *Nudt7*, *Nudt19A*, or *Nudt19B* silencing transgene. **B.** The RT-qPCR analysis was performed on ovaries of *NGT-Gal4/Nudt7* (CG11095)**,** *Nudt19A* (CG18094), or *Nudt19B* (CG10195) compared to control *NGT-Gal4/+*. Two RNAi lines targeting each *Nudt* gene were tested as indicated in the graphs. The statistical analysis was performed using GraphPad Prism9. *= $p<0.05$, **= $p<0.01$.

Genotypes*: NGT-Gal4/+* is the abbreviation of *NGT-Gal/+;+/+*, **(A)** *NGT-Gal4>Nudt7 RNAi(37911* or *67793)* is the abbreviation of *NGT-Gal4/+; UAS Nudt7 RNAi(37911 or 67793) /+,* **(B)** *NGT-Gal4>Nudt19A RNAi(62908* or *100138)* is the abbreviation of *NGT-Gal4/+;UAS Nudt19A RNAi(62908 or 100138)/+*, **(C)** *NGT-*

Gal4>Nudt19B RNAi(41170 or *107721)* is the abbreviation of *NGT-Gal4/+;UAS Nudt19B RNAi(41170* or *107721)/+*.

To evaluate the RNAi silencing efficiency, we tested the level of expression of *Nudt7*, *Nudt19A* or *Nudt19B* transcripts by RT-qPCR in total RNAs extracted from dissected ovaries. We noticed that the silencing of the candidate genes correlated with the phenotype even though the silencing efficiency range was modest, between 30 and 60% (Figure 34A, B, C). At this stage, given the fact that only one RNAi line produced a significant effect, we cannot exclude that reduced fertility could be due to off target effects.

It is worth mentioning that from sequence similarities, the Nudt7, Nudt19A and Nudt19B proteins are predicted to function as dpCoA hydrolases although it has not been shown experimentally whether or not they can indeed perform this role and if this enzymatic function is important for female fertility. *Nudt7* and *Nudt19* genes are conserved in mammals and it has been shown that recombinant mammalian NUDT7 and NUDT19 proteins hydrolyze dpCoA-RNA *in vitro*[160]. However, no decapping activity has been identified yet *in vivo*.

We then focused on a deeper analysis of the three *Nudt19* genes in *Drosophila*, named here as *Nudt19A*, *Nudt19B* and *Nudt19C* (Figure 35).

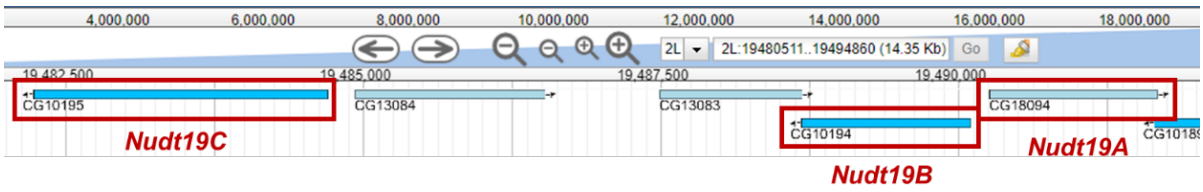


Figure 35. The *Nudt19A*, *Nudt19B* and *Nudt19C* genes localize very close to each other.

Schematic representation of *Nudt19A* (CG18094), *Nudt19B* (CG10194) and *Nudt19C* (CG10195) gene loci on the second chromosome in fly (Source: GBrowse/FlyBase).

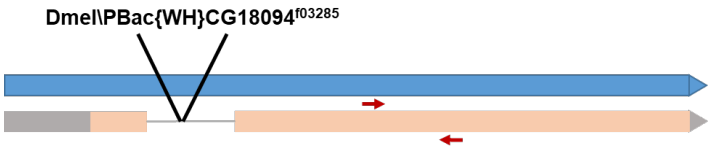
These three genes map to the left arm of the second chromosome (2L) in *Drosophila* and this triplication event is quite recent and common to the *Diptera* order. It is still unknown whether they have independent or complementary function in fly. By protein sequence alignment of the Nudt19 proteins, we found that Nudt19B displays the highest identity with the unique NUDT19 human orthologue (33%). Moreover, the fly Nudt19 proteins share approximately 34.5% of sequence identity (Suppl. Figure 3).

The previous experiments suggest a role of the Nudt19 proteins in fly development or physiology, although the low rate or absence of gene silencing and lack of reproducible phenotype between the two RNAi lines did not permit a clear conclusion.

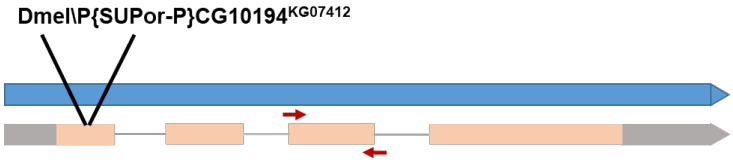
In order to further analyze the role of *Nudt19* genes in *Drosophila,* we ordered new lines carrying P element insertions in either the *Nudt19A (Nudt19A^{f03285})*, *Nudt19B (Nudt19B^{KG07412})* or *Nudt19C (Nudt19C^{f06699})* loci (Figure 36A) as well as a chromosomal deficiency covering the three loci *(Df(2L)ED1272)*. Using these lines, I generated trans-heterozygous flies carrying a P element inserted in either *Nudt19A*, *Nudt19B* or *Nudt19C* over the deficiency (*Nudt19A^{f03285}/Df(2L)ED1272;Nudt19B^{KG07412}/Df(2L)ED1272; Nudt19C^{f06699}/Df(2L)ED1272*). The purpose was to efficiently reduce the dosage of each *Nudt19* gene and study the resulting phenotype. Trans-heterozygous flies were fully viable in each case suggesting either that none of the *Nudt19* genes is essential for survival or that the P elements does not significantly affect gene expression (Suppl. Figure 4). No effect on male or female fertility was observed in the case of *Nudt19A* or *Nudt19C* trans-heterozygous flies (Figure 36B). On the contrary, male and female fertilities were significantly affected in *Nudt19B* trans-heterozygous flies (*Nudt19B^{KG07412}/Df(2L)ED1272*) (Figure 36B).

A.

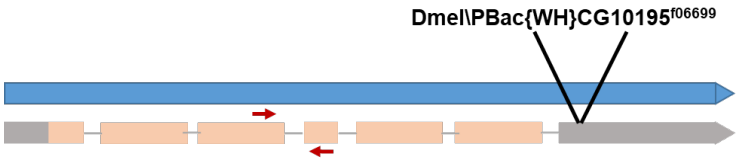
Nudt19A



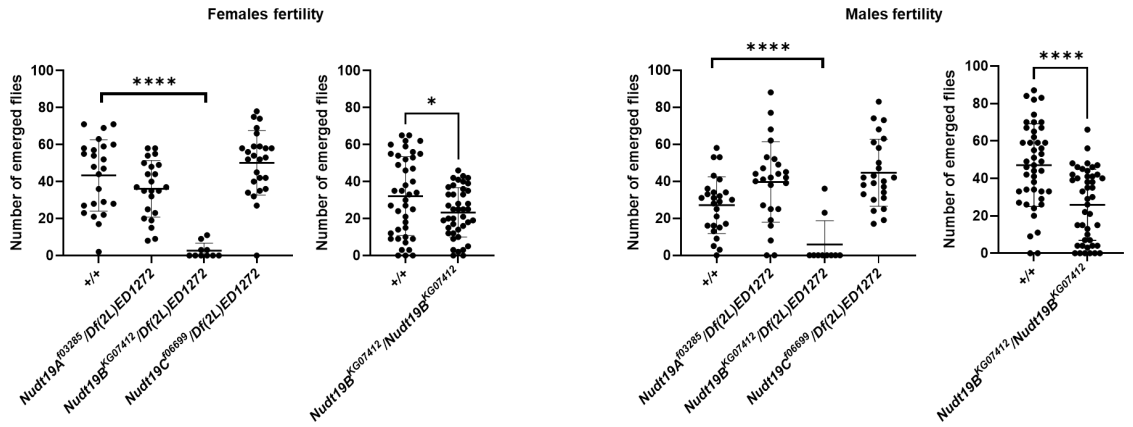
Nudt19B



Nudt19C



B.



C.

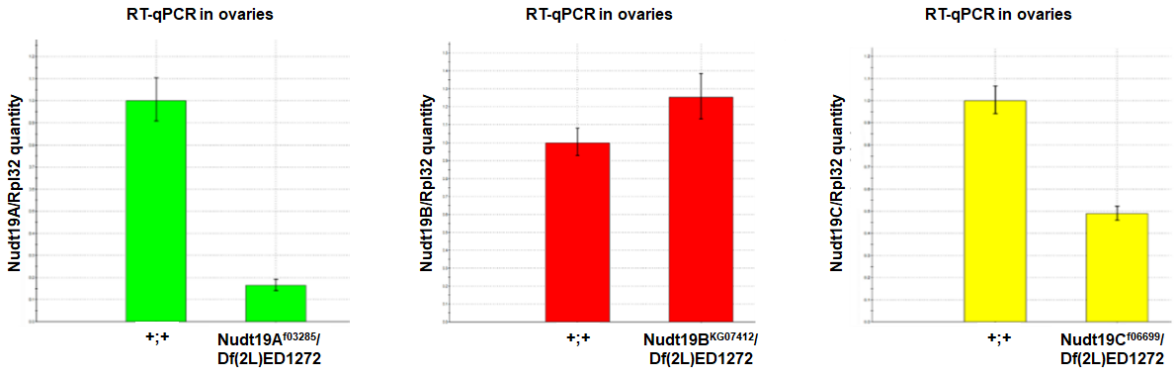


Figure 36. *Nudt19B* trans-heterozygotes display reduced fertility in *Drosophila*.

A. Schematic representation of the insertion of P element in the *Nudt19A*, *Nudt19B* and *Nudt19C* genes in the genome of the fly lines used in the study. The red arrows indicate the localization of the primers used for the RT-qPCR. **B.** The scattered plots show females (left) and males (right) fertility expressed as number of emerged flies from 25 individuals of the indicated genotypes (left). The statistical analysis was performed using Graph Pad Prism. *= $p<0.05$, ****= $p<0.0001$. **C.** The histograms show the level of expression of *Nudt19A*, *Nudt19B* or *Nudt19C* transcripts in the ovaries of the genotypes indicated.

Genotypes: +/+ is the abbreviation of w^{1118}/w^{1118};+/+, $Nudt19A^{f03285}/Df(2L)ED1272$ is the abbreviation of w^{1118}/w^{1118}; $Nudt19A^{f03285}/Df(2L)ED1272$, $Nudt19B^{KG07412}/Df(2L)ED1272$ is the abbreviation of w^{1118}/w^{1118}; $Nudt19B^{KG07412}/Df(2L)ED1272$, $Nudt19C^{f06699}/Df(2L)ED1272$ is the abbreviation of w^{1118}/w^{1118}; $Nudt19C^{f06699}/Df(2L)ED1272$ and $Nudt19B^{KG07412}/Nudt19B^{KG07412}$ is the abbreviation of w^{1118}/w^{1118}; $Nudt19B^{KG07412}/Nudt19B^{KG07412}$.

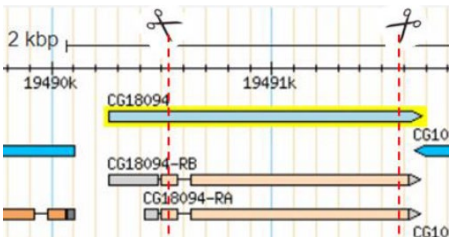
The fertility defect observed in $Nudt19B^{KG07412}/Df(2L)ED1272$ flies (Figure 36B) confirms the phenotype showed by RNAi silencing of *Nudt19B* in the ovaries (Figure 34B). Moreover, the fertility of $Nudt19B^{KG07412}/Nudt19B^{KG07412}$ homozygotes was reduced but not as much as that of $Nudt19B^{KG07412}/Df(2L)ED1272$ flies, indicating that it is not a null allele (Figure 36B). The level of the transcript did not decrease in $Nudt19B^{KG07412}/Df(2L)ED1272$ ovaries according to the

RT-qPCR analysis (Figure 36C center) which indicated that although the P element was inserted in the first exon of the *Nudt19B* gene, the mRNA is still expressed. Whether or not this mRNA is functional and able to produce the Nudt19B protein will be determined by Western blot analysis using antibodies directed towards Nudt19B that BIOTEM company is producing for this purpose. For *Nudt19A*^{f03285}*/Df(2L)ED1272* flies, we did not observed any fertility phenotype (Figure 36B), although the level of transcript was reduced by the 80% (Figure 36C left). In contrast, we had observed a fertility phenotype by RNAi silencing of *Nudt19A* in the ovaries, which was not confirmed in this analysis. Thus, we hypothesised that the fertility decrease observed by RNAi silencing of *Nudt19A* (Figure 34A) may be due to an off target effect and that Nudt19A may not be involved in fly fertility or that the P element insertion has not disrupted the gene. Finally, we observed no fertility phenotype in *Nudt19C*^{f06699}*/Df(2L)ED1272* flies (Figure 36B) in accordance with the absence of effects when expressing RNAi silencing transgenes of *Nudt19C* in the ovaries (Suppl. Figure 5). However, we observed only a partial reduction of Nudt19C transcript (Figure 36C right), probably due to fact that the P element is inserted in the 3′UTR of the gene. Thus, the level of transcription may be only slightly affected which may not allow to detect a putative physiological function of this gene. Moreover, there might be some functional redundancy between the three *Nudt19* genes. Overall, we have no strong indication to conclude whether Nudt19A, B and C are each essential or redundant (or even not contributing) to fly fertility at this stage of the study.

2.5. Generation of CRISPR-Cas9 mutant to characterize the *Nudt19A* and *Nudt19B* genes in *Drosophila*

In order to characterize further the function of *Nudt19* genes in *Drosophila*, sgRNAs driving the Cas9 endonuclease were designed to generate deletions of *Nudt19A*, *Nudt19B* or *Nudt19A* and *Nudt19B* together, (later on indicated as *Nudt19AB*) with the help of WellGenetics Inc., a company who took in charge the generation of CRISPR/Cas9 mutants (Figure 37A, B, C respectively).

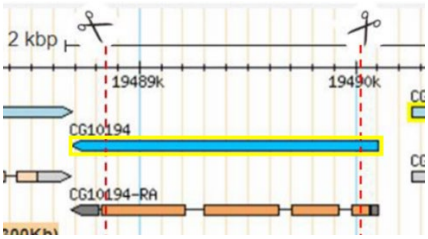
A. Nudt19A KO




```
>WT  ATGACTACTAAAACTGGAACTTATAGACCTTCGGCTAGCCTTATTGTAGCCGCCAAAGAGGATTCTTTAG/1020 nt/
>F29 ATGACTACTAAAACTGGAACTTATAGACCTTCGGCTAGCCTTATTGTAGCCGCCAAAGAGGATTCT-------------
>M46 ATGACTACTAAAACTGGAACTTATAGACCTTCGGCTAGCCTTATTGTAGCCGCCAAAGAGG-----------------
```

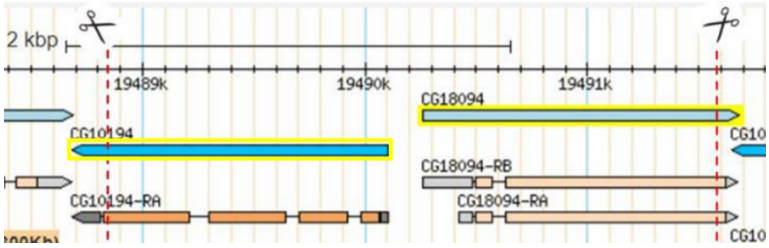
```
>WT  ACTGGCCTCCAATATACCGCCCTGCGATGGGCATCTGCCACTGCAGCCTCTTGTCAATAACAAGCTAGCCAAGTTGTAG
>F29 ----------------------------GGGCATCTGCCACTGCAGCCTCTTGTCAATAACAAGCTAACCAAGTTGTAG
>M46 -----------------------------GCATCTGCCACTGCAGCCTCTTGTCAATAACAAGCTAGCCAAGTTGTAG
```

B. *Nudt19B KO*




```
>WT  ATGTCGAGGCTTTTGCCCAAAATTCGCTCCTCCTCAAGTTTAATCCTTCTGGCTAAGAACCAAGTTGAAAAGTCCACG
>F8  ATGTCGAGGCTTTTGCCCAAAATTCGCTCCTCCTCAGGTT-----------------------------------
>M9  ATGTCGAGGCTTTTGCCCAAAATTCGCTCCTCCTCAAGTTTAATCCTTCTG-------------------------
>M20 ATGTCGAGGCTTTTGCCCAAAATTCGCTCCTCCTCAAGTTTAATCCTTC---------------------------
>M10 ATGTCGAGGCTTTTGCCCAAAATTCGCTCCTCCTCAAGTTTAATCCTTCT--------------------------


>WT  /1117 nt/TCGCAGCTGATAATTAAGTTTGAGCGGGATGATGGTCAGGTTCATCCATTGGATCCAACAAAATTATAG
>F8  ----------------------------------------------------CATCCATTGGATCCAACAAAATTATAG
>M9  ---------------------------------------------TCAGGTTCATCCATTGGATCCAACAAAATTATAG
>M20 ------------------------------------------------AGGTTCATCCATTGGATCCAACAAAATTATAG
>M10 -------------------------------AATCCCTAGGTTCCTTAGGTTCATCCATTGGATCCAACAAAATTATAG
```

C. *Nudt19AB KO*




```
>WT  CTAATAATTTTGTTGGATCCAATGGATGAACCTGACCATCATCCCGCTCAAACTTAATTATCAGCTGCGA/2699 nt/
>M17 CTAATAATTTTGTTGGATCCAATGGATGAACCTGACTTGTTGTTGTT-------------------------------

>WT  CTGGCCTCCAATATACCGCCCTGCGATGGGCATCTGCCACTGCAGCCTCTTGTCAATAACAAGCTAGCCAAGTTGTAG
>M17 ------------------------GATGGGCATCTGCCACTGCAGCCTCTTGTCAATAACAAGCTAACCAAGTTGTAG
```

Figure 37. Generation of CRISPR-Cas9 *Nudt19* knock out flies.

A-C. Schematic illustration of the cutting sites of the Cas9 enzyme to induce the mutation of **(A)** *Nudt19A* (CG18094), **(B)** *Nudt19B* (CG10194) or **(C)** *Nudt19A* and *Nudt19B* genes. The PAM sequences are highlighted in green, the sgRNAs sequences are highlighted in yellow and the inserted sequence in the *Nudt19AB KO* allele is highlighted in red. The start codon and the stop codons are indicated in bold.

Wellgenetics injected plasmids expressing these sgRNAs into transgenic embryos expressing the Cas9 endonuclease (Figure 37) and screened the putative mutants by genomic PCR. I characterize the mutant alleles by sequencing and confirmed the mutations for *Nudt19A* (two

independent alleles, Figure 37A), *Nudt19B* (four alleles, Figure 37B) and *Nudt19A* and *Nudt19B* together (one allele, Figure 37C).

All these alleles were then outcrossed for seven generations in order to remove off target mutations that the Cas9 endonuclease might have induced in the genome. These mutants will allow the in-depth analysis of the fertility and locomotion phenotypes identified by RNAi silencing or using the P element insertion in the *Nudt19* genes. Moreover, we will check if there are any detectable changes in CoA levels in these flies by Mass Spectrometry or in gene expression by RNA sequencing.

2.6. Cloning *Nudt19A*, *Nudt19B* and *Nudt19C* for recombinant protein expression to test dpCoA-decapping activity *in vitro* and for generating transgenic flies

In addition, I produced catalytic dead versions of the Nudt19A, Nudt19B and Ndt19C proteins by mutating the predicted catalytic residues, as shown in Shumar et al.[346]. In this work, the mutation of two Glutamate residues within the Nudix motif drastically reduced the catalytic activity of mouse NUDT19[346]. Thus, we induced missense mutations in the two conserved catalytic Glutamate residues within the Nudix motif of Nudt19A, Nudt19B and Nudt19C which were converted in two Alanine residues in order to abolish the enzymatic activity (Figure 38) using the Quick Change II Site-Directed mutagenesis Kit (see Materials and Methods).

Nudt19A

MTTKTGTYRPSASLIVAAKEDSLEDYDYRLLLIKRTEKTSYALNHCVFPGGVFDPIEDQSAKWITFFKSFGVTDEQLKMCRHN
QDSPRPEFLSGGDHISRDIALRLTALRETFEEVGILICTEQDDIQKWDSKSGHPRTVLLESSEHFEWQHRVHNDASQFLELFR
HYKVIPNIWSLQEWSIWRTAATANRKYDTVYYITMLDKYTRNIKLLLEPHEVASAHWLSPIEAWSSSQKAIIWLPFMLLYDIA
RLMNFYSFQELLNFSRQRSCNGSTMVQPVYYRCDDCMFGVLPGDELYPKEPGACTQTIILSGSVDDLHRKSKQYNRYIVYDFH
KVVLASNIPPCDGHLPLQPLVNNKLAKL

Nudt19B

MSRLLPKIRSSSSLILLAKNQVEKSTSCDYNALLLTRTQKSTFMPESSVFPGGVCDASDSSPAWLEHFQRNEFSAAKLRNVGH
VKGPRPDIFHTKADKKSLDPSLALRLTAIRETFEELGILLCRDSKSLTSTSDYGAFYDQFDRAHWQHIVHNNASQFLELCKQL
DVLPDVWSLHEWSVWRTPSTFKKRFETAFFMTALEQEPRVHIEPNEVKDSAWRSPLDYLQASLRKELWLPPPQFYELSRCLNF
SSLDNLRQFAAEREVKGIQLIHPVVHKCTNGLVHLLPGDDAYPADPDASNEKIEIDLSVEEFRSKTNAKLHRSEHWNQHQSQL
IIKFERDDGQVHPLDPTKL

Nudt19C

MSQPASPTKWRTSASVILVSRDDGNNKDYKVLMLKRSDATAIALNQTVFPGGLLDSGADESVAWLHYLEEFGVPQEALRRLVL
IREDRPAILAPQGTGCYDRFFKRSRIWAREITLRLTAVRECFEEVGLLLCRSRSQLDFGAVTCAQAVPDLESWQRRVHNKPAE
FLTLCRELNVVPDLWALHEWSAWASPGFIRKGHETVFFMAFVDKQPELLEEPSEVKETLWLTPVELLRLADLGNVWFMPPQVY
ELSRLMGIKAYQSLLEFAIKRSGLGTTMFLPIGYNCQGSMVFVLPGDDFYVPEPHLVHEIISFPGSEEEFRARSKHLHRYTYG
PGVRNLELNIAPPNGHLKPLKFQAERQKL

Figure 38. The predicted Glutamate catalytic residues are mutated in two Alanine residues.

The protein sequence of Nudt19A, Nudt19B and Nudt19C are represented. In yellow, the two predicted catalytic Glutamate residues which were mutated to Alanine residues according to sequence homology to the mammalian NUDT19 protein.

On one hand, the wild type and catalytic mutant *Nudt19* sequences were cloned in the pETM11SS-His6-Strep-SUMO-Tev expression vector for producing the recombinant proteins to perform an *in vitro* dpCoA-RNA decapping essay and check if they have a similar decapping activity *in vitro* as the mammalian NUDT7 and NUDT19. The advantage of using pETM11SS-His6-Strep-SUMO-Tev is double: first it expresses SUMO (Small Ubiquitin-like Modifier) which is a small protein of 12 kDa serving as a fused soluble protein; secondly the vector carries the Histidine and Streptavidin stretches as two alternative tags that can be used during the purification steps to increase the purification yield (Figure 39).

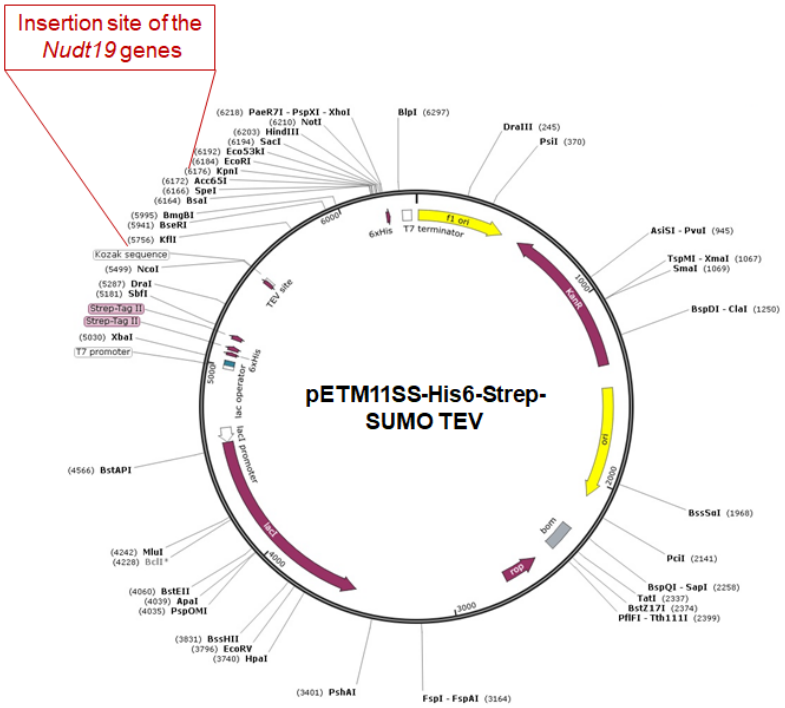


Figure 39. The pETM11SS-His6-Strep-SUMO-Tev vector contains the SUMO sequence and the His6 and Streptavidin tags.

The vector used for the generation of recombinant wild type and mutated Nudt19 proteins is shown in the figure. The insertion site of the *Nudt19* genes is indicated in red. The SUMO sequence localization is indicated together with the His6 and the Streptavidin tags. Tev is the recognition site for TEV protease to cleave the tags from the protein.

On the other hand, the wild type and catalytic mutant sequences were cloned in the pP[UAST] transgenesis vector (Figure 40) and sent to the BestGene company to generate transgenic flies expressing the Nudt19 proteins (wild type or mutated) under the UAS promoter.

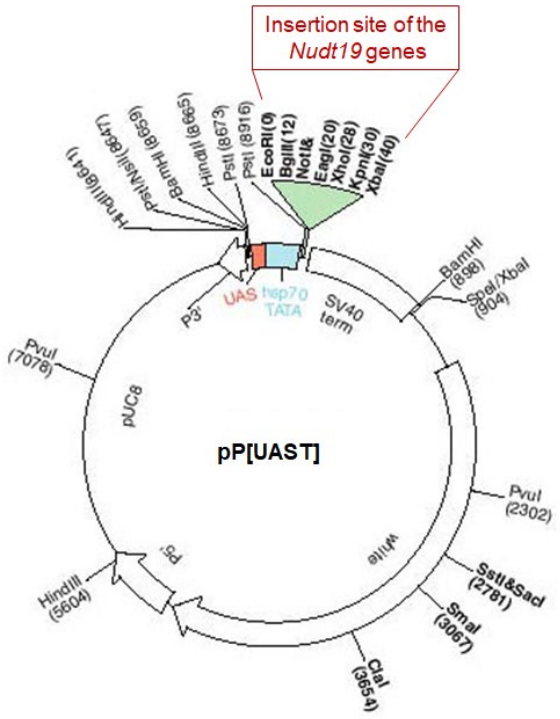


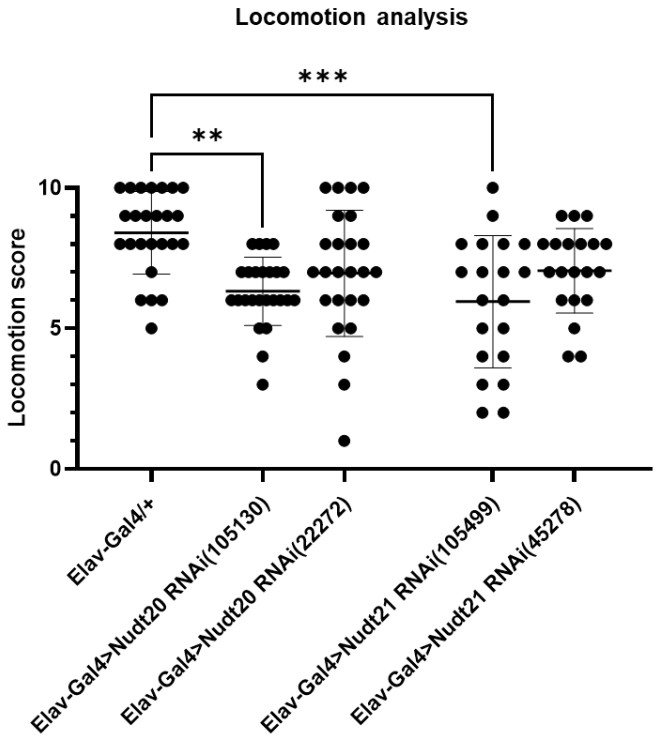
Figure 40. The pP[UAST] vector is used for generating transgenic flies.

The pP[UAST] vector used for transgenesis in *Drosophila* carries the UAS promoter upstream the insertion site of the *Nudt19* genes. P3′ and P5′ are P element ends that carry the *white* gene marker used for tracing the insertion of the gene in the fly acceptor genome.

These transgenic flies will be crossed with fly lines carrying a tissue specific driver (i.e. NGT-Gal4) in order to induce the expression of the Nudt19 proteins (wild type or mutated in a tissue-specific manner (i.e. in the ovaries)) in *Nudt19 KO* flies. The aim would be to rescue the fertility phenotype that we expect to see in *Nudt19 KO* mutants and that we have already observed in the *Nudt19B*^{KG07412}*/Df(2L)ED1272* and in *NGT-Gal4>Nudt19B RNAi (41170)*.

The phenotypic and molecular analysis of the *Nudt19* knock out mutants and *in vitro* testing of the Nudt19 decapping activity will contribute to shed light on the role of the conserved Nudt19 proteins in fly.

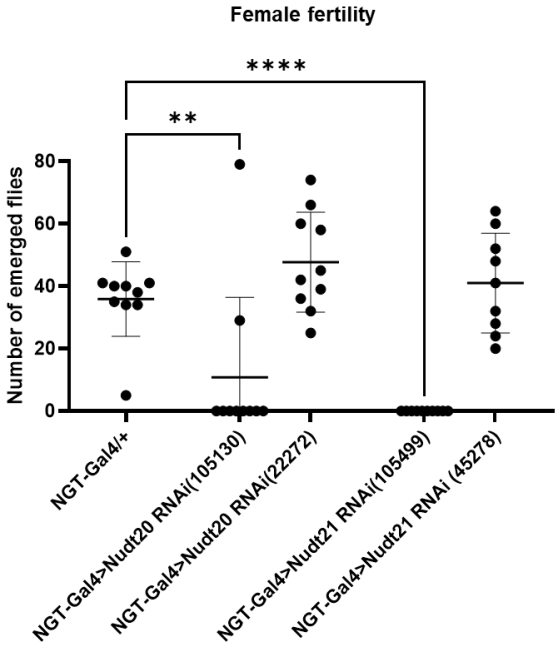
SUPPLEMENTAL FIGURES RESULTS-Chapter 2



Suppl. Figure 1. RNAi silencing of *Nudt20 (Dcp2)* and *Nudt21 (Cpsf5)* genes in *Drosophila* neurons affect fly locomotion.

The scattered plot indicate the locomotion score for each indicated genotype (see Materials and Methods). The statistical analysis was performed using GraphPad Prism. **= $p<0.01$, ***= $p<0.001$.

Genotypes: *Elav-Gal4/+* is the abbreviation of *Elav-Gal4/+ ;+/+, Elav-Gal4>Nudt20 RNAi (105130* or *22272)* is the abbreviation of *Elav-Gal4/+; UAS Nudt20 RNAi (105130* or *22272), Elav-Gal4>Nudt21 RNAi (105499* or *45278)* is the abbreviation of *Elav-Gal4/+; UAS Nudt21(105499* or *45278).*



Suppl. Figure 2. RNAi silencing of *Nudt20 (Dcp2)* and *Nudt21 (Cpsf5)* genes in *Drosophila* ovaries affect female fertility.

The scattered plot shows female fertility expressed as the number of emerged flies per each genotype indicated. The statistical analysis was performed using GraphPad Prism. **= p<0.01, ****= p<0.0001.

Genotypes: *NGT-Gal4/+* is the abbreviation of *NGT-Gal4/+ ;+/+, NGT-Gal4>Nudt20 RNAi (105130* or *22272)* is the abbreviation of *NGT-Gal4/+; UAS Nudt20 RNAi (105130* or *22272)/+, NGT-Gal4>Nudt21 RNAi (105499* or *45278)* is the abbreviation of *NGT-Gal4/+; UAS Nudt21(105499* or *45278)/+*.

A. CLUSTAL OMEGA (1.2.4) multiple sequence alignment

```
NUDT19        MSSSLRPGPSRWRRAASIVLAAGWSRPETATPPSRPPPAEGFRLLLLQRSPHQGFMPGAH    60
Nudt19A       ----MTTKTGTYRPSASLIVAAKE----------DSLEDYDYRLLLIKRTEKTSYALNHC    46
Nudt19C       --MSQPASPTKWRTSASVILVSR-----------DDGNNKDYKVLMLKRSDATAIALNQT    47
Nudt19B       ----MSRLLPKIRSSSSLILLAKNQV--------EKSTSCDYNALLLTRTQKSTFMPESS    48
                         * ::*::: :                  .:. *:: *:

NUDT19        VFSGGVLDAA-DRSADWLGLFAPHHGPPRFGLGPAPFSRTAFPSLPD----------TDD    109
Nudt19A       VFPGGVFDPIEDQSAKWITFFKSFG------VTDEQLKMCRHNQD-S----------PRP    89
Nudt19C       VFPGGLLDSGADESVAWLHYLEEFG------VPQEALRRLVLIREDRPAILAPQGTGCYD    101
Nudt19B       VFPGGVCDAS-DSSPAWLEHFQRNE------FSAAKLRNVGHVKGPRPDIFHTKA-----    96
              ** **: *   * *  *:  :          .    :

NUDT19        HKTDNTGTLPEDVAFRICAVREAFEEAGVLLLRPRTSPPGPAPG---PGLALEPPPGLAS    166
Nudt19A       EFLSGGDHISRDIALRLTALRETFEEVGILICTEQDDIQKWDSKSGHPRTVLLESSEHFE    149
Nudt19C       RFFKRSRIWAREITLRLTAVRECFEEVGLLLCRSRSQLDFGAVT------CAQAVPDLES    155
Nudt19B       ----DKKSLDPSLALRLTAIRETFEELGILLCRDSKSLTSTSD-----YGAFYDQFDRAH    147
                        .:::*: *:** *** *:*:     .

NUDT19        WRDRVRQDPRHFLRLCAHLDCTPDIWALHNWSAWLTPFLRGTTRRFDTAFFLCCLREP--    224
Nudt19A       WQHRVHNDASQFLELFRHYKVIPNIWSLQEWSIWRTAAT--ANRKYDTVYYITMLDKYTR    207
Nudt19C       WQRRVHNKPAEFLTLCRELNVVPDLWALHEWSAWASPGF--IRKGHETVFFMAFVDKQ--    211
Nudt19B       WQHIVHNNASQFLELCKQLDVLPDVWSLHEWSVWRTPST--FKKRFETAFFMTALEQE--    203
              *:  *::.  .** *  . .  *::*:*::** * :       : .:*.:::  : :

NUDT19        -PPVYPDLAEVVGYQWSSPSEATESFLSKEIWLPPPQFYEVRRLANFASLSDLHKFCLGR    283
Nudt19A       NIKLLLEPHEVASAHWLSPIEAWSSSQKAIIWLPFMLLYDIARLMNFYSFQELLNFSRQR    267
Nudt19C       -PELLEEPSEVKETLWLTPVELLRLADLGNVWFMPPQVYELSRLMGIKAYQSLLEFAIKR    270
Nudt19B       -PRVHIEPNEVKDSAWRSPLDYLQASLRKELWLPPPQFYELSRCLNFSSLDNLRQFAAER    262
                 :  :  **    * :* :         :*:    .*:: *  .: : ..* :*.  *
```

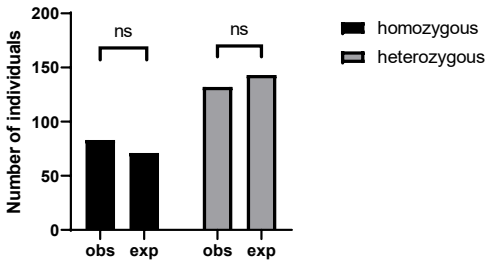
```
NUDT19       ALEGLERWLPIILLTADGMVHLLPGDELYLEDSDFLENLMSTEKKTEEIMKEGK-QFHRI    342
Nudt19A      SCNGSTMVQPVYYRCDDCMFGVLPGDELYPKEPGACTQTIILSGSVDDLHRKSK-QYNRY    326
Nudt19C      SGLGTTMFLPIGYNCQGSMVFVLPGDDFYVPEPHLVHEIISFPGSEEEFRARSK-HLHRY    329
Nudt19B      EVKGIQLIHPVVHKCTNGLVHLLPGDDAYPADPDASNEKIEIDLSVEEFRSKTNAKLHRS    322
                *     *:     . :. :****: *  :     : :    . :::  . : : :*

NUDT19       VTYHRHLYDIHVTVQPKYKHVYPKNSVVRKSH-L    375
Nudt19A      IVYDFHKVVLASNIPPCDGHLPLQPLVNNKLAKL    360
Nudt19C      TYG-PGVRNLELNIAPPNGHLKPLKFQAERQ-KL    361
Nudt19B      EHWNQHQSQLIIKFERDDGQVHPLDPTKL-----    351
                      :  ..     ::
```

B. Percentage of Nudt19 protein identity

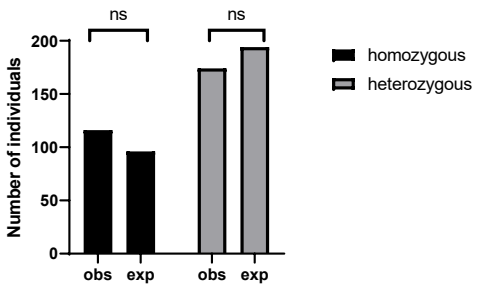
	Nudt19	Nudt19A	Nudt19C	Nudt19B
1: NUDT19	100.00	27.27	28.94	33.04
2: Nudt19A	27.27	100.00	34.48	33.04
3: Nudt19C	28.94	**34.48**	100.00	34.30
4: Nudt19B	**33.04**	33.04	34.30	100.00

Suppl. Figure 3. The Nudt19B protein shares the highest identity to the human NUDT19 orthologue.
A. Alignment of the Nudt19A, Nudt19B and Nudt19C proteins to the human NUDT19 was performed using Clustal Omega Multiple sequence alignment tool. * identical amino acids : similar amino acids . different amino acids. **B.** The matrix shows the percentage of identity of the four Nudt19 proteins considered. In bold red are indicated the highest percentages of identity.

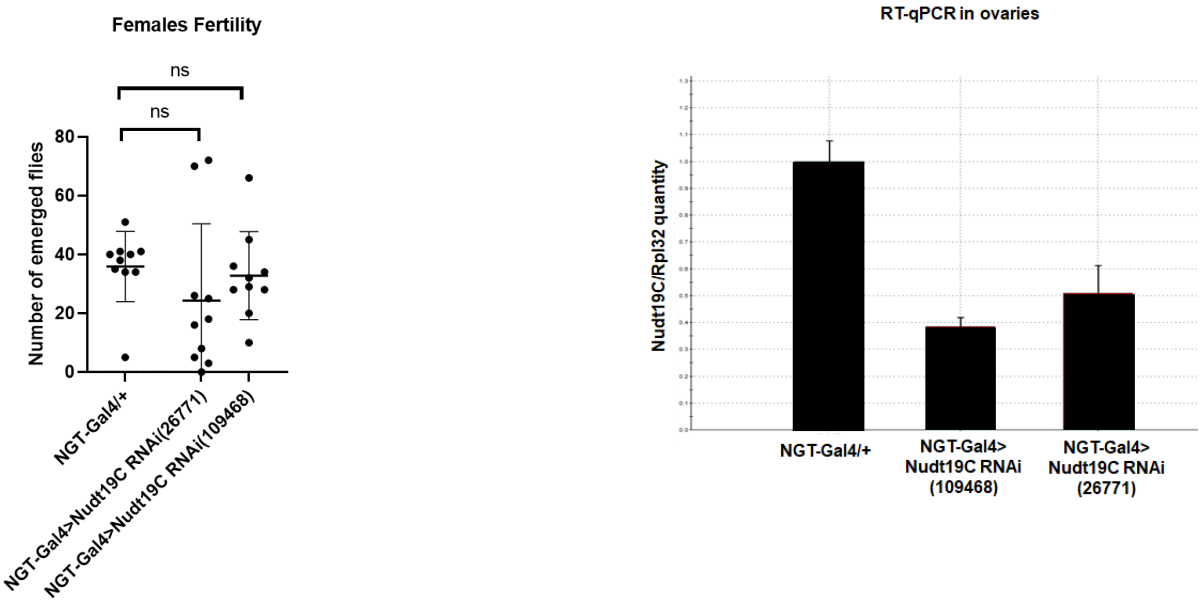
A.



B.



Suppl. Figure 4. *Nudt19A* and *Nudt19B* trans-heterozygous and heterozygous mutants are viable.
A-B. The histograms show the expected vs the observed number of individuals of **(A)** w^{1118}/w^{1118} ;*Nudt19A*^{f03285}*/Df(2L)ED1272* trans-heterozygotes (black bars) and w^{1118}/w^{1118} ;*Nudt19A*^{f03285}/+ or w^{1118}/w^{1118} ;*Df(2L)ED1272*/+ heterozygotes (grey bars) or **(B)** w^{1118}/w^{1118} ;*Nudt19B*^{KG07412}*/Df(2L)ED1272* trans-heterozygotes (black bars) and w^{1118}/w^{1118} ;*Nudt19B*^{KG07412}/+ or w^{1118}/w^{1118} ;*Df(2L)ED1272* /+ heterozygotes (grey bars). Chi-squared test analysis was performed using Graph Pad Prism9.



Suppl. Figure 5. RNAi silencing of the *Nudt19C* gene in *Drosophila* ovaries does not affect female fertility. The scattered plot shows female fertility expressed as the number of emerged flies per each indicated genotype (left). The RT-qPCR analysis was performed on ovaries of *NGT-Gal4>Nudt19C RNAi* compared to control *NGT-Gal4/+*. Two RNAi lines targeting *Nudt19C* gene were tested as indicated in the graphs. The statistical analysis was performed using GraphPad Prism9.

Genotypes*: NGT-Gal4/+* is the abbreviation of *NGT-Gal/+;+/+*, *NGT-Gal4>Nudt19C RNAi(109468* or *26771)* is the abbreviation of *NGT-Gal4/+; UAS Nudt19C RNAi(109468 or 26771) /+*.

DISCUSSION AND PERSPECTIVES

Chapter 1. Catalytic dead Pcif1 has a role in fertility and gene expression in *Drosophila*

1.1. Putative mechanism of action of the catalytic dead Pcif1 in gene regulation

One of the modifications that has attracted the attention of different groups in the recent years is the m^6Am modification, which has been detected on the first transcribed nucleotide if it is an adenosine[51,52,54]. It is 10 times less abundant in human cell and the writer protein involved in the deposition is the Phosphorylated CTD Interacting Factor 1 (PCIF1), previously known to bind to the CTD of the RNA Pol II[76,77]. It has been demonstrated that m^6Am modification contributes to mRNA stability, however it is still under debate whether it promotes or repress translation[52,54,78]. In mice, PCIF1 depletion does not affect viability nor fertility, but mutant animals show reduced body weight and RNA seq data indicate a set of dysregulated pseudogenes associated to the Y chromosome.

In *Drosophila melanogaster*, a single amino acidic substitution in the catalytic motif of Pcif1 renders it catalytically inactive[78]. Accordingly, no m^6Am modification exist in this organism. However, pull down essay reported that the Pcif1 WW-domain in fly is still able to bind the CTD of the RNA Pol II[78,145]. Thus, Pandey et al. suggested that Pcif1 in flies is probably involved in the interaction with the polymerase without any methyltransferase activity[78].

Other methyltransferase proteins are conserved during evolution but have developed additional roles as in the case of METTL3. It is a member of the nuclear methylation complex involved in the m^6A deposition on mRNA sites close to the 3' stop codon. When in the cytoplasm, it has shown the ability to promote translation by contributing to the formation of the mRNA loop structure by binding the eIFH at the 3'end of the mRNA[64]. Similarly, the cytoplasmic METTL16 is a methyltransferase involved in the deposition of the m^6A marks on U6 snRNAs as *MALAT1* and *XIST* and on *MAT2* mRNAs[68,71,347]. However, it is additionally involved in recruiting the eIF3a and b translation factors together with ribosomal RNAs to promote the assembly of the small and large ribosome subunits[348].

A possible mode of action of Pcif1 in regulating gene expression may be through the interaction with diphosphatases that are not able to dephosphorylate Ser5, inducing RNA Pol II stalling. Indeed, the CTD of RNA Pol II is first phosphorylated on Ser5 by TFIIH as soon as the transcription starts. Subsequently the RNA Pol II proceeds when Ser2 is phosphorylated by p-TEFb and Ser5 is dephosphorylated. A mistaken hyperphosphorylation status of the polymerase results in transcription blockage as it has been already demonstrated through the overexpression of Pin1[349]. Pin1 is able to inhibit the FCP1 phosphatase, which hydrolyzes the phosphate bond on Ser2, reducing the transcription rate. Similarly, PCIF1 is able to bind and inhibit the CTD dephosphorylation mediated by SCP1 phosphatase *in vitro*[350]. In our work, the upregulation of

genes upon *Pcif1* mutation observed by RNA sequencing and the de-repression of the heterochromatin-dependent w^{m4h} gene suggest that Pcif1 may contribute to gene expression repression. This repression may depend on interference with phosphatases in a similar fashion as Pin1.

In conclusion, this work has demonstrated that there is a link between Pcif1 role in gene expression and the phenotype observed in the PEV, although we missed the identification of the molecular mechanism which would explain the results we obtained.

1.2. Novel high-throughput technologies to unveil Pcif1 binding-enriched genes

As a perspective of my work to better characterize the role of *Pcif1* in gene expression regulation through its WW-domain, it should be analyzed whether its binding to the RNA Pol II is modulated by particular environmental conditions or cell types. This could be addressed by performing ChIP sequencing in S2 cells, which would identify the enriched genes of Pcif1 binding and integrate the results we obtained by RNA-seq in ovaries. Nonetheless, we have observed that Pcif1 binds along the chromosome rather uniformly so it might be trivial to determine whether there is any gene-specificity. Some recent technologies may help in unveiling Pcif1 specificity on RNA pol II transcribed genes binding. Native Elongating Transcript sequencing (NET-seq) does not use crosslinking step, the cells are lysed and the proteins involved in the transcription are bound to specific antibodies for immunoprecipitation. Then, the mRNAs trapped by the Pol II are reversed transcribed and then sequenced[351].

Alternatively, the Global Run-On sequencing (GRO-seq) exploits the *in vitro* transcription of the nuclear genome through the incorporation of Brome-Uridine step, giving a quantitative analysis of the RNA Pol II distribution and position[352].

With the phenotypic analysis performed during my PhD, we highlighted an undescribed involvement of the fly Pcif1 catalytic dead protein in gene expression regulation, which absolutely deserve further mechanistic studies. The use of S2 cell lines could be a preliminary step to perform the experiments proposed here and switch to *in vivo* analysis for validating the results obtained in cellulo.

1.3. Proximity labelling assay to identify Pcif1 interacting proteins in gene transcription regulation

Another perspective of my work would be to identify Pcif1 interacting proteins mediating gene transcription regulation. Apart from its interaction with the Pol II, mammalian PCIF1 has already been identified in binding and inhibiting the Pancreatic and Duodenal homeobox 1 (PDX1) protein, a transcription factor involved in the pancreatic β-cells differentiation which

secrete insulin[353,354]. Additionally, it has been shown that the WW-domain sequence of PCIF1 is conserved in the peptidyl-prolyl cis/trans isomerase 1 (Pin1) protein which also binds the P-Ser5 of RNA Pol II and inhibits the diphosphatase FCP1 by modulating the Cyclin Dependent Kinase 2 (CDK2) phosphorylation of Ser5[349]. These findings, together with the ability of the conserved WW-domain binding activity of Pin1 to some phosphorylated proteins such as the Tau and App involved in the onset of the Alzheimer disease[355], suggest that also PCIF1 may interact with other substrates.

Moreover, the interaction of PCIF1 with the polymerase is co-transcriptional and it is known that this process is modulated by the interaction with different factors including cap binding proteins, poly(A) binding proteins and translation initiating factors. It is thus reasonable to propose a proximity labelling assay using fused enzymes able to catalyze the biotinylation of proteins close to the protein of interest[356]. The two most common available enzymes are APEX (Ascorbate Peroxidase) or Biotin ligase and their modified versions to improve the labelling efficiency (Figure 41).

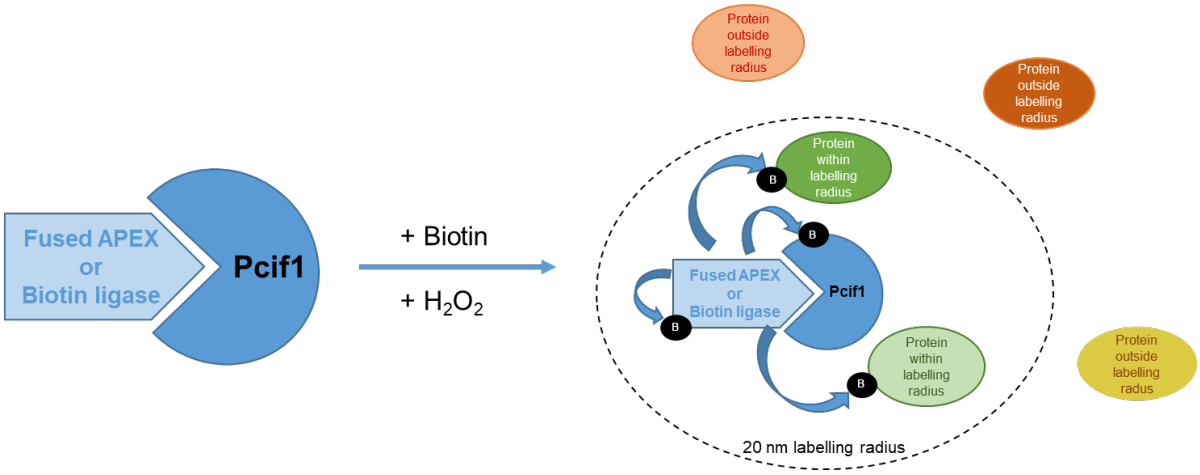


Figure 41. Protein labelling assay to identify putative Pcif1 interactors.

The graph illustrates a putative APEX protein or biotin ligase fused with the fly Pcif1 that in presence of biotin (and H_2O_2 if using APEX) in cellulo, it can drive the biotinylation of tyrosine (APEX) or lysine (biotin ligases) residues of proteins that are located within the labelling radius (20 nm).

In our case, APEX or biotin ligase would be fused with fly Pcif1 and in the presence of Biotin substrates (and H_2O_2, if using APEX) proteins within 20 nm would be biotinylated on their tyrosine (APEX) or lysine (biotin ligase) residues. The biotinylated proteins are then trapped with streptavidin beads and identified by Mass Spectrometry. In the last decade, the biotinylation binding assays have been applied in the identification of interacting proteins in bacteria, yeast, flies and mammals[356–359]. In conclusion, this system could shed light on a set of putative interacting proteins with fly Pcif1 starting from S2 cell culture, where putative

interacting proteins may not strictly binding to it but that are in its surrounding and probably involved in the modulation of Pcif1 activity.

Chapter 2. Functional role of Nudt19 proteins in fly physiology

As previously reported, DCP2 was considered the main decapping enzyme able to remove the 5' end m^7G cap. However, it has been recently demonstrated that other decapping enzymes exist in bacteria, yeast, plants and mammals suggesting that non-canonical cap structure can control the fate of RNA molecules[160]. Indeed, novel alternative cap structures have been identified *in vitro* or *in vivo* such as NAD, FAD, dpCoA, Ap_4A and UDP-Glc, suggesting a further layer of gene expression regulation by other members of the Nudix family[161].

The interest in studying the Nudix protein family members and their putative role in removing alternative caps form RNAs also rises from various evidences showing that there is a complex concerting system that regulates gene expression according to external stimuli and the metabolites availability in the cell. Moreover, various Nudix genes are conserved among prokaryotes and/or eukaryotes indicating that the selective pressure has maintained them for key role in gene expression regulation dependent on cellular metabolism that still need to be deciphered.

The aim of the second project of my PhD thesis was to identify novel decapping enzymes targeting alternative cap-structures by taking advantage of the *Drosophila melanogaster* model. First, I found that 12 out of 22 Nudix mammalian genes are conserved in the fly genome. To characterize their unknown function in fly, we used RNAi silencing strategy to knock down their expression and look for a potential effect on fly phenotype.

2.1. The *Nudt20 (Dcp2)* gene in *Drosophila* is required for proper photoreceptor shape

The genetic screen was designed to induce first the silencing of the *Nudix* genes in the photoreceptors. We observed that the silencing of *Nudt20* (*Dcp2*) induced the 'rough eye' phenotype, indicating that the *Dcp2* is essential for cell organization and survival in the developing eye. The eye development begins during the third instar larvae in the so called 'eye imaginal disc' where the eye differentiation starts by the cross talks of hedgehog (hh), decapentaplegic (dpp) and wingless (wg) signaling pathways[360]. The 'rough eye' phenotype has been identified as a consequence of silencing or impaired expression of many different genes, indicating that eye formation is a sensitive model for studying the function of proteins and signaling pathways involved in cell growth and differentiation.

In our study, we demonstrated that silencing the canonical decapping enzymes disrupts the normal eye conformation. Although we did not explored the effects at the molecular level, we can hypothesize that the stability of a large set of mRNAs is profoundly affected in the absence of Dcp2 thus perturbing normal eye development.

Surprisingly, we did not observed any eye phenotype when silencing the other Nudix genes in the photoreceptors. We hypothesized that Nudix proteins other than Nudt20 are redundant or not active in eye development or that the other RNAi lines were not efficient enough to reveal their functional requirement.

2.2. *Nudt19A* may be required in post-mitotic neurons

Later, we investigated the effects of the 12 *Nudix* orthologues silencing induced by Elav driver in the adult fly neurons. We observed that the flies carrying the silencing of *Nudt20 (Dcp2)* or *Nudt21 (Cpsf5)* showed a significant defect in locomotion indicating that the impairment of genes involved in mRNA metabolism are detrimental for fly physiology as expected. Moreover, it has been shown that worms carrying loss of function mutation in *Dcp1* and *Dcp2* are viable but demonstrated reduced locomotion[361] while neuronal-specific depletion of *Dcp1* and *Dcp2* affect normal axon development and regrowth after injury[362]. In accordance with these results, *Dcp1* specific silencing in fly neurons result in shorter lifespan and defective wing morphogenesis[207].

Here, we additionally identified a defective locomotion following the expression of a *Nudt19A* silencing transgene, whose phenotype correlated with the decreased transcript levels as observed by RT-qPCR. This suggests that Nudt19A may be required for proper neuronal function in post-mitotic neurons, a hypothesis that stays to be investigated with the recently obtained CRISPR/Cas9 mutants given the fact that only one out of the two RNAi line presented this defect.

2.3. *Nudt19B* is required for full fertility in fly

Lastly, we have induced the 12 *Nudix* orthologues silencing in the ovaries through the NGT driver and observed that the fertility was significantly affected when silencing transgene targeting either *Nudt7*, *Nudt19A* or *Nudt19B* was expressed. Although, also in this case, only one out of the two silencing transgene produced a phenotype. Reassuringly, we observed that reduced expression of either the *Nudt7*, *Nudt19A* or *Nudt19B* correlated with the decreased fertility phenotype strongly suggesting a role of the corresponding genes in full fertility. We speculated that the second RNAi line probably had poor or null silencing efficiency although we cannot exclude at this stage that reduced fertility would be due to off-target effects of the efficient silencing transgenes. Strikingly, the RNAi line targeting *Nudt19A,* which was efficient in post-mitotic neurons and induced a locomotion phenotype, was inefficient in silencing *Nudt19A* in the ovaries (and viceversa) further questioning the relevance of the phenotypic defects observed with these two RNAi lines. Alternatively, we may induce suboptimal silencing

generating a poorly robust phenotype, pointing out the absolute requirement to produce null mutants.

2.4. Nudix proteins inducing fly fertility defect all share a common putative enzymatic function

The predicted function of Nudt7 and Nudt19 proteins in *Drosophila* is to hydrolyse CoA metabolites from sequence similarities. Despite there are no experimental evidences demonstrating this enzymatic activity in fly model (FlyBase database), the mammalian NUDT7 and NUDT19 showed dpCoA decapping activity *in vitro*[160]. NUDT7 is a peroxisomal protein and it is highly represented in liver while NUDT19 localizes in the peroxisomes but it is mostly expressed in the kidney[264]. In peroxisomes, oxidation of fatty acids occurs and this is a relevant step in cellular metabolism the dynamic of which is finely regulated. It is difficult to say at this stage of the study and according to the literature if fly Nudt7 and Nudt19 proteins are hydrolases involved only in the fatty acid metabolism or if they can play a role in mRNA regulation. In both cases, this could profoundly affect fertility and/or neuronal function leading to the phenotypes observed in this study. At the present day, only in *E. coli* and *S. venezuelae* CoA-RNAs have been isolated but their role is still unknown[266].

Our screen provides an unprecedented general overview of the effect of silencing of the 12 Nudix orthologues in flies from a phenotypic point of view, despite some limitations in the interpretations of the phenotypes obtained or of the absence of phenotypes observed following the expression or silencing transgenes. Indeed, it is important to highlight that for all the positive genes we identified, only one line out of two RNAi lines tested produced a phenotype, indicating that the other line may be poorly or not efficient in reducing gene expression, or raising the question of a putative off-target effect of the silencing transgene that produced a phenotype. These results therefore point out towards some limitations in the use of RNAi-mediated gene silencing with poorly efficient gene extinction in some cases.

Still, this screen allowed to select few candidate genes belonging to the CoA hydrolase family for a deeper phenotypic characterization. Indeed, we ended up with the three *Nudt19* genes, which are orthologues of the mammalian *nudt19* and specific to the *Diptera* order.

In conclusion, we found that Nudt19B depletion in fly ovaries affects fertility but the identification of the molecular causes of this phenotype needs further investigation. It is important to note that any fertility phenotype associated to the NUDT19 proteins has never been demonstrated, opening up new avenues in the understanding of their biological function in living organisms.

2.4. CRISPR-Cas9 knock out mutants may validate the fertility phenotype

The phenotype observed by depleting Nudt19B in fly was promising and the triplication of the *Nudt19* genes observed in the *Drosophila* genome was intriguing. We thus designed a CRISPR-Cas9 knock out mutagenesis strategy to create single or combined mutants of *Nudt19A* and *Nudt19B* genes that were generated by WellGenetics Inc. In particular, we designed the coupled deletion of the two *Nudt19A* and *Nudt19B* loci in case of redundancy of the two genes. These knock out mutants should allow us to characterize deeply the fly Nudt19A and Nudt19B function in physiology and determine if they play a metabolic and/or a decapping role in *Drosophila* by testing the putative accumulation of CoA-RNA.

Fertility and locomotion analysis will be thus performed using the *Nudt19 KO* flies in order to check if the phenotypes observed by RNAi silencing or using P elements can be reproduced and validate their functional roles in fly physiology. This would give a primary information about the function of *Nudt19* genes in *Drosophila*. Obviously, it must be checked by RT-qPCR that the transcripts levels are abrogated and that there is no expression of the corresponding protein by Western blot. To go further, it would be also very informative to perform RNA-Mass Spectrometry in order to verify whether there are detectable CoA-RNAs in *Nudt19 KO* mutants since we expect the depletion of the putative decapping enzymes. However, the interpretation of the results may be misleading since we may not be able to discriminate if there is no CoA-RNAs at all or if we are not able to detect them because of their low abundance. A complemental analysis would be furnished by metabolomics studies, in order to check whether lack of Nudt19 proteins may alter the fatty acid oxidation and recycling. In addition, we may observe dysregulation of genes involved in fertility or locomotion by RNA sequencing.

As a parallel study, we plan to perform rescueing experiments to check whether the fertility or locomotion phenotypes are truly due to the disruption of *Nudt19A* or *Nudt19B* genes respectively. This experiment tool will allow us to express the wild type *Nudt19* genes in *Nudt19* knock out mutants and likely confirm the fertility and locomotion analysis.

2.5. Biochemical assays to test dpCoA activity *in vitro* and *in vivo*

A sophisticated technique has allowed the detection of NAD-capped RNAs and FAD-capped RNAs: the NAD (or FAD)-capture sequencing[287]. This protocol includes RNA extraction and fragmentation followed by a click reaction which ensures the attachment of the biotin-alternatively capped-RNAs to streptavidin beads. Later, a reverse transcription step occurs for quantification of cDNA and NGS sequencing[287,363] (Figure 42).

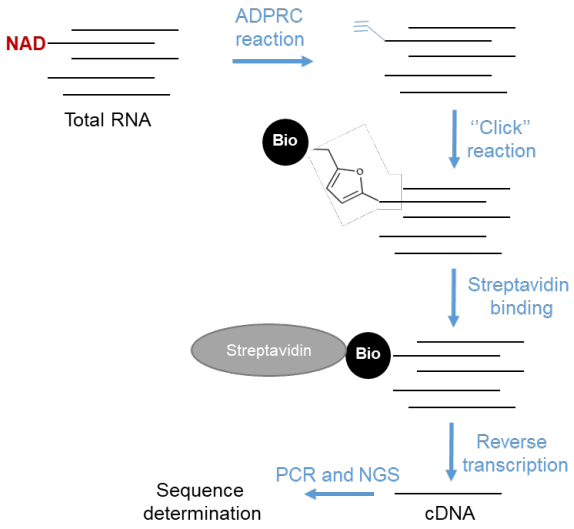


Figure 42. The NAD-capture sequencing.

The scheme shows the steps characterizing the NAD-capture sequencing. Total RNA is isolated and fragmented followed by Adenosine Diphosphate Ribocyclase (ADPRC) reaction which replace the nicotinamide with a pentynyl group. The substrate is then bound to biotin in the so called ''click'' reaction in order to pull down bound RNAs with streptavidin beads. Then, the reverse transcription occurs to generate cDNA ready to be amplified by PCR. The product is sequenced by NGS and the results are analyzed (Adapted from Cahova et al. 2015)[287].

Unfortunately, we have no idea of the amount of CoA-RNAs, if any, that may be present in *Nudt19 KO* flies which means that this technique could be applied only as a second step of analysis if RNA-Mass Spectrometry identifies the existence of CoA-RNAs in *Nudt19* mutants. Moreover, an additional obstacle could be choosing the best click reaction for specifically sequestering CoA-RNA and reduce the background noise. An alternative method or prior to the cap-capture sequencing could be the small molecule cleavage method which led to the identification of CoA-RNAs in *E. coli* and *S. venezuelae*[266]. In this protocol, total RNA is extracted, fragmented and subjected to a tandem size exclusion chromatography for larger and for smaller molecules. The aim is to obtain small capped RNAs for MS analysis. However, the results may include other cap-structures attached to small RNAs which could reduce the total yield of CoA-RNAs, if they are way less abundant than others.

In order to support the phenotypic and molecular experimentation with biochemical analysis of the Nudt19 proteins, we produced vectors expressing the wild type and the catalytic mutant sequences of the fly *Nudt19* genes in order to produce recombinant proteins for testing their decapping activity *in vitro*, if any. As previously proposed by Sharma et al., we may radiolabel the β-phosphate (^{32}P) of the dpCoA cap synthesized within the RNA stretch and perform decapping activity by adding the Nudt19 proteins. The resulting reaction will be loaded on Thin Layer Chromatography (TLC) or analyzed by Mass Spectrometry for checking if the decapping

activity has occurred. In this condition, the difficulty of the synthesis of (^{32}P)-dpCoA RNAs should be taken into account.

Quantitative analysis of the decapping activity may be performed by adding a colorimetric or fluorescent dye to the capped RNAs that would release a certain wavelength of fluorescence proportional to the amount of free caps and so to the decapping efficiency[293].

In conclusion, the combination of a huge variety of biochemical, molecular and genetic tools will be used for complementing and supporting the phenotypic investigation on the *Nudt19* genes that I have begun during my PhD thesis and would probably help in shed a light on their physiological function in living organisms.

MATERIALS AND METHODS

Table 2. Primers for PCR

TARGET GENE	SEQUENCE	SOURCE
Primer_CG7939 for	AAGAAGTTCCTGGTGCACAACGTG	Sigma
Primer_CG7939 rev	AATCTCCTTGCTTCTTGGAGGA	Sigma
Primer_CG11095 for	CGGCTACAGCGGTCCAGTATTTGT	Eurofins Genomics
Primer_CG11095 rev	TTGAGGCAATCGGGCAGCAAATTG	Eurofins Genomics
Primer_CG18094 for	TATTTGGCTCCCGTTCATGCTGCT	Eurofins Genomics
Primer_CG18094 rev	CCATTGTGCTTCCATTGCAGCTCC	Eurofins Genomics
Primer_CG10194 for	TGCCCAAAATTCGCTCCTCCTCCTCAA	Eurofins Genomics
Primer_CG10194 rev	AAGTGGACTCTGCGTGCGTGTAA	Eurofins Genomics
Primer-CG10195 for	ATTTGTGGGCTCGCATGAGTGGT	Eurofins Genomics
Primer-CG10195 rev	AAGGTTCTTCCAGCAATTCGGGCT	Eurofins Genomics
Primer-CG10194_XhoI for	CCCCCCTCGAGTTACCTAAATATGTCGAGGCTTTTGC	Eurofins Genomics
Primer-CG10194_NotI rev	CCCCGCGGCCGCAATGCACCTGACATAGAGGAAACC	Eurofins Genomics
Primer-GR-N19Afor	TTTCAGGGCGCCATGATGACTACTAAAACTGGACTT	Eurofins Genomics
Primer-GR-N19Arev	TCGAATTCGGATCCGCTACAACTTGGCTAGCTTGTT	Eurofins Genomics
Primer-CG10194_NcoI for	TTTCAGGGCGCCATGATGTCGAGGCTTTTGCCCA	Eurofins Genomics
Primer-CG10194_KpnI rev	TCGAATTCGGATCCGCTATAATTTTGTTGGATCCAATGGA	Eurofins Genomics
Primer-CG10195_NcoI for	TTTCAGGGCGCCATGATGAGCCAACCTGCATCTCC	Eurofins Genomics
Primer-CG10195_KpnI for	TCGAATTCGGATCCGTTACAGCTTTTGGCGCTCCG	Eurofins Genomics
Primer-Bgl-N19Afor	CCCCAGATCTATGACTACTAAAACTGGAAC	Eurofins Genomics
Primer-Not-N19Arev	CCCCGCGGCCGCCTACAACTTGGCTAGCTTGT	Eurofins Genomics
Primer-Bgl-N19Bfor	CCCCAGATCTATGTCGAGGCTTTTGCCCA	Eurofins Genomics
Primer-Not-N19Brev	CCCCGCGGCCGCCTATAATTTTGTTGGATCCAATGGA	Eurofins Genomics
Primer-Bgl-N19Cfor	CCCCAGATCTATGAGCCAACCTGCATCTCC	Eurofins Genomics
Primer-Not-N19Crev	CCCCGCGGCCGCTTACAGCTTTTGGCGCTCCG	Eurofins Genomics
Primer-PIPEN19Afor	[PHO]TTATTGTTGATAAACGAACTGAAAAAACTTCCTATGC	Eurofins Genomics
Primer-PIPEN19Arev	GCATAGGAAGTTTTTTCAGTTCGTTTAATCAACAATAATCTGTAG TCGTAGTCCTCTAAAGAATCC	Eurofins Genomics
Primer-N19A_KO for	CAAGCTCTTATTGCTTTGAGAC	Eurofins Genomics
Primer-N19A_KO rev	CAGACCGAGTTGGATAGTGG	Eurofins Genomics
Primer-N19B_KO for	ACCCTCGGATACCTCTGACA	Eurofins Genomics
Primer-N19B_KO rev	TGCACCTGACATAGAGGAAAC	Eurofins Genomics
Primer-N19A-N19BKO_for	AAATCCCAGCATGTCGATTC	Eurofins Genomics
Primer-N19A-N19BKO_rev	AGAACAAGGACGGTTTGTGC	Eurofins Genomics

Table 3. Primers for the Site directed mutagenesis

TARGET GENE	SEQUENCE	SOURCE
19A_E111A_E115A_for	CAAGATGCCCACCGCTTCAAAGGTTGCCCTGAGCGCAGTC	Eurofins Genomics
19A_E111A_E115A_rev	GACTGCGCTCAGGGCAACCTTTGAAGCGGTGGGCATCTTG	Eurofins Genomics
19B_E114A_E118A_for	CAGGATACCCAACGCCTCGAACCTTGCTCGTATGGCAGTC	Eurofins Genomics
19B_E114A_E118A_rev	GACTGCCATACGAGCAACCTTCGAGGCGTTGGGTATCCTG	Eurofins Genomics
19C_E123A_E127A_for	CAACAGACCCACCGCTTCAAAGCATGCCCTCACGGCGGTTAG	Eurofins Genomics
19C_E123A_E127A_rev	CTAACCGCCGTGAGGGCATGCTTTGAAGCGGTGGGTCTGTTG	Eurofins Genomics

Table 4. Reagent or resource

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial strains		
One Shot TOP10 competent cells	Invitrogen	Cat. No. C4040
Chemicals		
NucleoZOL	Macherey-Nagel	Cat. No. 740404-200
Critical Commercial Assays/Kits		
DNA, RNA and protein purification RNA	Macherey-Nagel	Cat. No. 740406-50
Affinity Script qPCR cDNA synthesis kit	Agilent Technologies	Cat. No. 600559
Brilliant II SYBR Green qPCR Master Mix	Agilent Technologies	Cat. No. 600828
Qia Miniprep Spin Kit (250)	Qiagen	Cat. No. 27104
NucleoBond Xtra Midi Prep kit	Macherey-Nagel	Cat. No. 740410-100
Gel and PCR Clean up kit	Macherey-Nagel	Cat. No. 740609-250
Rapid DNA Ligation kit	Thermo Fisher	Cat. No. K1422
Quick Change II site-directed kit	Agilent Technologies	Cat. No. 200523
Gibson Assembly Master Mix	NEBio Labs	Cat. No. E2611S
Experimental Models: Organisms/Strains		
Fly: Nudt knockdown (Table 4)	Bloomington stock center Vienna stock center	BDSC VDRC
w^{1118};PBac{w[+mC]=WH}CG18094[f03285]	BDSC	85214
y[1]w[67c23];P{y+mDint2]w[BR.E.BR]=SUPor-P}CG10194[KG0412]	BDSC	14625
w^{1118};PBac{w[+mC]=WH}CG10195[f06699]	BDSC	19001
w^{1118};Df(2L)ED1272,P{w[+mW.Scer\FRT.hs]=3'.RS5+3.3'}ED1272/SM6a	BDSC	24116
Oligonucleotides		
DNA oligos		See Table 1
Recombinant DNA		

pU6-BbsI-chiRNA		
pCRII vector		
pBlue Script KII vector		
pOT2 vector		
pFCLI vector		
pETM11-His_6 Strep SUMO TEV vector		Pr. Pillai Lab
pP[UAST] vector		
Software and Algorithms		
GraphPad Prism v.9		
Adobe Photoshop		
ImageJ		
CRISPR Optimal Target Finder		http://tools.flycrispr.molbio.wisc.edu/targetFinder/
Takara Primer Designing tool		https://www.takarabio.com/learning-centers/cloning/primer-design-and-other-tools

2.1. RNAi silencing of the *Nudix* family genes in *Drosophila melanogaster*

To induced a tissue-specific silencing of each of the 12 *Nudix* orthologues in *Drosophila,* we took advantage of three different existing fly lines carrying the yeast *Gal4* gene under the regulation of tissue-specific promoter: the Glass Multimer Reporter Gal4 (GMR-Gal4) line expresses the GAL4 protein in the photoreceptors , the *nanos*-Gal4 (NGT-Gal4) line expresses the GAL4 protein in the ovaries and the Embryonic lethal abnormal vision Gal4 (Elav-Gal4) expresses the GAL4 protein in post-mitotic neurons. Females carrying one of these constructs were crossed with males carrying the silencing transgene containing an inverted repeated sequence (IR) coding for a dsRNA or shRNA targeting each of the 12 *Nudix* orthologues under the control of the UAS promoter (*Nudt RNAi*). Two different RNAi lines targeting the same *Nudix* orthologue were tested for the phenotypic analyses (Table 1). The flies were grown on regular medium containing 60 g/L corn flour , 40 g/L yeast extract and 7g/L agar.

2.2. Phenotypic analysis of flies carrying *Nudix* gene silencing

2.2.1. Eye phenotype

Five virgin *GMR-Gal4/GMR-Gal4* were crossed with three *UAS Nudt RNAi* males at 25°C and flipped each 2 days for one week. The *GMR-Gal4/+; UAS Nudt RNAi/+* progeny was screened according to the eye phenotype and compared to control *GMR-Gal4/+;+/+* flies.

2.2.2. Fertility analysis

Five virgin *NGT-Gal4/NGT-Gal4* females were crossed with three *UAS Nudt RNAi* males at 25°C and flipped each 2 days for one week. In the progeny, 10 to 25 *NGT-Gal4/+; UAS Nudt RNAi/+* virgin females were individually crossed with three wild type $w^{1118}/w^{1118};+/+$ males at 25°C. After a mating period of 2 days, they were transferred to a new vial for 5 days and discarded. The emerging flies were counted and the statistical analysis was performed using the

one-way ANOVA test in GraphPad Prism9. This experiment was done for the two RNAi lines targeting each of the 12 *Nudix* orthologues.

2.2.3. Locomotion analysis

Five virgin *Elav-Gal4/Elav-Gal4* females were crossed with three *UAS Nudt RNAi/+* males at 25°C and flipped each 2 days for one week. In the progeny, five independent groups of ten *Elav-Gal4/+;UAS Nudt RNAi* males were collected for performing the locomotion assay (Protocol adapted from Nichols *et al*. 2012)[6].

Briefly, the locomotion test consisted in grouping ten adult males in a vial and after a gentle tapping to ensure that all the 10 flies were at the bottom, the number of flies overcoming 6 cm run in 5 seconds was recorded (locomotion score). Each group of ten flies was tested five times, with a resting time of 30 seconds after each run. Five independent replicates were performed for each genotype. The statistical analysis was performed using the one-way ANOVA test in GraphPad Prism9. The test was performed for the two RNAi lines targeting each of the 12 *Nudix* fly orthologues.

2.2.4. RT-qPCR

In order to validate the silencing efficiency, five pairs of *NGT-Gal4/+;UAS Nudt RNAi/+* ovaries or 10 *Elav-Gal4/+;UAS Nudt RNAi/+* male heads were dissected and used for RT-qPCR analysis. A couple of primers targeting each of the *Nudix* orthologue were designed to quantify the level of expression of the specific gene (Table 2). The analysis was conducted in the ovaries or heads of flies showing a significant decreased fertility or locomotion after *Nudix* silencing. Five pairs of ovaries or ten male heads were lysed in 300µl of NucleoZOL and mechanically homogenised using an electric pestler. The Nucleo Spin RNA Set for NucleoZOL kit (Macherey-NAGEL, Ca.No.740.406.50) was used for RNA purification according to the manufacturer instructions. The RNA concentration was measured by Nano Drop.

For the Reverse transcription step, 150 ng of RNA were reverse transcribed in a final volume of 10µl. The reaction mix was composed by 5µl of Master Mix 2X, 1,5µl of oligo-dT and 0.5µl of Reverse Transcriptase. The reaction lasts for 25 min (5 min at 25°C, 15 min at 42°C, 5 min at 95°C) in BIO-RAD T100 Thermal Cycler.

The qPCR reaction was performed in a 96 well plate in the STRAGENE Mx-3005P. The 20µl reaction mix contained: 12.5µl of SYBR-Brilliant qPCR mix, 10µg/µl of Primer forwards and 10µg/µl of Primer reverse and 6.3µl of deionised water. 100ng of cDNA were added in each reaction loaded on a 96 well plate. The plate was centrifuged and the cDNA was amplified by 45 cycle steps 10min at 95°C, 10s at 95°C and 30s at 60°C for 45 cycles, 1min at 95°C, 30s at

55°C and 30s at 95°C. The analysis of RNA quantification was performed using the MxPro-Mx3005P Comparative quantitation program.

2.3. Generation of flies carrying a P element insertion in one of the *Nudt19* genes and a chromosomal deletion covering the *Nudt19* locus

In order to study the phenotype of flies carrying a P element insertion in one of the *Nudt19* genes and a chromosomal deletion covering the *Nudt19* locus, the fly line were provided by the Bloomington Stock Center (BSC) and the genotypes are reported in Table 4.

25 individual females or males were crossed with three wild type $w^{1118}/w^{1118};+/+$ flies of the opposite sex as described in 2.2.2 and the RT-qPCR on fly ovaries was performed as in 2.2.4.

2.4. Cloning of *Nudt19A*, *Nudt19B* and *Nudt19C* genes for protein purification

In order to produce recombinant proteins for the *in vitro* catalytic activity analysis, we order the cDNAs coding for the *Nudt19* orthologues from the *Drosophila* Genomics Resource Center (DGRC) on Whatman FTA discs (cDNA #LD43867 for *Nudt19A* (CG18094) cloned in pOT2, cDNA #GH04973 for *Nudt19B* (CG10194) cloned in pOT2 and cDNA #RE56218 for *Nudt19C* (CG10195) cloned in pFLCI). The cDNA #LD43867 was mis-annotated and we could not use for *Nudt19A* construct. For this reason, we amplified the genomic sequence by genomic PCR of wild type fly genome using Bgl-N19A for and Not-N19A rev primers (Table 2). In order to remove the intron from the genomic sequence, we performed an additional PCR using PIPEN19Afor and PIPEN19Arev primers.

The Whatman discs containing the *Nudt19B* (CG10194) cloned in pOT2 or *Nudt19C* (CG10195) cloned in pFLCI cDNA were used for the excision of the cDNA by PCR. The primers used for the PCR amplification of the #GH04973 were Primer-Bgl-N19Bfor and Primer-Not-N19Brev (Table 2) while the primers used for #RE56218 were Primer-Bgl-N19Cfor Primer-Not-N19Crev (Table 3).

The cDNAs corresponding to Nudt19A, Nudt19B or nudt19C were subcloned in the pBlueScriptII SK+ vector (here referred to as pBS vector).

2.5. Cloning of *Nudt19A, Nudt19B* and *Nudt19C* sequences in pBS vector

Digestion of pBS vector using BamHI and NotI restriction enzymes was performed in a final volume of 30µl for 3h at 37°C. Subsequently the overhanging ends of pBS were dephosphorylated in order to avoid the closure of linearized plasmid by the addition of 1µl of Antartyk Phosphatase for 30 min at 37°C. Digestion of *Nudt19A, Nudt19B* or *Nudt19C* using BglII and NotI restriction enzymes was performed in a final volume of 30µl for 3h at 37°C. In

order to purify the pBS vector, and *Nudt19A*, *Nudt19B* and *Nudt19C*, we loaded the digestion volume onto 0.7% agarose gel 0.5% TBE and the bands corresponding to the DNA of interested were excised and purified as before. The DNA concentrations were measured at Nano Drop and adjusted for the ligation step.

The ligation of the *Nudt19A*, *Nudt19B* or *Nudt19C* genes to the pBS vector was performed using the Rapid Ligation kit according to the manufacturer instructions (Thermo Fisher, Cat. No. K1422). 25µl of One Shot TOP10 competent bacteria were transformed using 2.5µl ligated DNA for 30min at 4°C as before. The bacteria were then heated-shocked for 30min at 42°C and placed back for 2min at 4°C as reported before.

Six independent colonies were randomly dissolved in 3ml of LB + Amp medium and shake overnight at 37°C. 2ml of liquid bacteria were used to purify the pBS-*Nudt19A*, pBS-*Nudt19B* and pBS-*Nudt19C* construct as before.

In order to check the size of the insert, the pBS-*Nudt19A*, pBS-*Nudt19B* and pBS-*Nudt19C* constructs were digested using XhoI and NotI in a final volume of 10µl for 1h at 37°C. The digested DNA was then loaded onto 0.7% agarose gel 0.5% TBE. The DNA was send to Eurofins for sequencing in order to confirm the nucleotide sequence of wild type *Nudt19A*, *Nudt19B* and *Nudt19C* genes.

2.5. Site directed mutagenesis for generating catalytically-dead Nudt19 fly proteins

In order to produce catalytic dead version of Nudt19 fly proteins, we generated point mutations in the predicted catalytic motif of the *Drosophila* Nudt19 proteins using the Quik Change II Site-Directed Mutagenesis Kit (Agilent Technologies, Cat. No. 200523). According to Shumar *et al.* we induced point mutations E->A on the two Aspartates in the catalytic residues putatively involved in the CoA hydrolysis of Nudt19 *in vitro*[9]. The Quik Change II Site-Directed Mutagenesis Kit consists in the PCR amplification of the DNA sequence of interest using primers containing the desired mutation. Briefly, we designed a couple of primers for each of the three fly *Nudt19* orthologues (Table 3).

The primers amplify the template through the *Pfu Ultra HF DNA polymerase* (2.5 U/µl) and the reaction buffer 10X to a final volume of 50µl. The PCR amplification protocol was adapted to the length of the dsDNA template *Nudt19A*, *Nudt19B* and *Nudt19C* cloned in pBS vector (4kb->16 cycles): 95°C 30 sec, 55°C 1 min, 68°C 4min, ice. The PCR products were than digested using Dpn1 restriction enzymes, in order to degrade the remaining original template for 1h at 37°C. 1µl of digested dsDNA was used to transform XLC1 competent bacteria for 30min on ice. Than they were heat-shocked for 45sec at 42°C and placed on ice for 2min as reported before.

Liquid cultures were used to purify the mutagenized DNA constructs by the Qia Miniprep kit as reported before. The DNA concentrations were measured at the Nano Drop and adapted for further digestion in order to check the size of the inserts as presented before. The *Nudt19* genes clones in pBS vector were sent for sequencing for validation.

2.6. Cloning of *Nudt19A*, *Nudt19B* and *Nudt19C* sequences in pETM11-His6-Strep-SUMO TEV vector (pETM11)

In order to express the recombinant Nudt19 proteins, we used the pETM11-His6-Strep-SUMO TEV vector (here after referred as pETM11) to clone the *Nudt19* genes.

The pETM11 vector was digested using KpnI and NcoI restriction enzymes to a final volume of 10 µl. Briefly, the digestion occurred in the presence of 0.5 µl of each of the two restriction enzymes, 1 µl of the buffer 3.1 10X and 500 ng of purified vector.

The insertion of the *Drosophila* purified *Nudt19* genes in pETM11 was performed by Gibson assembly due to recombination and ligation of overhanging complementary sequences of the insert and the vector (NEBio Labs, Cat. No. E2611S). Briefly, a couple of PCR primers were designed according to the Takara Primers designing Tool https://www.takarabio.com/learning-centers/cloning/primer-design-and-other-tools. The forward primer contains the NcoI restriction site and the complementary sequence close to the ATG of the *Nudt19A, Nudt19B* or *Nudt19C* gene respectively. The reverse primer contains the KpnI restriction site and the complementary sequence close to the stop codon of the *Nudt19A, Nudt19B* or *Nudt19C* gene respectively. The PCR was performed by: Denaturation at 95°C for 30 sec, Annealing at 63°C for 30 sec, Elongation at 72°C for 1 min in a final volume of 50µl. The PCR products were purified using the Gel and PCR Purification kit as before. The DNA was eluted in 30 µl of NE buffer and the DNA concentration was measured at the Nano Drop. A further step of purification was performed by loading the purified PCR products onto 0.7% agarose gel. The bands corresponding to *Nudt19A* (1200bp), *Nudt19B* (1080bp) and *Nudt19C* (1060bp) were excised from the gel and purified for the Gibson assembly reaction.

The ligation of *Nudt19* orthologues in the pETM11 vector occurred in a final volume of 20 µl according to the manufacturer instructions. Briefly, 0.5 pmol of DNA were added (vector + insert), 10 µl of Gibson assembly master mix 2X (NEBio Labs, Cat. No. E2611S) and deionised water were supplied till the final volume. The reaction was held at 50°C for 15 min. As suggested by the manufacturer, we diluted the ligated DNA 4 times in deionised water and used 4µl for transforming 50µl of One shot TOP10 competent bacteria as described before.

200 µl of transformants were plated on LB + Kan agar plate and incubated overnight at 37°C. Single colonies were diluted in small cultures to check the efficiency of cloning and the positive

clones were used for producing large culture of bacteria expressing the proper constructs as previously described.

For the mutated *Nudt19A*, *Nudt19B* and Nudt19C cloning in pETM11 vector, a PCR reaction was performed by using GR_N19Afor and GR_N19Arev, CG10194_KpnIfor and CG10194_NcoIrev and CG10195_KpnIfor and CG10195_NcoIrev (Table 2). The PCR products were ligated to pETM11 vector by the Gibson assembly reaction as presented before and competent bacteria were used to express the constructs.

The *Nudt19A, Nudt19B* and *Nudt19C* (wild type and mutated) cloned in pETM11 vector were sent to sequence for validation. The constructs were sent to EMBL Heidelberg for further protein expression and purification.

2.7. Cloning of *Nudt19A, Nudt19B* and *Nudt19C* sequences in p[PUAST] vector for fly transgenesis

In order to inject *Nudt19A, Nudt19B* and *Nudt19C* DNA in flies, we cloned the three *Nudt19* genes in the transgenesis vector pP[UAST]. Each of the three *Nudt19* genes were amplified by specific primers generating overhanging sequences recognised by BglII and NotI (Table 2). The pP[UAST] vector and the three PCR products were digested by BglII and NotI in a final volume 30µl for 3h at 37°C. The pP[UAST] vector was dephosphorylated using 1µl of Antartyk Phosphatase for 30min at 37°C. The digested products were loaded on 0.7% agarose gel 0.5% TBE and the bands corresponding to the insert or to the vector were cut and purified using the Gel and PCR clean up kit as previously reported. The concentration of the three *Nudt19* genes and of the vector were measured by the Nano Drop.

The ligation of each of the three *Nudt19* genes with the pP[UAST] vector occurred in a final volume 20µl according to the rapid Ligation kit as reported before.

25µl of One Shot TOP10 competent bacteria were transformed with 2.5µl of ligated product for 30min on ice as reported before. The liquid cultures were used to purify the putative DNA using the Qia Miniprep kit and the purified DNA were digested using HindIII and NotI for 1h at 37°C.

For the mutant *Nudt19A*, *Nudt19B* and *Nudt19C* cloning in pP[UAST] vector, a PCR reaction was performed by using the Bgl-N19Afor and Not-N19Arev, Bgl-N19Bfor and Bgli-N19Brev and Bgl-N19Cfor and Bgl-N19Crev (Table 2). The PCR products were ligated to pP[UAST] vector by the Rapid ligation reaction as presented before and competent bacteria were used to express the constructs.

The *Nudt19A, Nudt19B* and *Nudt19C* (wild type and mutated) cloned in pP[UAST] vector were sent to sequence for validation. The constructs were sent to Best to produce the transgenic lines.

The genotype of the flies injected is *y[1] w[67c23]; P{y[+t7.7]=CaryP}attP2*. The flies were amplified at 25°C and stocked at 17°C.

2.8. Generation of *Nudt19* CRISPR-Cas9 mutant flies

The generation of *Nudt19* gene knock out flies was performed by WellGenetics Inc. Briefly they designed two sgRNAs targeting the *Drosophila melanogaster Nudt19* orthologues *Nudt19A* (CG18094), *Nudt19B* (CG10194) and *Nudt19A* and *Nudt19B* together in order to delete the target gene. The sgRNAs were injected in *nanos-Cas9* expressing embryos and the adult flies emerging were sent to us for further genotyping and sequencing.

Genomic PCR with specific primers (Table 2) were run to check the nature of the indel mutation induced (Annex). Reaction mix for 20 µl PCR reactions: 4 µl of 5X Phusion HF buffer (Thermo Scientific, Cat. No. F126L), 0.2 µl Phusion Hot StartII DNA Polymerase (2U/µl) (Thermo Scientific, Cat. No. F549S), 2.0 µl primer mix (stock 10 µM each), 0.4 µl oligo-dNTP (100µM), 2.0 µl genomic DNA, and 11.4 µl water to make up to 20 µl final volume. Reactions were run using the following conditions: 94°C for 4 min, 35 cycles of [94°C for 30 s, 52°C for 30 s and 72°C for 30 s], 72°C for 5 min, and finally at 12°C to hold the reaction. Reactions were examined by 2% agarose gel electrophoresis followed by sequencing.

In order to clean the genome of KO mutants from putative cryptic mutations induced by the CRISPR-Cas9 mutagenesis, all the alleles were outcrossed with *w[1118]; PBac{w[+mC]=WH}CG10195[f06699]* fly line, in the work indicated as $Nudt19C^{f06699}$ (Table 3). This fly line carries a P element in the *Nudt19C* (CG10195) gene which shows no phenotypic defects and it carries a red eye color marker used for tracing the mutation along the generation since the mutants carry a white eye color instead.

This process was repeated five times in order to remove the putative mutations all over the genome. Then, $Nudt19KO/Nudt19C^{f06699}$ males females were crossed with sibling females and *Nudt19KO/Nudt19KO* homozygous mutant flies were collected for further phenotypic and molecular analysis.

REFERENCES

(1) Boccaletto, P.; Machnicka, M. A.; Purta, E.; Piatkowski, P.; Baginski, B.; Wirecki, T. K.; de Crécy-Lagard, V.; Ross, R.; Limbach, P. A.; Kotter, A.; Helm, M.; Bujnicki, J. M. MODOMICS: A Database of RNA Modification Pathways. 2017 Update. *Nucleic Acids Res* **2018**, *46* (D1), D303–D307. https://doi.org/10.1093/nar/gkx1030.

(2) Ontiveros, R. J.; Stoute, J.; Liu, K. F. The Chemical Diversity of RNA Modifications. *Biochem J* **2019**, *476* (8), 1227–1245. https://doi.org/10.1042/BCJ20180445.

(3) Helm, M.; Motorin, Y. Detecting RNA Modifications in the Epitranscriptome: Predict and Validate. *Nat Rev Genet* **2017**, *18* (5), 275–291. https://doi.org/10.1038/nrg.2016.169.

(4) Ranjan, N.; Leidel, S. A. The Epitranscriptome in Translation Regulation: MRNA and TRNA Modifications as the Two Sides of the Same Coin? *FEBS Lett* **2019**, *593* (13), 1483–1493. https://doi.org/10.1002/1873-3468.13491.

(5) *Epitranscriptomics: RNA revisited*. https://www.science.org/content/article/epitranscriptomics-rna-revisited (accessed 2022-11-02).

(6) Gonatopoulos-Pournatzis, T.; Cowling, V. H. Cap-Binding Complex (CBC). *Biochem J* **2014**, *457* (2), 231–242. https://doi.org/10.1042/BJ20131214.

(7) Galloway, A.; Atrih, A.; Grzela, R.; Darzynkiewicz, E.; Ferguson, M. A. J.; Cowling, V. H. CAP-MAP: Cap Analysis Protocol with Minimal Analyte Processing, a Rapid and Sensitive Approach to Analysing MRNA Cap Structures. *Open Biol* **2020**, *10* (2), 190306. https://doi.org/10.1098/rsob.190306.

(8) Zhao, B. S.; He, C. Pseudouridine in a New Era of RNA Modifications. *Cell Res* **2015**, *25* (2), 153–154. https://doi.org/10.1038/cr.2014.143.

(9) Addepalli, B.; Limbach, P. A. Pseudouridine in the Anticodon of Escherichia Coli TRNATyr(QΨA) Is Catalyzed by the Dual Specificity Enzyme RluF* *This Work Was Supported by NIGMS, National Institutes of Health Grant R01 GM 058843 (to P. A. L.). The Authors Declare That They Have No Conflicts of Interest with the Contents of This Article. The Content Is Solely the Responsibility of the Authors and Does Not Necessarily Represent the Official Views of the National Institutes of Health. This Article Contains Supplemental Tables S1–S3 and Figs. S1–S8. *Journal of Biological Chemistry* **2016**, *291* (42), 22327–22337. https://doi.org/10.1074/jbc.M116.747865.

(10) Schwartz, S.; Bernstein, D. A.; Mumbach, M. R.; Jovanovic, M.; Herbst, R. H.; León-Ricardo, B. X.; Engreitz, J. M.; Guttman, M.; Satija, R.; Lander, E. S.; Fink, G.; Regev, A. Transcriptome-Wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of NcRNA and MRNA. *Cell* **2014**, *159* (1), 148–162. https://doi.org/10.1016/j.cell.2014.08.028.

(11) Toh, S.-M.; Mankin, A. S. An Indigenous Posttranscriptional Modification in the Ribosomal Peptidyl Transferase Center Confers Resistance to an Array of Protein Synthesis Inhibitors. *J Mol Biol* **2008**, *380* (4), 593–597. https://doi.org/10.1016/j.jmb.2008.05.027.

(12) Fujiwara, T.; Harigae, H. Pathophysiology and Genetic Mutations in Congenital Sideroblastic Anemia. *Pediatr Int* **2013**, *55* (6), 675–679. https://doi.org/10.1111/ped.12217.

(13) Carlile, T. M.; Rojas-Duran, M. F.; Zinshteyn, B.; Shin, H.; Bartoli, K. M.; Gilbert, W. V. Pseudouridine Profiling Reveals Regulated MRNA Pseudouridylation in Yeast and Human Cells. *Nature* **2014**, *515* (7525), 143–146. https://doi.org/10.1038/nature13802.

(14) Nance, K. D.; Meier, J. L. Modifications in an Emergency: The Role of N1-Methylpseudouridine in COVID-19 Vaccines. *ACS Cent Sci* **2021**, *7* (5), 748–756. https://doi.org/10.1021/acscentsci.1c00197.

(15) Li, X.; Xiong, X.; Wang, K.; Wang, L.; Shu, X.; Ma, S.; Yi, C. Transcriptome-Wide Mapping Reveals Reversible and Dynamic N(1)-Methyladenosine Methylome. *Nat Chem Biol* **2016**, *12* (5), 311–316. https://doi.org/10.1038/nchembio.2040.

(16) Dai, X.; Wang, T.; Gonzalez, G.; Wang, Y. Identification of YTH Domain-Containing Proteins as the Readers for N1-Methyladenosine in RNA. *Anal Chem* **2018**, *90* (11), 6380–6384. https://doi.org/10.1021/acs.analchem.8b01703.

(17) Zhang, C.; Jia, G. Reversible RNA Modification N1-Methyladenosine (M1A) in MRNA and TRNA. *Genomics Proteomics Bioinformatics* **2018**, *16* (3), 155–161. https://doi.org/10.1016/j.gpb.2018.03.003.

(18) Esteve-Puig, R.; Climent, F.; Piñeyro, D.; Domingo-Domènech, E.; Davalos, V.; Encuentra, M.; Rea, A.; Espejo-Herrera, N.; Soler, M.; Lopez, M.; Ortiz-Barahona, V.; Tapia, G.; Navarro, J.-T.; Cid, J.; Farré, L.; Villanueva, A.; Casanova, I.; Mangues, R.; Santamarina-Ojeda, P.; Fernández, A. F.; Fraga, M. F.; Piris, M.

A.; Kol, N.; Avrahami, C.; Moshitch-Moshkovitz, S.; Rechavi, G.; Sureda, A.; Esteller, M. Epigenetic Loss of M1A RNA Demethylase ALKBH3 in Hodgkin Lymphoma Targets Collagen, Conferring Poor Clinical Outcome. *Blood* **2021**, *137* (7), 994–999. https://doi.org/10.1182/blood.2020005823.

(19) Shi, Q.; Xue, C.; Yuan, X.; He, Y.; Yu, Z. Gene Signatures and Prognostic Values of M1A-Related Regulatory Genes in Hepatocellular Carcinoma. *Sci Rep* **2020**, *10* (1), 15083. https://doi.org/10.1038/s41598-020-72178-1.

(20) Gao, L.; Chen, R.; Sugimoto, M.; Mizuta, M.; Kishimoto, Y.; Omori, K. The Impact of M1A Methylation Modification Patterns on Tumor Immune Microenvironment and Prognosis in Oral Squamous Cell Carcinoma. *Int J Mol Sci* **2021**, *22* (19), 10302. https://doi.org/10.3390/ijms221910302.

(21) Bélanger, F.; Stepinski, J.; Darzynkiewicz, E.; Pelletier, J. Characterization of HMTr1, a Human Cap1 2'-O-Ribose Methyltransferase. *J Biol Chem* **2010**, *285* (43), 33037–33044. https://doi.org/10.1074/jbc.M110.155283.

(22) Langberg, S. R.; Moss, B. Post-Transcriptional Modifications of MRNA. Purification and Characterization of Cap I and Cap II RNA (Nucleoside-2'-)-Methyltransferases from HeLa Cells. *J Biol Chem* **1981**, *256* (19), 10054–10060.

(23) Bangs, J. D.; Crain, P. F.; Hashizume, T.; McCloskey, J. A.; Boothroyd, J. C. Mass Spectrometry of MRNA Cap 4 from Trypanosomatids Reveals Two Novel Nucleosides. *J Biol Chem* **1992**, *267* (14), 9805–9815.

(24) Dix, T. C.; Haussmann, I. U.; Brivio, S.; Nallasivan, M. P.; HadzHiev, Y.; Müller, F.; Müller, B.; Pettitt, J.; Soller, M. CMTr Mediated 2'-O-Ribose Methylation Status of Cap-Adjacent Nucleotides across Animals. *RNA* **2022**, *28* (10), 1377–1390. https://doi.org/10.1261/rna.079317.122.

(25) Haussmann, I. U.; Wu, Y.; Nallasivan, M. P.; Archer, N.; Bodi, Z.; Hebenstreit, D.; Waddell, S.; Fray, R.; Soller, M. CMTr Cap-Adjacent 2'-O-Ribose MRNA Methyltransferases Are Required for Reward Learning and MRNA Localization to Synapses. *Nat Commun* **2022**, *13* (1), 1209. https://doi.org/10.1038/s41467-022-28549-5.

(26) Ferron, F.; Decroly, E.; Selisko, B.; Canard, B. The Viral RNA Capping Machinery as a Target for Antiviral Drugs. *Antiviral Res* **2012**, *96* (1), 21–31. https://doi.org/10.1016/j.antiviral.2012.07.007.

(27) Netzband, R.; Pager, C. T. Epitranscriptomic Marks: Emerging Modulators of RNA Virus Gene Expression. *Wiley Interdiscip Rev RNA* **2020**, *11* (3), e1576. https://doi.org/10.1002/wrna.1576.

(28) Daffis, S.; Szretter, K. J.; Schriewer, J.; Li, J.; Youn, S.; Errett, J.; Lin, T.-Y.; Schneller, S.; Zust, R.; Dong, H.; Thiel, V.; Sen, G. C.; Fensterl, V.; Klimstra, W. B.; Pierson, T. C.; Buller, R. M.; Gale, M.; Shi, P.-Y.; Diamond, M. S. 2'-O Methylation of the Viral MRNA Cap Evades Host Restriction by IFIT Family Members. *Nature* **2010**, *468* (7322), 452–456. https://doi.org/10.1038/nature09489.

(29) Case, J. B.; Ashbrook, A. W.; Dermody, T. S.; Denison, M. R. Mutagenesis of S-Adenosyl-l-Methionine-Binding Residues in Coronavirus Nsp14 N7-Methyltransferase Demonstrates Differing Requirements for Genome Translation and Resistance to Innate Immunity. *J Virol* **2016**, *90* (16), 7248–7256. https://doi.org/10.1128/JVI.00542-16.

(30) Angelova, M. T.; Dimitrova, D. G.; Da Silva, B.; Marchand, V.; Jacquier, C.; Achour, C.; Brazane, M.; Goyenvalle, C.; Bourguignon-Igel, V.; Shehzada, S.; Khouider, S.; Lence, T.; Guerineau, V.; Roignant, J.-Y.; Antoniewski, C.; Teysset, L.; Bregeon, D.; Motorin, Y.; Schaefer, M. R.; Carré, C. TRNA 2'-O-Methylation by a Duo of TRM7/FTSJ1 Proteins Modulates Small RNA Silencing in Drosophila. *Nucleic Acids Res* **2020**, *48* (4), 2050–2072. https://doi.org/10.1093/nar/gkaa002.

(31) Dimitrova, D. G.; Teysset, L.; Carré, C. RNA 2′-O-Methylation (Nm) Modification in Human Diseases. *Genes* **2019**, *10* (2), 117. https://doi.org/10.3390/genes10020117.

(32) Agris, P. F. Bringing Order to Translation: The Contributions of Transfer RNA Anticodon-Domain Modifications. *EMBO Rep* **2008**, *9* (7), 629–635. https://doi.org/10.1038/embor.2008.104.

(33) Schaefer, M.; Pollex, T.; Hanna, K.; Lyko, F. RNA Cytosine Methylation Analysis by Bisulfite Sequencing. *Nucleic Acids Res* **2009**, *37* (2), e12. https://doi.org/10.1093/nar/gkn954.

(34) Edelheit, S.; Schwartz, S.; Mumbach, M. R.; Wurtzel, O.; Sorek, R. Transcriptome-Wide Mapping of 5-Methylcytidine RNA Modifications in Bacteria, Archaea, and Yeast Reveals M5C within Archaeal MRNAs. *PLoS Genet* **2013**, *9* (6), e1003602. https://doi.org/10.1371/journal.pgen.1003602.

(35) Squires, J. E.; Patel, H. R.; Nousch, M.; Sibbritt, T.; Humphreys, D. T.; Parker, B. J.; Suter, C. M.; Preiss, T. Widespread Occurrence of 5-Methylcytosine in Human Coding and Non-Coding RNA. *Nucleic Acids Res* **2012**, *40* (11), 5023–5033. https://doi.org/10.1093/nar/gks144.

(36) Khoddami, V.; Cairns, B. R. Identification of Direct Targets and Modified Bases of RNA Cytosine Methyltransferases. *Nat Biotechnol* **2013**, *31* (5), 458–464. https://doi.org/10.1038/nbt.2566.

(37) Hussain, S.; Sajini, A. A.; Blanco, S.; Dietmann, S.; Lombard, P.; Sugimoto, Y.; Paramor, M.; Gleeson, J. G.; Odom, D. T.; Ule, J.; Frye, M. NSun2-Mediated Cytosine-5 Methylation of Vault Noncoding RNA Determines Its Processing into Regulatory Small RNAs. *Cell Rep* **2013**, *4* (2), 255–261. https://doi.org/10.1016/j.celrep.2013.06.029.

(38) Yang, X.; Yang, Y.; Sun, B.-F.; Chen, Y.-S.; Xu, J.-W.; Lai, W.-Y.; Li, A.; Wang, X.; Bhattarai, D. P.; Xiao, W.; Sun, H.-Y.; Zhu, Q.; Ma, H.-L.; Adhikari, S.; Sun, M.; Hao, Y.-J.; Zhang, B.; Huang, C.-M.; Huang, N.; Jiang, G.-B.; Zhao, Y.-L.; Wang, H.-L.; Sun, Y.-P.; Yang, Y.-G. 5-Methylcytosine Promotes MRNA Export - NSUN2 as the Methyltransferase and ALYREF as an M5C Reader. *Cell Res* **2017**, *27* (5), 606–625. https://doi.org/10.1038/cr.2017.55.

(39) Rai, K.; Chidester, S.; Zavala, C. V.; Manos, E. J.; James, S. R.; Karpf, A. R.; Jones, D. A.; Cairns, B. R. Dnmt2 Functions in the Cytoplasm to Promote Liver, Brain, and Retina Development in Zebrafish. *Genes Dev* **2007**, *21* (3), 261–266. https://doi.org/10.1101/gad.1472907.

(40) Blanco, S.; Kurowski, A.; Nichols, J.; Watt, F. M.; Benitah, S. A.; Frye, M. The RNA-Methyltransferase Misu (NSun2) Poises Epidermal Stem Cells to Differentiate. *PLoS Genet* **2011**, *7* (12), e1002403. https://doi.org/10.1371/journal.pgen.1002403.

(41) Saletore, Y.; Meyer, K.; Korlach, J.; Vilfan, I. D.; Jaffrey, S.; Mason, C. E. The Birth of the Epitranscriptome: Deciphering the Function of RNA Modifications. *Genome Biol* **2012**, *13* (10), 175. https://doi.org/10.1186/gb-2012-13-10-175.

(42) Dominissini, D.; Moshitch-Moshkovitz, S.; Schwartz, S.; Salmon-Divon, M.; Ungar, L.; Osenberg, S.; Cesarkas, K.; Jacob-Hirsch, J.; Amariglio, N.; Kupiec, M.; Sorek, R.; Rechavi, G. Topology of the Human and Mouse M6A RNA Methylomes Revealed by M6A-Seq. *Nature* **2012**, *485* (7397), 201–206. https://doi.org/10.1038/nature11112.

(43) Ke, S.; Alemu, E. A.; Mertens, C.; Gantman, E. C.; Fak, J. J.; Mele, A.; Haripal, B.; Zucker-Scharff, I.; Moore, M. J.; Park, C. Y.; Vågbø, C. B.; Kusśnierczyk, A.; Klungland, A.; Darnell, J. E.; Darnell, R. B. A Majority of M6A Residues Are in the Last Exons, Allowing the Potential for 3' UTR Regulation. *Genes Dev* **2015**, *29* (19), 2037–2053. https://doi.org/10.1101/gad.269415.115.

(44) Beemon, K.; Keith, J. Localization of N6-Methyladenosine in the Rous Sarcoma Virus Genome. *J Mol Biol* **1977**, *113* (1), 165–179. https://doi.org/10.1016/0022-2836(77)90047-x.

(45) Chen-Kiang, S.; Nevins, J. R.; Darnell, J. E. N-6-Methyl-Adenosine in Adenovirus Type 2 Nuclear RNA Is Conserved in the Formation of Messenger RNA. *J Mol Biol* **1979**, *135* (3), 733–752. https://doi.org/10.1016/0022-2836(79)90174-8.

(46) Kane, S. E.; Beemon, K. Precise Localization of M6A in Rous Sarcoma Virus RNA Reveals Clustering of Methylation Sites: Implications for RNA Processing. *Mol Cell Biol* **1985**, *5* (9), 2298–2306. https://doi.org/10.1128/mcb.5.9.2298-2306.1985.

(47) Meyer, K. D.; Saletore, Y.; Zumbo, P.; Elemento, O.; Mason, C. E.; Jaffrey, S. R. Comprehensive Analysis of MRNA Methylation Reveals Enrichment in 3' UTRs and near Stop Codons. *Cell* **2012**, *149* (7), 1635–1646. https://doi.org/10.1016/j.cell.2012.05.003.

(48) He, L.; Li, H.; Wu, A.; Peng, Y.; Shu, G.; Yin, G. Functions of N6-Methyladenosine and Its Role in Cancer. *Mol Cancer* **2019**, *18* (1), 176. https://doi.org/10.1186/s12943-019-1109-9.

(49) Liu, L.; Wang, Y.; Wu, J.; Liu, J.; Qin, Z.; Fan, H. N6-Methyladenosine: A Potential Breakthrough for Human Cancer. *Mol Ther Nucleic Acids* **2020**, *19*, 804–813. https://doi.org/10.1016/j.omtn.2019.12.013.

(50) Adams, J. M.; Cory, S. Modified Nucleosides and Bizarre 5'-Termini in Mouse Myeloma MRNA. *Nature* **1975**, *255* (5503), 28–33. https://doi.org/10.1038/255028a0.

(51) Sun, H.; Zhang, M.; Li, K.; Bai, D.; Yi, C. Cap-Specific, Terminal N6-Methylation by a Mammalian M6Am Methyltransferase. *Cell Res* **2019**, *29* (1), 80–82. https://doi.org/10.1038/s41422-018-0117-4.

(52) Sendinc, E.; Valle-Garcia, D.; Dhall, A.; Chen, H.; Henriques, T.; Navarrete-Perea, J.; Sheng, W.; Gygi, S. P.; Adelman, K.; Shi, Y. PCIF1 Catalyzes M6Am MRNA Methylation to Regulate Gene Expression. *Mol Cell* **2019**, *75* (3), 620-630.e9. https://doi.org/10.1016/j.molcel.2019.05.030.

(53) Wei, J.; Liu, F.; Lu, Z.; Fei, Q.; Ai, Y.; He, P. C.; Shi, H.; Cui, X.; Su, R.; Klungland, A.; Jia, G.; Chen, J.; He, C. Differential M6A, M6Am, and M1A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. *Mol Cell* **2018**, *71* (6), 973-985.e5. https://doi.org/10.1016/j.molcel.2018.08.011.

(54) Akichika, S.; Hirano, S.; Shichino, Y.; Suzuki, T.; Nishimasu, H.; Ishitani, R.; Sugita, A.; Hirose, Y.; Iwasaki, S.; Nureki, O.; Suzuki, T. Cap-Specific Terminal N 6-Methylation of RNA by an RNA Polymerase II-Associated Methyltransferase. *Science* **2019**, *363* (6423), eaav0080. https://doi.org/10.1126/science.aav0080.

(55) Wang, P.; Doxtader, K. A.; Nam, Y. Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. *Mol Cell* **2016**, *63* (2), 306–317. https://doi.org/10.1016/j.molcel.2016.05.041.

(56) Wang, X.; Feng, J.; Xue, Y.; Guan, Z.; Zhang, D.; Liu, Z.; Gong, Z.; Wang, Q.; Huang, J.; Tang, C.; Zou, T.; Yin, P. Structural Basis of N(6)-Adenosine Methylation by the METTL3-METTL14 Complex. *Nature* **2016**, *534* (7608), 575–578. https://doi.org/10.1038/nature18298.

(57) Ping, X.-L.; Sun, B.-F.; Wang, L.; Xiao, W.; Yang, X.; Wang, W.-J.; Adhikari, S.; Shi, Y.; Lv, Y.; Chen, Y.-S.; Zhao, X.; Li, A.; Yang, Y.; Dahal, U.; Lou, X.-M.; Liu, X.; Huang, J.; Yuan, W.-P.; Zhu, X.-F.; Cheng, T.; Zhao, Y.-L.; Wang, X.; Rendtlew Danielsen, J. M.; Liu, F.; Yang, Y.-G. Mammalian WTAP Is a Regulatory Subunit of the RNA N6-Methyladenosine Methyltransferase. *Cell Res* **2014**, *24* (2), 177–189. https://doi.org/10.1038/cr.2014.3.

(58) Liu, J.; Yue, Y.; Han, D.; Wang, X.; Fu, Y.; Zhang, L.; Jia, G.; Yu, M.; Lu, Z.; Deng, X.; Dai, Q.; Chen, W.; He, C. A METTL3-METTL14 Complex Mediates Mammalian Nuclear RNA N6-Adenosine Methylation. *Nat Chem Biol* **2014**, *10* (2), 93–95. https://doi.org/10.1038/nchembio.1432.

(59) Yue, Y.; Liu, J.; Cui, X.; Cao, J.; Luo, G.; Zhang, Z.; Cheng, T.; Gao, M.; Shu, X.; Ma, H.; Wang, F.; Wang, X.; Shen, B.; Wang, Y.; Feng, X.; He, C.; Liu, J. VIRMA Mediates Preferential M6A MRNA Methylation in 3'UTR and near Stop Codon and Associates with Alternative Polyadenylation. *Cell Discov* **2018**, *4*, 10. https://doi.org/10.1038/s41421-018-0019-0.

(60) Wen, J.; Lv, R.; Ma, H.; Shen, H.; He, C.; Wang, J.; Jiao, F.; Liu, H.; Yang, P.; Tan, L.; Lan, F.; Shi, Y. G.; He, C.; Shi, Y.; Diao, J. Zc3h13 Regulates Nuclear RNA M6A Methylation and Mouse Embryonic Stem Cell Self-Renewal. *Mol Cell* **2018**, *69* (6), 1028-1038.e6. https://doi.org/10.1016/j.molcel.2018.02.015.

(61) Patil, D. P.; Chen, C.-K.; Pickering, B. F.; Chow, A.; Jackson, C.; Guttman, M.; Jaffrey, S. R. M(6)A RNA Methylation Promotes XIST-Mediated Transcriptional Repression. *Nature* **2016**, *537* (7620), 369–373. https://doi.org/10.1038/nature19342.

(62) Knuckles, P.; Carl, S. H.; Musheev, M.; Niehrs, C.; Wenger, A.; Bühler, M. RNA Fate Determination through Cotranscriptional Adenosine Methylation and Microprocessor Binding. *Nat Struct Mol Biol* **2017**, *24* (7), 561–569. https://doi.org/10.1038/nsmb.3419.

(63) Xiang, Y.; Laurent, B.; Hsu, C.-H.; Nachtergaele, S.; Lu, Z.; Sheng, W.; Xu, C.; Chen, H.; Ouyang, J.; Wang, S.; Ling, D.; Hsu, P.-H.; Zou, L.; Jambhekar, A.; He, C.; Shi, Y. RNA M6A Methylation Regulates the Ultraviolet-Induced DNA Damage Response. *Nature* **2017**, *543* (7646), 573–576. https://doi.org/10.1038/nature21671.

(64) Choe, J.; Lin, S.; Zhang, W.; Liu, Q.; Wang, L.; Ramirez-Moya, J.; Du, P.; Kim, W.; Tang, S.; Sliz, P.; Santisteban, P.; George, R. E.; Richards, W. G.; Wong, K.-K.; Locker, N.; Slack, F. J.; Gregory, R. I. MRNA Circularization by METTL3-EIF3h Enhances Translation and Promotes Oncogenesis. *Nature* **2018**, *561* (7724), 556–560. https://doi.org/10.1038/s41586-018-0538-8.

(65) Geula, S.; Moshitch-Moshkovitz, S.; Dominissini, D.; Mansour, A. A.; Kol, N.; Salmon-Divon, M.; Hershkovitz, V.; Peer, E.; Mor, N.; Manor, Y. S.; Ben-Haim, M. S.; Eyal, E.; Yunger, S.; Pinto, Y.; Jaitin, D. A.; Viukov, S.; Rais, Y.; Krupalnik, V.; Chomsky, E.; Zerbib, M.; Maza, I.; Rechavi, Y.; Massarwa, R.; Hanna, S.; Amit, I.; Levanon, E. Y.; Amariglio, N.; Stern-Ginossar, N.; Novershtern, N.; Rechavi, G.; Hanna, J. H. Stem Cells. M6A MRNA Methylation Facilitates Resolution of Naïve Pluripotency toward Differentiation. *Science* **2015**, *347* (6225), 1002–1006. https://doi.org/10.1126/science.1261417.

(66) Yoon, K.-J.; Ringeling, F. R.; Vissers, C.; Jacob, F.; Pokrass, M.; Jimenez-Cyrus, D.; Su, Y.; Kim, N.-S.; Zhu, Y.; Zheng, L.; Kim, S.; Wang, X.; Doré, L. C.; Jin, P.; Regot, S.; Zhuang, X.; Canzar, S.; He, C.; Ming, G.-L.; Song, H. Temporal Control of Mammalian Cortical Neurogenesis by M6A Methylation. *Cell* **2017**, *171* (4), 877-889.e17. https://doi.org/10.1016/j.cell.2017.09.003.

(67) Zhang, C.; Chen, Y.; Sun, B.; Wang, L.; Yang, Y.; Ma, D.; Lv, J.; Heng, J.; Ding, Y.; Xue, Y.; Lu, X.; Xiao, W.; Yang, Y.-G.; Liu, F. M6A Modulates Haematopoietic Stem and Progenitor Cell Specification. *Nature* **2017**, *549* (7671), 273–276. https://doi.org/10.1038/nature23883.

(68) Pendleton, K. E.; Chen, B.; Liu, K.; Hunter, O. V.; Xie, Y.; Tu, B. P.; Conrad, N. K. The U6 SnRNA M6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. *Cell* **2017**, *169* (5), 824-835.e14. https://doi.org/10.1016/j.cell.2017.05.003.

(69) H, S.; M, M.; Y, I.; M, E.; A, M.; Y, S.; S, K.; K, O.; T, S.; K, I. S-Adenosylmethionine Synthesis Is Regulated by Selective N6-Adenosine Methylation and MRNA Degradation Involving METTL16 and YTHDC1. *Cell reports* **2017**, *21* (12). https://doi.org/10.1016/j.celrep.2017.11.092.

(70) Doxtader, K. A.; Wang, P.; Scarborough, A. M.; Seo, D.; Conrad, N. K.; Nam, Y. Structural Basis for Regulation of METTL16, an S-Adenosylmethionine Homeostasis Factor. *Mol Cell* **2018**, *71* (6), 1001-1011.e4. https://doi.org/10.1016/j.molcel.2018.07.025.

(71) Mendel, M.; Chen, K.-M.; Homolka, D.; Gos, P.; Pandey, R. R.; McCarthy, A. A.; Pillai, R. S. Methylation of Structured RNA by the M6A Writer METTL16 Is Essential for Mouse Embryonic Development. *Mol Cell* **2018**, *71* (6), 986-1000.e11. https://doi.org/10.1016/j.molcel.2018.08.004.

(72) Wang, T.; Birsoy, K.; Hughes, N. W.; Krupczak, K. M.; Post, Y.; Wei, J. J.; Lander, E. S.; Sabatini, D. M. Identification and Characterization of Essential Genes in the Human Genome. *Science* **2015**, *350* (6264), 1096–1101. https://doi.org/10.1126/science.aac7041.

(73) Warda, A. S.; Kretschmer, J.; Hackert, P.; Lenz, C.; Urlaub, H.; Höbartner, C.; Sloan, K. E.; Bohnsack, M. T. Human METTL16 Is a N6-Methyladenosine (M6A) Methyltransferase That Targets Pre-MRNAs and Various Non-Coding RNAs. *EMBO Rep* **2017**, *18* (11), 2004–2014. https://doi.org/10.15252/embr.201744940.

(74) Sun, H.; Zhang, M.; Li, K.; Bai, D.; Yi, C. Cap-Specific, Terminal N6-Methylation by a Mammalian M6Am Methyltransferase. *Cell Res* **2019**, *29* (1), 80–82. https://doi.org/10.1038/s41422-018-0117-4.

(75) Cowling, V. H. CAPAM: The MRNA Cap Adenosine N6-Methyltransferase. *Trends Biochem Sci* **2019**, *44* (3), 183–185. https://doi.org/10.1016/j.tibs.2019.01.002.

(76) Fan, H.; Sakuraba, K.; Komuro, A.; Kato, S.; Harada, F.; Hirose, Y. PCIF1, a Novel Human WW Domain-Containing Protein, Interacts with the Phosphorylated RNA Polymerase II. *Biochem Biophys Res Commun* **2003**, *301* (2), 378–385. https://doi.org/10.1016/s0006-291x(02)03015-2.

(77) Hirose, Y.; Iwamoto, Y.; Sakuraba, K.; Yunokuchi, I.; Harada, F.; Ohkuma, Y. Human Phosphorylated CTD-Interacting Protein, PCIF1, Negatively Modulates Gene Expression by RNA Polymerase II. *Biochem Biophys Res Commun* **2008**, *369* (2), 449–455. https://doi.org/10.1016/j.bbrc.2008.02.042.

(78) Pandey, R. R.; Delfino, E.; Homolka, D.; Roithova, A.; Chen, K.-M.; Li, L.; Franco, G.; Vågbø, C. B.; Taillebourg, E.; Fauvarque, M.-O.; Pillai, R. S. The Mammalian Cap-Specific M6Am RNA Methyltransferase PCIF1 Regulates Transcript Levels in Mouse Tissues. *Cell Rep* **2020**, *32* (7), 108038. https://doi.org/10.1016/j.celrep.2020.108038.

(79) Levis, R.; Penman, S. 5'-Terminal Structures of Poly(A)+ Cytoplasmic Messenger RNA and of Poly(A)+ and Poly(A)- Heterogeneous Nuclear RNA of Cells of the Dipteran Drosophila Melanogaster. *J Mol Biol* **1978**, *120* (4), 487–515. https://doi.org/10.1016/0022-2836(78)90350-9.

(80) Banerjee, A. K. 5'-Terminal Cap Structure in Eucaryotic Messenger Ribonucleic Acids. *Microbiol Rev* **1980**, *44* (2), 175–205. https://doi.org/10.1128/mr.44.2.175-205.1980.

(81) Du, H.; Zhao, Y.; He, J.; Zhang, Y.; Xi, H.; Liu, M.; Ma, J.; Wu, L. YTHDF2 Destabilizes m(6)A-Containing RNA through Direct Recruitment of the CCR4-NOT Deadenylase Complex. *Nat Commun* **2016**, *7*, 12626. https://doi.org/10.1038/ncomms12626.

(82) Shi, H.; Wang, X.; Lu, Z.; Zhao, B. S.; Ma, H.; Hsu, P. J.; Liu, C.; He, C. YTHDF3 Facilitates Translation and Decay of N6-Methyladenosine-Modified RNA. *Cell Res* **2017**, *27* (3), 315–328. https://doi.org/10.1038/cr.2017.15.

(83) Tirumuru, N.; Zhao, B. S.; Lu, W.; Lu, Z.; He, C.; Wu, L. Correction: N6-Methyladenosine of HIV-1 RNA Regulates Viral Infection and HIV-1 Gag Protein Expression. *eLife 6*, e31482. https://doi.org/10.7554/eLife.31482.

(84) Bai, Y.; Yang, C.; Wu, R.; Huang, L.; Song, S.; Li, W.; Yan, P.; Lin, C.; Li, D.; Zhang, Y. YTHDF1 Regulates Tumorigenicity and Cancer Stem Cell-Like Activity in Human Colorectal Carcinoma. *Front Oncol* **2019**, *9*, 332. https://doi.org/10.3389/fonc.2019.00332.

(85) Han, B.; Yan, S.; Wei, S.; Xiang, J.; Liu, K.; Chen, Z.; Bai, R.; Sheng, J.; Xu, Z.; Gao, X. YTHDF1-Mediated Translation Amplifies Wnt-Driven Intestinal Stemness. *EMBO Rep* **2020**, *21* (4), e49229. https://doi.org/10.15252/embr.201949229.

(86) Orouji, E.; Peitsch, W. K.; Orouji, A.; Houben, R.; Utikal, J. Oncogenic Role of an Epigenetic Reader of M6A RNA Modification: YTHDF1 in Merkel Cell Carcinoma. *Cancers (Basel)* **2020**, *12* (1), E202. https://doi.org/10.3390/cancers12010202.

(87) Mapperley, C.; van de Lagemaat, L. N.; Lawson, H.; Tavosanis, A.; Paris, J.; Campos, J.; Wotherspoon, D.; Durko, J.; Sarapuu, A.; Choe, J.; Ivanova, I.; Krause, D. S.; von Kriegsheim, A.; Much, C.; Morgan, M.; Gregory, R. I.; Mead, A. J.; O'Carroll, D.; Kranc, K. R. The MRNA M6A Reader YTHDF2 Suppresses Proinflammatory Pathways and Sustains Hematopoietic Stem Cell Function. *J Exp Med* **2021**, *218* (3), e20200829. https://doi.org/10.1084/jem.20200829.

(88) Zaccara, S.; Jaffrey, S. R. A Unified Model for the Function of YTHDF Proteins in Regulating M6A-Modified MRNA. *Cell* **2020**, *181* (7), 1582-1595.e18. https://doi.org/10.1016/j.cell.2020.05.012.

(89) Xiao, W.; Adhikari, S.; Dahal, U.; Chen, Y.-S.; Hao, Y.-J.; Sun, B.-F.; Sun, H.-Y.; Li, A.; Ping, X.-L.; Lai, W.-Y.; Wang, X.; Ma, H.-L.; Huang, C.-M.; Yang, Y.; Huang, N.; Jiang, G.-B.; Wang, H.-L.; Zhou, Q.; Wang, X.-J.; Zhao, Y.-L.; Yang, Y.-G. Nuclear m(6)A Reader YTHDC1 Regulates MRNA Splicing. *Mol Cell* **2016**, *61* (4), 507–519. https://doi.org/10.1016/j.molcel.2016.01.012.

(90) Liu, T.; Wei, Q.; Jin, J.; Luo, Q.; Liu, Y.; Yang, Y.; Cheng, C.; Li, L.; Pi, J.; Si, Y.; Xiao, H.; Li, L.; Rao, S.; Wang, F.; Yu, J.; Yu, J.; Zou, D.; Yi, P. The M6A Reader YTHDF1 Promotes Ovarian Cancer Progression via Augmenting EIF3C Translation. *Nucleic Acids Res* **2020**, *48* (7), 3816–3831. https://doi.org/10.1093/nar/gkaa048.

(91) Liu, R.; Kasowitz, S. D.; Homolka, D.; Leu, N. A.; Shaked, J. T.; Ruthel, G.; Jain, D.; Lin, H.; Keeney, S.; Luo, M.; Pillai, R. S.; Wang, P. J. YTHDC2 Is Essential for Pachytene Progression and Prevents Aberrant Microtubule-Driven Telomere Clustering in Male Meiosis. *Cell Rep* **2021**, *37* (11), 110110. https://doi.org/10.1016/j.celrep.2021.110110.

(92) Liu, N.; Dai, Q.; Zheng, G.; He, C.; Parisien, M.; Pan, T. N(6)-Methyladenosine-Dependent RNA Structural Switches Regulate RNA-Protein Interactions. *Nature* **2015**, *518* (7540), 560–564. https://doi.org/10.1038/nature14234.

(93) Wu, B.; Su, S.; Patil, D. P.; Liu, H.; Gan, J.; Jaffrey, S. R.; Ma, J. Molecular Basis for the Specific and Multivariant Recognitions of RNA Substrates by Human HnRNP A2/B1. *Nat Commun* **2018**, *9* (1), 420. https://doi.org/10.1038/s41467-017-02770-z.

(94) Huang, H.; Weng, H.; Sun, W.; Qin, X.; Shi, H.; Wu, H.; Zhao, B. S.; Mesquita, A.; Liu, C.; Yuan, C. L.; Hu, Y.-C.; Hüttelmaier, S.; Skibbe, J. R.; Su, R.; Deng, X.; Dong, L.; Sun, M.; Li, C.; Nachtergaele, S.; Wang, Y.; Hu, C.; Ferchen, K.; Greis, K. D.; Jiang, X.; Wei, M.; Qu, L.; Guan, J.-L.; He, C.; Yang, J.; Chen, J. Recognition of RNA N6-Methyladenosine by IGF2BP Proteins Enhances MRNA Stability and Translation. *Nat Cell Biol* **2018**, *20* (3), 285–295. https://doi.org/10.1038/s41556-018-0045-z.

(95) R, W.; A, L.; B, S.; Jg, S.; J, Z.; T, Z.; Y, C.; Y, X.; Y, G.; Q, Z.; J, M.; X, Y.; Y, L.; Wy, L.; X, Q.; S, W.; Y, S.; Hl, W.; F, W.; Yg, Y.; Z, Y. A Novel M6A Reader Prrc2a Controls Oligodendroglial Specification and Myelination. *Cell research* **2019**, *29* (1). https://doi.org/10.1038/s41422-018-0113-8.

(96) Jia, G.; Fu, Y.; Zhao, X.; Dai, Q.; Zheng, G.; Yang, Y.; Yi, C.; Lindahl, T.; Pan, T.; Yang, Y.-G.; He, C. N6-Methyladenosine in Nuclear RNA Is a Major Substrate of the Obesity-Associated FTO. *Nat Chem Biol* **2011**, *7* (12), 885–887. https://doi.org/10.1038/nchembio.687.

(97) Wei, C.; Gershowitz, A.; Moss, B. N6, O2'-Dimethyladenosine a Novel Methylated Ribonucleoside next to the 5' Terminal of Animal Cell and Virus MRNAs. *Nature* **1975**, *257* (5523), 251–253. https://doi.org/10.1038/257251a0.

(98) Zhang, X.; Wei, L.-H.; Wang, Y.; Xiao, Y.; Liu, J.; Zhang, W.; Yan, N.; Amu, G.; Tang, X.; Zhang, L.; Jia, G. Structural Insights into FTO's Catalytic Mechanism for the Demethylation of Multiple RNA Substrates. *Proc Natl Acad Sci U S A* **2019**, *116* (8), 2919–2924. https://doi.org/10.1073/pnas.1820574116.

(99) Wei, J.; Liu, F.; Lu, Z.; Fei, Q.; Ai, Y.; He, P. C.; Shi, H.; Cui, X.; Su, R.; Klungland, A.; Jia, G.; Chen, J.; He, C. Differential M6A, M6Am, and M1A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. *Mol Cell* **2018**, *71* (6), 973-985.e5. https://doi.org/10.1016/j.molcel.2018.08.011.

(100) Berulava, T.; Ziehe, M.; Klein-Hitpass, L.; Mladenov, E.; Thomale, J.; Rüther, U.; Horsthemke, B. FTO Levels Affect RNA Modification and the Transcriptome. *Eur J Hum Genet* **2013**, *21* (3), 317–323. https://doi.org/10.1038/ejhg.2012.168.

(101) Gulati, P.; Cheung, M. K.; Antrobus, R.; Church, C. D.; Harding, H. P.; Tung, Y.-C. L.; Rimmington, D.; Ma, M.; Ron, D.; Lehner, P. J.; Ashcroft, F. M.; Cox, R. D.; Coll, A. P.; O'Rahilly, S.; Yeo, G. S. H. Role for the Obesity-Related FTO Gene in the Cellular Sensing of Amino Acids. *Proc Natl Acad Sci U S A* **2013**, *110* (7), 2557–2562. https://doi.org/10.1073/pnas.1222796110.

(102) Zhu, T.; Yong, X. L. H.; Xia, D.; Widagdo, J.; Anggono, V. Ubiquitination Regulates the Proteasomal Degradation and Nuclear Translocation of the Fat Mass and Obesity-Associated (FTO) Protein. *J Mol Biol* **2018**, *430* (3), 363–371. https://doi.org/10.1016/j.jmb.2017.12.003.

(103) Zheng, G.; Dahl, J. A.; Niu, Y.; Fedorcsak, P.; Huang, C.-M.; Li, C. J.; Vågbø, C. B.; Shi, Y.; Wang, W.-L.; Song, S.-H.; Lu, Z.; Bosmans, R. P. G.; Dai, Q.; Hao, Y.-J.; Yang, X.; Zhao, W.-M.; Tong, W.-M.; Wang, X.-J.; Bogdan, F.; Furu, K.; Fu, Y.; Jia, G.; Zhao, X.; Liu, J.; Krokan, H. E.; Klungland, A.; Yang, Y.-G.; He, C. ALKBH5 Is a Mammalian RNA Demethylase That Impacts RNA Metabolism and Mouse Fertility. *Mol Cell* **2013**, *49* (1), 18–29. https://doi.org/10.1016/j.molcel.2012.10.015.

(104) Xu, C.; Liu, K.; Tempel, W.; Demetriades, M.; Aik, W.; Schofield, C. J.; Min, J. Structures of Human ALKBH5 Demethylase Reveal a Unique Binding Mode for Specific Single-Stranded N6-Methyladenosine RNA Demethylation. *J Biol Chem* **2014**, *289* (25), 17299–17311. https://doi.org/10.1074/jbc.M114.550350.

(105) Tang, C.; Klukovich, R.; Peng, H.; Wang, Z.; Yu, T.; Zhang, Y.; Zheng, H.; Klungland, A.; Yan, W. ALKBH5-Dependent M6A Demethylation Controls Splicing and Stability of Long 3'-UTR MRNAs in Male Germ Cells. *Proc Natl Acad Sci U S A* **2018**, *115* (2), E325–E333. https://doi.org/10.1073/pnas.1717794115.

(106) Li, Z.; Weng, H.; Su, R.; Weng, X.; Zuo, Z.; Li, C.; Huang, H.; Nachtergaele, S.; Dong, L.; Hu, C.; Qin, X.; Tang, L.; Wang, Y.; Hong, G.-M.; Huang, H.; Wang, X.; Chen, P.; Gurbuxani, S.; Arnovitz, S.; Li, Y.; Li, S.; Strong, J.; Neilly, M. B.; Larson, R. A.; Jiang, X.; Zhang, P.; Jin, J.; He, C.; Chen, J. FTO Plays an Oncogenic Role in Acute Myeloid Leukemia as a N6-Methyladenosine RNA Demethylase. *Cancer Cell* **2017**, *31* (1), 127–141. https://doi.org/10.1016/j.ccell.2016.11.017.

(107) Yamaji, T.; Iwasaki, M.; Sawada, N.; Shimazu, T.; Inoue, M.; Tsugane, S. Fat Mass and Obesity-Associated Gene Polymorphisms, Pre-Diagnostic Plasma Adipokine Levels and the Risk of Colorectal Cancer: The Japan Public Health Center-Based Prospective Study. *PLoS One* **2020**, *15* (2), e0229005. https://doi.org/10.1371/journal.pone.0229005.

(108) Shafik, A. M.; Allen, E. G.; Jin, P. Dynamic N6-Methyladenosine RNA Methylation in Brain and Diseases. *Epigenomics* **2020**, *12* (4), 371–380. https://doi.org/10.2217/epi-2019-0260.

(109) Choudhry, Z.; Sengupta, S. M.; Grizenko, N.; Thakur, G. A.; Fortier, M.-E.; Schmitz, N.; Joober, R. Association between Obesity-Related Gene FTO and ADHD. *Obesity (Silver Spring)* **2013**, *21* (12), E738-744. https://doi.org/10.1002/oby.20444.

(110) Milaneschi, Y.; Lamers, F.; Mbarek, H.; Hottenga, J.-J.; Boomsma, D. I.; Penninx, B. W. J. H. The Effect of FTO Rs9939609 on Major Depression Differs across MDD Subtypes. *Mol Psychiatry* **2014**, *19* (9), 960–962. https://doi.org/10.1038/mp.2014.4.

(111) Keller, L.; Xu, W.; Wang, H.-X.; Winblad, B.; Fratiglioni, L.; Graff, C. The Obesity Related Gene, FTO, Interacts with APOE, and Is Associated with Alzheimer's Disease Risk: A Prospective Cohort Study. *J Alzheimers Dis* **2011**, *23* (3), 461–469. https://doi.org/10.3233/JAD-2010-101068.

(112) Reitz, C.; Tosto, G.; Mayeux, R.; Luchsinger, J. A.; NIA-LOAD/NCRAD Family Study Group; Alzheimer's Disease Neuroimaging Initiative. Genetic Variants in the Fat and Obesity Associated (FTO) Gene and Risk of Alzheimer's Disease. *PLoS One* **2012**, *7* (12), e50354. https://doi.org/10.1371/journal.pone.0050354.

(113) Han, M.; Liu, Z.; Xu, Y.; Liu, X.; Wang, D.; Li, F.; Wang, Y.; Bi, J. Abnormality of M6A MRNA Methylation Is Involved in Alzheimer's Disease. *Front Neurosci* **2020**, *14*, 98. https://doi.org/10.3389/fnins.2020.00098.

(114) Semenza, G. L. The Hypoxic Tumor Microenvironment: A Driving Force for Breast Cancer Progression. *Biochim Biophys Acta* **2016**, *1863* (3), 382–391. https://doi.org/10.1016/j.bbamcr.2015.05.036.

(115) Regan Anderson, T. M.; Peacock, D. L.; Daniel, A. R.; Hubbard, G. K.; Lofgren, K. A.; Girard, B. J.; Schörg, A.; Hoogewijs, D.; Wenger, R. H.; Seagroves, T. N.; Lange, C. A. Breast Tumor Kinase (Brk/PTK6) Is a Mediator of Hypoxia-Associated Breast Cancer Progression. *Cancer Res* **2013**, *73* (18), 5810–5820. https://doi.org/10.1158/0008-5472.CAN-13-0523.

(116) Zhang, C.; Samanta, D.; Lu, H.; Bullen, J. W.; Zhang, H.; Chen, I.; He, X.; Semenza, G. L. Hypoxia Induces the Breast Cancer Stem Cell Phenotype by HIF-Dependent and ALKBH5-Mediated M^6A-Demethylation of NANOG MRNA. *Proc Natl Acad Sci U S A* **2016**, *113* (14), E2047-2056. https://doi.org/10.1073/pnas.1602883113.

(117) S, Z.; Bs, Z.; A, Z.; K, L.; S, Z.; Z, L.; Y, C.; Ep, S.; K, X.; O, B.; S, M.; C, H.; S, H. M6A Demethylase ALKBH5 Maintains Tumorigenicity of Glioblastoma Stem-like Cells by Sustaining FOXM1 Expression and Cell Proliferation Program. *Cancer cell* **2017**, *31* (4). https://doi.org/10.1016/j.ccell.2017.02.013.

(118) Chao, Y.; Shang, J.; Ji, W. ALKBH5-M6A-FOXM1 Signaling Axis Promotes Proliferation and Invasion of Lung Adenocarcinoma Cells under Intermittent Hypoxia. *Biochem Biophys Res Commun* **2020**, *521* (2), 499–506. https://doi.org/10.1016/j.bbrc.2019.10.145.

(119) Carrell, D. T.; Emery, B. R.; Hammoud, S. Altered Protamine Expression and Diminished Spermatogenesis: What Is the Link? *Hum Reprod Update* **2007**, *13* (3), 313–327. https://doi.org/10.1093/humupd/dml057.

(120) Lence, T.; Akhtar, J.; Bayer, M.; Schmid, K.; Spindler, L.; Ho, C. H.; Kreim, N.; Andrade-Navarro, M. A.; Poeck, B.; Helm, M.; Roignant, J.-Y. M6A Modulates Neuronal Functions and Sex Determination in Drosophila. *Nature* **2016**, *540* (7632), 242–247. https://doi.org/10.1038/nature20568.

(121) Bailey, A. S.; Batista, P. J.; Gold, R. S.; Chen, Y. G.; de Rooij, D. G.; Chang, H. Y.; Fuller, M. T. The Conserved RNA Helicase YTHDC2 Regulates the Transition from Proliferation to Differentiation in the Germline. *Elife* **2017**, *6*, e26116. https://doi.org/10.7554/eLife.26116.

(122) Lence, T.; Soller, M.; Roignant, J.-Y. A Fly View on the Roles and Mechanisms of the M6A MRNA Modification and Its Players. *RNA Biol* **2017**, *14* (9), 1232–1240. https://doi.org/10.1080/15476286.2017.1307484.

(123) Kan, L.; Grozhik, A. V.; Vedanayagam, J.; Patil, D. P.; Pang, N.; Lim, K.-S.; Huang, Y.-C.; Joseph, B.; Lin, C.-J.; Despic, V.; Guo, J.; Yan, D.; Kondo, S.; Deng, W.-M.; Dedon, P. C.; Jaffrey, S. R.; Lai, E. C. The M6A Pathway Facilitates Sex Determination in Drosophila. *Nat Commun* **2017**, *8*, 15737. https://doi.org/10.1038/ncomms15737.

(124) Hongay, C. F.; Orr-Weaver, T. L. Drosophila Inducer of MEiosis 4 (IME4) Is Required for Notch Signaling during Oogenesis. *Proc Natl Acad Sci U S A* **2011**, *108* (36), 14855–14860. https://doi.org/10.1073/pnas.1111577108.

(125) Iu, H.; Z, B.; E, S.-M.; Np, M.; N, A.; Rg, F.; M, S. M6A Potentiates Sxl Alternative Pre-MRNA Splicing for Robust Drosophila Sex Determination. *Nature* **2016**, *540* (7632). https://doi.org/10.1038/nature20577.

(126) Guo, J.; Tang, H.-W.; Li, J.; Perrimon, N.; Yan, D. Xio Is a Component of the Drosophila Sex Determination Pathway and RNA N6-Methyladenosine Methyltransferase Complex. *Proc Natl Acad Sci U S A* **2018**, *115* (14), 3674–3679. https://doi.org/10.1073/pnas.1720945115.

(127) Penalva, L. O.; Ruiz, M. F.; Ortega, A.; Granadino, B.; Vicente, L.; Segarra, C.; Valcárcel, J.; Sánchez, L. The Drosophila Fl(2)d Gene, Required for Female-Specific Splicing of Sxl and Tra Pre-MRNAs, Encodes a Novel Nuclear Protein with a HQ-Rich Domain. *Genetics* **2000**, *155* (1), 129–139. https://doi.org/10.1093/genetics/155.1.129.

(128) Haussmann, I. U.; Bodi, Z.; Sanchez-Moran, E.; Mongan, N. P.; Archer, N.; Fray, R. G.; Soller, M. M6A Potentiates Sxl Alternative Pre-MRNA Splicing for Robust Drosophila Sex Determination. *Nature* **2016**, *540* (7632), 301–304. https://doi.org/10.1038/nature20577.

(129) Ortega, A.; Niksic, M.; Bachi, A.; Wilm, M.; Sánchez, L.; Hastie, N.; Valcárcel, J. Biochemical Function of Female-Lethal (2)D/Wilms' Tumor Suppressor-1-Associated Proteins in Alternative Pre-MRNA Splicing. *J Biol Chem* **2003**, *278* (5), 3040–3047. https://doi.org/10.1074/jbc.M210737200.

(130) Anderson, A. M.; Weasner, B. P.; Weasner, B. M.; Kumar, J. P. The Drosophila Wilms′ Tumor 1-Associating Protein (WTAP) Homolog Is Required for Eye Development. *Dev Biol* **2014**, *390* (2), 170–180. https://doi.org/10.1016/j.ydbio.2014.03.012.

(131) Granadino, B.; Campuzano, S.; Sánchez, L. The Drosophila Melanogaster Fl(2)d Gene Is Needed for the Female-Specific Splicing of Sex-Lethal RNA. *EMBO J* **1990**, *9* (8), 2597–2602. https://doi.org/10.1002/j.1460-2075.1990.tb07441.x.

(132) Penn, J. K. M.; Graham, P.; Deshpande, G.; Calhoun, G.; Chaouki, A. S.; Salz, H. K.; Schedl, P. Functioning of the Drosophila Wilms'-Tumor-1-Associated Protein Homolog, Fl(2)d, in Sex-Lethal-Dependent Alternative Splicing. *Genetics* **2008**, *178* (2), 737–748. https://doi.org/10.1534/genetics.107.081679.

(133) Weasner, B.; Salzer, C.; Kumar, J. P. Sine Oculis, a Member of the SIX Family of Transcription Factors, Directs Eye Formation. *Dev Biol* **2007**, *303* (2), 756–771. https://doi.org/10.1016/j.ydbio.2006.10.040.

(134) Halder, G.; Callaerts, P.; Flister, S.; Walldorf, U.; Kloter, U.; Gehring, W. J. Eyeless Initiates the Expression of Both Sine Oculis and Eyes Absent during Drosophila Compound Eye Development. *Development* **1998**, *125* (12), 2181–2191. https://doi.org/10.1242/dev.125.12.2181.

(135) Pauli, T.; Seimiya, M.; Blanco, J.; Gehring, W. J. Identification of Functional Sine Oculis Motifs in the Autoregulatory Element of Its Own Gene, in the Eyeless Enhancer and in the Signalling Gene Hedgehog. *Development* **2005**, *132* (12), 2771–2782. https://doi.org/10.1242/dev.01841.

(136) Pappu, K. S.; Ostrin, E. J.; Middlebrooks, B. W.; Sili, B. T.; Chen, R.; Atkins, M. R.; Gibbs, R.; Mardon, G. Dual Regulation and Redundant Function of Two Eye-Specific Enhancers of the Drosophila Retinal Determination Gene Dachshund. *Development* **2005**, *132* (12), 2895–2905. https://doi.org/10.1242/dev.01869.

(137) Kenyon, K. L.; Li, D. J.; Clouser, C.; Tran, S.; Pignoni, F. Fly SIX-Type Homeodomain Proteins Sine Oculis and Optix Partner with Different Cofactors during Eye Development. *Dev Dyn* **2005**, *234* (3), 497–504. https://doi.org/10.1002/dvdy.20442.

(138) Neilson, K. M.; Pignoni, F.; Yan, B.; Moody, S. A. Developmental Expression Patterns of Candidate Cofactors for Vertebrate Six Family Transcription Factors. *Dev Dyn* **2010**, *239* (12), 3446–3466. https://doi.org/10.1002/dvdy.22484.

(139) Knuckles, P.; Lence, T.; Haussmann, I. U.; Jacob, D.; Kreim, N.; Carl, S. H.; Masiello, I.; Hares, T.; Villaseñor, R.; Hess, D.; Andrade-Navarro, M. A.; Biggiogera, M.; Helm, M.; Soller, M.; Bühler, M.; Roignant, J.-Y. Zc3h13/Flacc Is Required for Adenosine Methylation by Bridging the MRNA-Binding Factor Rbm15/Spenito to the M6A Machinery Component Wtap/Fl(2)d. *Genes Dev* **2018**, *32* (5–6), 415–429. https://doi.org/10.1101/gad.309146.117.

(140) Jemc, J.; Rebay, I. Characterization of the Split Ends-Like Gene Spenito Reveals Functional Antagonism Between SPOC Family Members During Drosophila Eye Development. *Genetics* **2006**, *173* (1), 279–286. https://doi.org/10.1534/genetics.106.055558.

(141) Chang, J. L.; Lin, H. V.; Blauwkamp, T. A.; Cadigan, K. M. Spenito and Split Ends Act Redundantly to Promote Wingless Signaling. *Dev Biol* **2008**, *314* (1), 100–111. https://doi.org/10.1016/j.ydbio.2007.11.023.

(142) Yan, D.; Perrimon, N. Spenito Is Required for Sex Determination in Drosophila Melanogaster. *Proc Natl Acad Sci U S A* **2015**, *112* (37), 11606–11611. https://doi.org/10.1073/pnas.1515891112.

(143) Hilfiker, A.; Amrein, H.; Dübendorfer, A.; Schneiter, R.; Nöthiger, R. The Gene Virilizer Is Required for Female-Specific Splicing Controlled by Sxl, the Master Gene for Sexual Development in Drosophila. *Development* **1995**, *121* (12), 4017–4026. https://doi.org/10.1242/dev.121.12.4017.

(144) Schwartz, S.; Mumbach, M. R.; Jovanovic, M.; Wang, T.; Maciag, K.; Bushkin, G. G.; Mertins, P.; Ter-Ovanesyan, D.; Habib, N.; Cacchiarelli, D.; Sanjana, N. E.; Freinkman, E.; Pacold, M. E.; Satija, R.; Mikkelsen, T. S.; Hacohen, N.; Zhang, F.; Carr, S. A.; Lander, E. S.; Regev, A. Perturbation of M6A Writers Reveals Two Distinct Classes of MRNA Methylation at Internal and 5' Sites. *Cell Rep* **2014**, *8* (1), 284–296. https://doi.org/10.1016/j.celrep.2014.05.048.

(145) Kachaev, Z. M.; Lebedeva, L. A.; Shaposhnikov, A. V.; Moresco, J. J.; Yates, J. R.; Schedl, P.; Shidlovskii, Y. V. Paip2 Cooperates with Cbp80 at an Active Promoter and Participates in RNA Polymerase II Phosphorylation in Drosophila. *FEBS Lett* **2019**, *593* (10), 1102–1112. https://doi.org/10.1002/1873-3468.13391.

(146) Xiao, W.; Adhikari, S.; Dahal, U.; Chen, Y.-S.; Hao, Y.-J.; Sun, B.-F.; Sun, H.-Y.; Li, A.; Ping, X.-L.; Lai, W.-Y.; Wang, X.; Ma, H.-L.; Huang, C.-M.; Yang, Y.; Huang, N.; Jiang, G.-B.; Wang, H.-L.; Zhou, Q.; Wang, X.-J.; Zhao, Y.-L.; Yang, Y.-G. Nuclear m(6)A Reader YTHDC1 Regulates MRNA Splicing. *Mol Cell* **2016**, *61* (4), 507–519. https://doi.org/10.1016/j.molcel.2016.01.012.

(147) Patil, D. P.; Chen, C.-K.; Pickering, B. F.; Chow, A.; Jackson, C.; Guttman, M.; Jaffrey, S. R. M(6)A RNA Methylation Promotes XIST-Mediated Transcriptional Repression. *Nature* **2016**, *537* (7620), 369–373. https://doi.org/10.1038/nature19342.

(148) Xu, C.; Liu, K.; Ahmed, H.; Loppnau, P.; Schapira, M.; Min, J. Structural Basis for the Discriminative Recognition of N6-Methyladenosine RNA by the Human YT521-B Homology Domain Family of Proteins. *J Biol Chem* **2015**, *290* (41), 24902–24913. https://doi.org/10.1074/jbc.M115.680389.

(149) Wojtas, M. N.; Pandey, R. R.; Mendel, M.; Homolka, D.; Sachidanandam, R.; Pillai, R. S. Regulation of M6A Transcripts by the 3'→5' RNA Helicase YTHDC2 Is Essential for a Successful Meiotic Program in the Mammalian Germline. *Mol Cell* **2017**, *68* (2), 374-387.e12. https://doi.org/10.1016/j.molcel.2017.09.021.

(150) Hsu, P. J.; Zhu, Y.; Ma, H.; Guo, Y.; Shi, X.; Liu, Y.; Qi, M.; Lu, Z.; Shi, H.; Wang, J.; Cheng, Y.; Luo, G.; Dai, Q.; Liu, M.; Guo, X.; Sha, J.; Shen, B.; He, C. Ythdc2 Is an N6-Methyladenosine Binding Protein That Regulates Mammalian Spermatogenesis. *Cell Res* **2017**, *27* (9), 1115–1127. https://doi.org/10.1038/cr.2017.99.

(151) Shen, R.; Xie, T. Stem Cell Self-Renewal versus Differentiation: Tumor Suppressor Mei-P26 and MiRNAs Control the Balance. *Cell Res* **2008**, *18* (7), 713–715. https://doi.org/10.1038/cr.2008.79.

(152) Chau, J.; Kulnane, L. S.; Salz, H. K. Sex-Lethal Enables Germline Stem Cell Differentiation by down-Regulating Nanos Protein Levels during Drosophila Oogenesis. *Proc Natl Acad Sci U S A* **2012**, *109* (24), 9465–9470. https://doi.org/10.1073/pnas.1120473109.

(153) Li, Y.; Zhang, Q.; Carreira-Rosario, A.; Maines, J. Z.; McKearin, D. M.; Buszczak, M. Mei-P26 Cooperates with Bam, Bgcn and Sxl to Promote Early Germline Development in the Drosophila Ovary. *PLoS One* **2013**, *8* (3), e58301. https://doi.org/10.1371/journal.pone.0058301.

(154) Chen, D.; Wu, C.; Zhao, S.; Geng, Q.; Gao, Y.; Li, X.; Zhang, Y.; Wang, Z. Three RNA Binding Proteins Form a Complex to Promote Differentiation of Germline Stem Cell Lineage in Drosophila. *PLoS Genet* **2014**, *10* (11), e1004797. https://doi.org/10.1371/journal.pgen.1004797.

(155) Insco, M. L.; Bailey, A. S.; Kim, J.; Olivares, G. H.; Wapinski, O. L.; Tam, C. H.; Fuller, M. T. A Self-Limiting Switch Based on Translational Control Regulates the Transition from Proliferation to Differentiation in an Adult Stem Cell Lineage. *Cell Stem Cell* **2012**, *11* (5), 689–700. https://doi.org/10.1016/j.stem.2012.08.012.

(156) Ohlstein, B.; Lavoie, C. A.; Vef, O.; Gateff, E.; McKearin, D. M. The Drosophila Cystoblast Differentiation Factor, Benign Gonial Cell Neoplasm, Is Related to DExH-Box Proteins and Interacts Genetically with Bag-of-Marbles. *Genetics* **2000**, *155* (4), 1809–1819. https://doi.org/10.1093/genetics/155.4.1809.

(157) Dunckley, T.; Parker, R. The DCP2 Protein Is Required for MRNA Decapping in Saccharomyces Cerevisiae and Contains a Functional MutT Motif. *EMBO J* **1999**, *18* (19), 5411–5422. https://doi.org/10.1093/emboj/18.19.5411.

(158) Zinder, J. C.; Lima, C. D. Targeting RNA for Processing or Destruction by the Eukaryotic RNA Exosome and Its Cofactors. *Genes Dev.* **2017**, *31* (2), 88–100. https://doi.org/10.1101/gad.294769.116.

(159) Mugridge, J. S.; Gross, J. D. Decapping Enzymes STOP "Cancer" Ribosomes in Their Tracks. *The EMBO Journal* **2018**, *37* (23), e100801. https://doi.org/10.15252/embj.2018100801.

(160) Sharma, S.; Grudzien-Nogalska, E.; Hamilton, K.; Jiao, X.; Yang, J.; Tong, L.; Kiledjian, M. Mammalian Nudix Proteins Cleave Nucleotide Metabolite Caps on RNAs. *Nucleic Acids Res* **2020**, *48* (12), 6788–6798. https://doi.org/10.1093/nar/gkaa402.

(161) Knaap, J. A. van der; Verrijzer, C. P. Undercover: Gene Control by Metabolites and Metabolic Enzymes. *Genes Dev.* **2016**, *30* (21), 2345–2369. https://doi.org/10.1101/gad.289140.116.

(162) McLennan, A. G. The Nudix Hydrolase Superfamily. *Cell. Mol. Life Sci.* **2006**, *63* (2), 123–143. https://doi.org/10.1007/s00018-005-5386-7.

(163) Carreras-Puigvert, J.; Zitnik, M.; Jemth, A.-S.; Carter, M.; Unterlass, J. E.; Hallström, B.; Loseva, O.; Karem, Z.; Calderón-Montaño, J. M.; Lindskog, C.; Edqvist, P.-H.; Matuszewski, D. J.; Ait Blal, H.; Berntsson, R. P. A.; Häggblad, M.; Martens, U.; Studham, M.; Lundgren, B.; Wählby, C.; Sonnhammer, E. L. L.; Lundberg, E.; Stenmark, P.; Zupan, B.; Helleday, T. A Comprehensive Structural, Biochemical and Biological Profiling of the Human NUDIX Hydrolase Family. *Nat Commun* **2017**, *8* (1), 1541. https://doi.org/10.1038/s41467-017-01642-w.

(164) Yoshimura, K.; Shigeoka, S. Versatile Physiological Functions of the Nudix Hydrolase Family in Arabidopsis. *Biosci Biotechnol Biochem* **2015**, *79* (3), 354–366. https://doi.org/10.1080/09168451.2014.987207.

(165) Galloway, A.; Cowling, V. H. MRNA Cap Regulation in Mammalian Cell Function and Fate. *Biochim Biophys Acta Gene Regul Mech* **2019**, *1862* (3), 270–279. https://doi.org/10.1016/j.bbagrm.2018.09.011.

(166) Devarkar, S. C.; Wang, C.; Miller, M. T.; Ramanathan, A.; Jiang, F.; Khan, A. G.; Patel, S. S.; Marcotrigiano, J. Structural Basis for M7G Recognition and 2′-O-Methyl Discrimination in Capped RNAs by the Innate Immune Receptor RIG-I. *Proc Natl Acad Sci U S A* **2016**, *113* (3), 596–601. https://doi.org/10.1073/pnas.1515152113.

(167) Furuichi, Y. Discovery of m(7)G-Cap in Eukaryotic MRNAs. *Proc Jpn Acad Ser B Phys Biol Sci* **2015**, *91* (8), 394–409. https://doi.org/10.2183/pjab.91.394.

(168) Fonseca, B. D.; Zakaria, C.; Jia, J.-J.; Graber, T. E.; Svitkin, Y.; Tahmasebi, S.; Healy, D.; Hoang, H.-D.; Jensen, J. M.; Diao, I. T.; Lussier, A.; Dajadian, C.; Padmanabhan, N.; Wang, W.; Matta-Camacho, E.; Hearnden, J.; Smith, E. M.; Tsukumo, Y.; Yanagiya, A.; Morita, M.; Petroulakis, E.; González, J. L.; Hernández, G.; Alain, T.; Damgaard, C. K. La-Related Protein 1 (LARP1) Represses Terminal Oligopyrimidine (TOP) MRNA Translation Downstream of MTOR Complex 1 (MTORC1). *J Biol Chem* **2015**, *290* (26), 15996–16020. https://doi.org/10.1074/jbc.M114.621730.

(169) Cao, Q.; Padmanabhan, K.; Richter, J. D. Pumilio 2 Controls Translation by Competing with EIF4E for 7-Methyl Guanosine Cap Recognition. *RNA* **2010**, *16* (1), 221–227. https://doi.org/10.1261/rna.1884610.

(170) Kim, J. Y.; Lee, Y. C.; Kim, C. Direct Inhibition of Pumilo Activity by Bam and Bgcn in Drosophila Germ Line Stem Cell Differentiation*. *Journal of Biological Chemistry* **2010**, *285* (7), 4741–4746. https://doi.org/10.1074/jbc.M109.002014.

(171) Menon, K. P.; Sanyal, S.; Habara, Y.; Sanchez, R.; Wharton, R. P.; Ramaswami, M.; Zinn, K. The Translational Repressor Pumilio Regulates Presynaptic Morphology and Controls Postsynaptic Accumulation of Translation Factor EIF-4E. *Neuron* **2004**, *44* (4), 663–676. https://doi.org/10.1016/j.neuron.2004.10.028.

(172) Van Etten, J.; Schagat, T. L.; Hrit, J.; Weidmann, C. A.; Brumbaugh, J.; Coon, J. J.; Goldstrohm, A. C. Human Pumilio Proteins Recruit Multiple Deadenylases to Efficiently Repress Messenger RNAs. *J Biol Chem* **2012**, *287* (43), 36370–36383. https://doi.org/10.1074/jbc.M112.373522.

(173) Chuang, T.-W.; Chang, W.-L.; Lee, K.-M.; Tarn, W.-Y. The RNA-Binding Protein Y14 Inhibits MRNA Decapping and Modulates Processing Body Formation. *Mol Biol Cell* **2013**, *24* (1), 1–13. https://doi.org/10.1091/mbc.E12-03-0217.

(174) Mazza, C.; Segref, A.; Mattaj, I. W.; Cusack, S. Large-Scale Induced Fit Recognition of an m(7)GpppG Cap Analogue by the Human Nuclear Cap-Binding Complex. *EMBO J* **2002**, *21* (20), 5548–5557. https://doi.org/10.1093/emboj/cdf538.

(175) Calero, G.; Wilson, K. F.; Ly, T.; Rios-Steiner, J. L.; Clardy, J. C.; Cerione, R. A. Structural Basis of M7GpppG Binding to the Nuclear Cap-Binding Protein Complex. *Nat Struct Biol* **2002**, *9* (12), 912–917. https://doi.org/10.1038/nsb874.

(176) Fortes, P.; Kufel, J.; Fornerod, M.; Polycarpou-Schwarz, M.; Lafontaine, D.; Tollervey, D.; Mattaj, I. W. Genetic and Physical Interactions Involving the Yeast Nuclear Cap-Binding Complex. *Mol Cell Biol* **1999**, *19* (10), 6543–6553. https://doi.org/10.1128/MCB.19.10.6543.

(177) Narita, T.; Yung, T. M. C.; Yamamoto, J.; Tsuboi, Y.; Tanabe, H.; Tanaka, K.; Yamaguchi, Y.; Handa, H. NELF Interacts with CBC and Participates in 3' End Processing of Replication-Dependent Histone MRNAs. *Mol Cell* **2007**, *26* (3), 349–365. https://doi.org/10.1016/j.molcel.2007.04.011.

(178) Swevers, L.; Liu, J.; Smagghe, G. Defense Mechanisms against Viral Infection in Drosophila: RNAi and Non-RNAi. *Viruses* **2018**, *10* (5), E230. https://doi.org/10.3390/v10050230.

(179) Hossain, M. A.; Chung, C.; Pradhan, S. K.; Johnson, T. L. The Yeast Cap Binding Complex Modulates Transcription Factor Recruitment and Establishes Proper Histone H3K36 Trimethylation during Active Transcription. *Mol Cell Biol* **2013**, *33* (4), 785–799. https://doi.org/10.1128/MCB.00947-12.

(180) Nielsen, A. F.; Gloggnitzer, J.; Martinez, J. Ars2 and the Cap-Binding Complex Team up for Silencing. *Cell* **2009**, *138* (2), 224–226. https://doi.org/10.1016/j.cell.2009.07.009.

(181) Izaurralde, E.; Lewis, J.; McGuigan, C.; Jankowska, M.; Darzynkiewicz, E.; Mattaj, I. W. A Nuclear Cap Binding Protein Complex Involved in Pre-MRNA Splicing. *Cell* **1994**, *78* (4), 657–668. https://doi.org/10.1016/0092-8674(94)90530-4.

(182) Izaurralde, E.; Lewis, J.; Gamberi, C.; Jarmolowski, A.; McGuigan, C.; Mattaj, I. W. A Cap-Binding Protein Complex Mediating U SnRNA Export. *Nature* **1995**, *376* (6542), 709–712. https://doi.org/10.1038/376709a0.

(183) Laubinger, S.; Sachsenberg, T.; Zeller, G.; Busch, W.; Lohmann, J. U.; Rätsch, G.; Weigel, D. Dual Roles of the Nuclear Cap-Binding Complex and SERRATE in Pre-MRNA Splicing and MicroRNA Processing in Arabidopsis Thaliana. *Proc Natl Acad Sci U S A* **2008**, *105* (25), 8795–8800. https://doi.org/10.1073/pnas.0802493105.

(184) Lewis, J. D.; Görlich, D.; Mattaj, I. W. A Yeast Cap Binding Protein Complex (YCBC) Acts at an Early Step in Pre-MRNA Splicing. *Nucleic Acids Res* **1996**, *24* (17), 3332–3336. https://doi.org/10.1093/nar/24.17.3332.

(185) Maquat, L. E.; Tarn, W.-Y.; Isken, O. The Pioneer Round of Translation: Features and Functions. *Cell* **2010**, *142* (3), 368–374. https://doi.org/10.1016/j.cell.2010.07.022.

(186) Worch, R.; Niedzwiecka, A.; Stepinski, J.; Mazza, C.; Jankowska-Anyszka, M.; Darzynkiewicz, E.; Cusack, S.; Stolarski, R. Specificity of Recognition of MRNA 5' Cap by Human Nuclear Cap-Binding Complex. *RNA* **2005**, *11* (9), 1355–1363. https://doi.org/10.1261/rna.2850705.

(187) Yi, W.; Gupta, S.; Ricker, E.; Manni, M.; Jessberger, R.; Chinenov, Y.; Molina, H.; Pernis, A. B. The MTORC1-4E-BP-EIF4E Axis Controls de Novo Bcl6 Protein Synthesis in T Cells and Systemic Autoimmunity. *Nat Commun* **2017**, *8* (1), 254. https://doi.org/10.1038/s41467-017-00348-3.

(188) Thoreen, C. C.; Chantranupong, L.; Keys, H. R.; Wang, T.; Gray, N. S.; Sabatini, D. M. A Unifying Model for MTORC1-Mediated Regulation of MRNA Translation. *Nature* **2012**, *485* (7396), 109–113. https://doi.org/10.1038/nature11083.

(189) Roux, P. P.; Topisirovic, I. Signaling Pathways Involved in the Regulation of MRNA Translation. *Mol Cell Biol* **2018**, *38* (12), e00070-18. https://doi.org/10.1128/MCB.00070-18.

(190) Ho, J. J. D.; Lee, S. A Cap for Every Occasion: Alternative EIF4F Complexes. *Trends Biochem Sci* **2016**, *41* (10), 821–823. https://doi.org/10.1016/j.tibs.2016.05.009.

(191) Osborne, M. J.; Volpon, L.; Kornblatt, J. A.; Culjkovic-Kraljacic, B.; Baguet, A.; Borden, K. L. B. EIF4E3 Acts as a Tumor Suppressor by Utilizing an Atypical Mode of Methyl-7-Guanosine Cap Recognition. *Proc Natl Acad Sci U S A* **2013**, *110* (10), 3877–3882. https://doi.org/10.1073/pnas.1216862110.

(192) Rozovsky, N.; Butterworth, A. C.; Moore, M. J. Interactions between EIF4AI and Its Accessory Factors EIF4B and EIF4H. *RNA* **2008**, *14* (10), 2136–2148. https://doi.org/10.1261/rna.1049608.

(193) Bi, X.; Ren, J.; Goss, D. J. Wheat Germ Translation Initiation Factor EIF4B Affects EIF4A and EIFiso4F Helicase Activity by Increasing the ATP Binding Affinity of EIF4A. *Biochemistry* **2000**, *39* (19), 5758–5765. https://doi.org/10.1021/bi992322p.

(194) Dmitriev, S. E.; Terenin, I. M.; Dunaevsky, Y. E.; Merrick, W. C.; Shatsky, I. N. Assembly of 48S Translation Initiation Complexesfrom Purified Components with MRNAs That Have Some Base Pairingwithin Their 5′ UntranslatedRegions. *Molecular and Cellular Biology* **2003**, *23* (24), 8925–8933. https://doi.org/10.1128/MCB.23.24.8925-8933.2003.

(195) Kahvejian, A.; Svitkin, Y. V.; Sukarieh, R.; M'Boutchou, M.-N.; Sonenberg, N. Mammalian Poly(A)-Binding Protein Is a Eukaryotic Translation Initiation Factor, Which Acts via Multiple Mechanisms. *Genes Dev.* **2005**, *19* (1), 104–113. https://doi.org/10.1101/gad.1262905.

(196) Park, E.-H.; Walker, S. E.; Lee, J. M.; Rothenburg, S.; Lorsch, J. R.; Hinnebusch, A. G. Multiple Elements in the EIF4G1 N-Terminus Promote Assembly of EIF4G1•PABP MRNPs in Vivo. *The EMBO Journal* **2011**, *30* (2), 302–316. https://doi.org/10.1038/emboj.2010.312.

(197) Feoktistova, K.; Tuvshintogs, E.; Do, A.; Fraser, C. S. Human EIF4E Promotes MRNA Restructuring by Stimulating EIF4A Helicase Activity. *Proceedings of the National Academy of Sciences* **2013**, *110* (33), 13339–13344. https://doi.org/10.1073/pnas.1303781110.

(198) *The translation factor eIF-4E promotes tumor formation and cooperates with c-Myc in lymphomagenesis | Nature Medicine*. https://www.nature.com/articles/nm1042 (accessed 2022-11-21).

(199) Zhou, S.; Wang, G.-P.; Liu, C.; Zhou, M. Eukaryotic Initiation Factor 4E (EIF4E) and Angiogenesis: Prognostic Markers for Breast Cancer. *BMC Cancer* **2006**, *6* (1), 231. https://doi.org/10.1186/1471-2407-6-231.

(200) Graff, J. R.; Boghaert, E. R.; De Benedetti, A.; Tudor, D. L.; Zimmer, C. C.; Chan, S. K.; Zimmer, S. G. Reduction of Translation Initiation Factor 4E Decreases the Malignancy of Ras -Transformed Cloned Rat Embryo Fibroblasts. *International Journal of Cancer* **1995**, *60* (2), 255–263. https://doi.org/10.1002/ijc.2910600221.

(201) He, F.; Wu, C.; Jacobson, A. Dcp2 C-Terminal Cis-Binding Elements Control Selective Targeting of the Decapping Enzyme by Forming Distinct Decapping Complexes. *Elife* **2022**, *11*, e74410. https://doi.org/10.7554/eLife.74410.

(202) Sheth, U.; Parker, R. Decapping and Decay of Messenger RNA Occur in Cytoplasmic Processing Bodies. *Science* **2003**, *300* (5620), 805–808. https://doi.org/10.1126/science.1082320.

(203) Borja, M. S.; Piotukh, K.; Freund, C.; Gross, J. D. Dcp1 Links Coactivators of MRNA Decapping to Dcp2 by Proline Recognition. *RNA* **2011**, *17* (2), 278–290. https://doi.org/10.1261/rna.2382011.

(204) Lai, T.; Cho, H.; Liu, Z.; Bowler, M. W.; Piao, S.; Parker, R.; Kim, Y. K.; Song, H. Structural Basis of the PNRC2-Mediated Link between Mrna Surveillance and Decapping. *Structure* **2012**, *20* (12), 2025–2037. https://doi.org/10.1016/j.str.2012.09.009.

(205) Mugridge, J. S.; Ziemniak, M.; Jemielity, J.; Gross, J. D. Structural Basis of MRNA-Cap Recognition by Dcp1-Dcp2. *Nat Struct Mol Biol* **2016**, *23* (11), 987–994. https://doi.org/10.1038/nsmb.3301.

(206) Grudzien-Nogalska, E.; Kiledjian, M. New Insights into Decapping Enzymes and Selective MRNA Decay. *Wiley Interdiscip Rev RNA* **2017**, *8* (1). https://doi.org/10.1002/wrna.1379.

(207) Borbolis, F.; Rallis, J.; Kanatouris, G.; Kokla, N.; Karamalegkos, A.; Vasileiou, C.; Vakaloglou, K. M.; Diallinas, G.; Stravopodis, D. J.; Zervas, C. G.; Syntichaki, P. MRNA Decapping Is an Evolutionarily Conserved Modulator of Neuroendocrine Signaling That Controls Development and Ageing. *Elife* **2020**, *9*, e53757. https://doi.org/10.7554/eLife.53757.

(208) Li, R.-M.; Zhang, M.-N.; Tang, Q.-Y.; Song, M.-G. Global Deletion of the RNA Decapping Enzyme Dcp2 Postnatally in Male Mice Results in Infertility. *Biochem Biophys Res Commun* **2020**, *526* (2), 512–518. https://doi.org/10.1016/j.bbrc.2020.03.101.

(209) Ibayashi, M.; Aizawa, R.; Tsukamoto, S. MRNA Decapping Factor Dcp1a Is Essential for Embryonic Growth in Mice. *Biochem Biophys Res Commun* **2021**, *555*, 128–133. https://doi.org/10.1016/j.bbrc.2021.03.117.

(210) Takagi, Y.; Setoyama, D.; Ito, R.; Kamiya, H.; Yamagata, Y.; Sekiguchi, M. Human MTH3 (NUDT18) Protein Hydrolyzes Oxidized Forms of Guanosine and Deoxyguanosine Diphosphates: Comparison with MTH1 and MTH2. *J Biol Chem* **2012**, *287* (25), 21541–21549. https://doi.org/10.1074/jbc.M112.363010.

(211) Kang, D.; Nishida, J.; Iyama, A.; Nakabeppu, Y.; Furuichi, M.; Fujiwara, T.; Sekiguchi, M.; Takeshige, K. Intracellular Localization of 8-Oxo-DGTPase in Human Cells, with Special Reference to the Role of the Enzyme in Mitochondria ∗. *Journal of Biological Chemistry* **1995**, *270* (24), 14659–14665. https://doi.org/10.1074/jbc.270.24.14659.

(212) Yoshimura, D.; Sakumi, K.; Ohno, M.; Sakai, Y.; Furuichi, M.; Iwai, S.; Nakabeppu, Y. An Oxidized Purine Nucleoside Triphosphatase, MTH1, Suppresses Cell Death Caused by Oxidative Stress*. *Journal of Biological Chemistry* **2003**, *278* (39), 37965–37973. https://doi.org/10.1074/jbc.M306201200.

(213) Nakabeppu, Y.; Kajitani, K.; Sakamoto, K.; Yamaguchi, H.; Tsuchimoto, D. MTH1, an Oxidized Purine Nucleoside Triphosphatase, Prevents the Cytotoxicity and Neurotoxicity of Oxidized Purine Nucleotides. *DNA Repair* **2006**, *5* (7), 761–772. https://doi.org/10.1016/j.dnarep.2006.03.003.

(214) Tchou, J.; Kasai, H.; Shibutani, S.; Chung, M. H.; Laval, J.; Grollman, A. P.; Nishimura, S. 8-Oxoguanine (8-Hydroxyguanine) DNA Glycosylase and Its Substrate Specificity. *Proc Natl Acad Sci U S A* **1991**, *88* (11), 4690–4694. https://doi.org/10.1073/pnas.88.11.4690.

(215) Kennedy, C. H.; Cueto, R.; Belinsky, S. A.; Lechner, J. F.; Pryor, W. A. Overexpression of HMTH1 MRNA: A Molecular Marker of Oxidative Stress in Lung Cancer Cells. *FEBS Letters* **1998**, *429* (1), 17–20. https://doi.org/10.1016/S0014-5793(98)00505-5.

(216) Carter, M.; Jemth, A.-S.; Hagenkort, A.; Page, B. D. G.; Gustafsson, R.; Griese, J. J.; Gad, H.; Valerie, N. C. K.; Desroses, M.; Boström, J.; Warpman Berglund, U.; Helleday, T.; Stenmark, P. Crystal Structure, Biochemical and Cellular Activities Demonstrate Separate Functions of MTH1 and MTH2. *Nat Commun* **2015**, *6*, 7871. https://doi.org/10.1038/ncomms8871.

(217) Okamoto, K.; Toyokuni, S.; Uchida, K.; Ogawa, O.; Takenewa, J.; Kakehi, Y.; Kinoshita, H.; Hattori-Nakakuki, Y.; Hiai, H.; Yoshida, O. Formation of 8-Hydroxy-2′-Deoxyguanosine and 4-Hydroxy-2-Nonenal-Modified Proteins in Human Renal-Cell Carcinoma. *International Journal of Cancer* **1994**, *58* (6), 825–829. https://doi.org/10.1002/ijc.2910580613.

(218) Iida, T.; Furuta, A.; Kawashima, M.; Nishida, J.-I.; Nakabeppu, Y.; Iwaki, T. Accumulation of 8-Oxo-2′-Deoxyguanosine and Increased Expression of HMTH1 Protein in Brain Tumors. *Neuro-Oncology* **2001**, *3* (2), 73–81. https://doi.org/10.1215/15228517-3-2-73.

(219) Maki, H.; Sekiguchi, M. MutT Protein Specifically Hydrolyses a Potent Mutagenic Substrate for DNA Synthesis. *Nature* **1992**, *355* (6357), 273–275. https://doi.org/10.1038/355273a0.

(220) Bjorâs, M.; Luna, L.; Johnsen, B.; Hoff, E.; Haug, T.; Rognes, T.; Seeberg, E. Opposite Base-Dependent Reactions of a Human Base Excision Repair Enzyme on DNA Containing 7,8-Dihydro-8-Oxoguanine and Abasic Sites. *EMBO J* **1997**, *16* (20), 6314–6322. https://doi.org/10.1093/emboj/16.20.6314.

(221) Bruner, S. D.; Norman, D. P.; Verdine, G. L. Structural Basis for Recognition and Repair of the Endogenous Mutagen 8-Oxoguanine in DNA. *Nature* **2000**, *403* (6772), 859–866. https://doi.org/10.1038/35002510.

(222) Lin, J.; Hu, Y.; Tian, B.; Hua, Y. Evolution of Double MutT/Nudix Domain-Containing Proteins: Similar Domain Architectures from Independent Gene Duplication-Fusion Events. *Journal of Genetics and Genomics* **2009**, *36* (10), 603–610. https://doi.org/10.1016/S1673-8527(08)60152-6.

(223) Okumura, K.; Nishihara, S.; Inoue, Y. H. Genetic Identification and Characterization of Three Genes That Prevent Accumulation of Oxidative DNA Damage in Drosophila Adult Tissues. *DNA Repair* **2019**, *78*, 7–19. https://doi.org/10.1016/j.dnarep.2019.02.013.

(224) Moriyama, T.; Nishii, R.; Perez-Andreu, V.; Yang, W.; Klussmann, F. A.; Zhao, X.; Lin, T.-N.; Hoshitsuki, K.; Nersting, J.; Kihira, K.; Hofmann, U.; Komada, Y.; Kato, M.; McCorkle, R.; Li, L.; Koh, K.; Najera, C. R.; Kham, S. K.-Y.; Isobe, T.; Chen, Z.; Chiew, E. K.-H.; Bhojwani, D.; Jeffries, C.; Lu, Y.; Schwab, M.; Inaba, H.; Pui, C.-H.; Relling, M. V.; Manabe, A.; Hori, H.; Schmiegelow, K.; Yeoh, A. E. J.; Evans, W. E.; Yang, J. J. NUDT15 Polymorphisms Alter Thiopurine Metabolism and Hematopoietic Toxicity. *Nat Genet* **2016**, *48* (4), 367–373. https://doi.org/10.1038/ng.3508.

(225) Khaeso, K.; Udayachalerm, S.; Komvilaisak, P.; Chainansamit, S.-O.; Suwannaying, K.; Laoaroon, N.; Kuwatjanakul, P.; Nakkam, N.; Sukasem, C.; Puangpetch, A.; Tassaneeyakul, W.; Chaiyakunapruk, N. Meta-Analysis of NUDT15 Genetic Polymorphism on Thiopurine-Induced Myelosuppression in Asian Populations. *Front Pharmacol* **2021**, *12*, 784712. https://doi.org/10.3389/fphar.2021.784712.

(226) Timmer, A.; Patton, P. H.; Chande, N.; McDonald, J. W. D.; MacDonald, J. K. Azathioprine and 6-Mercaptopurine for Maintenance of Remission in Ulcerative Colitis. *Cochrane Database Syst Rev* **2016**, No. 5, CD000478. https://doi.org/10.1002/14651858.CD000478.pub4.

(227) Kakuta, Y.; Kawai, Y.; Okamoto, D.; Takagawa, T.; Ikeya, K.; Sakuraba, H.; Nishida, A.; Nakagawa, S.; Miura, M.; Toyonaga, T.; Onodera, K.; Shinozaki, M.; Ishiguro, Y.; Mizuno, S.; Takahara, M.; Yanai, S.; Hokari, R.; Nakagawa, T.; Araki, H.; Motoya, S.; Naito, T.; Moroi, R.; Shiga, H.; Endo, K.; Kobayashi, T.; Naganuma, M.; Hiraoka, S.; Matsumoto, T.; Nakamura, S.; Nakase, H.; Hisamatsu, T.; Sasaki, M.; Hanai, H.; Andoh, A.; Nagasaki, M.; Kinouchi, Y.; Shimosegawa, T.; Masamune, A.; Suzuki, Y.; MENDEL study group. NUDT15 Codon 139 Is the Best Pharmacogenetic Marker for Predicting Thiopurine-Induced Severe Adverse Events in Japanese Patients with Inflammatory Bowel Disease: A Multicenter Study. *J Gastroenterol* **2018**, *53* (9), 1065–1078. https://doi.org/10.1007/s00535-018-1486-7.

(228) Tanaka, Y.; Saito, Y. Importance of NUDT15 Polymorphisms in Thiopurine Treatments. *J Pers Med* **2021**, *11* (8), 778. https://doi.org/10.3390/jpm11080778.

(229) Hayakawa, H.; Hofer, A.; Thelander, L.; Kitajima, S.; Cai, Y.; Oshiro, S.; Yakushiji, H.; Nakabeppu, Y.; Kuwano, M.; Sekiguchi, M. Metabolic Fate of Oxidized Guanine Ribonucleotides in Mammalian Cells. *Biochemistry* **1999**, *38* (12), 3610–3614. https://doi.org/10.1021/bi9823611.

(230) Randerath, K.; Janeway, C. M.; Stephenson, M. L.; Zamecnik, P. C. Isolation and Characterization of Dinucleoside Tetra- and Tri-Phosphates Formed in the Presence of Lysyl-SRNA Synthetase. *Biochem Biophys Res Commun* **1966**, *24* (1), 98–105. https://doi.org/10.1016/0006-291x(66)90416-5.

(231) Guo, R.-T.; Chong, Y. E.; Guo, M.; Yang, X.-L. Crystal Structures and Biochemical Analyses Suggest a Unique Mechanism and Role for Human Glycyl-TRNA Synthetase in Ap4A Homeostasis. *J Biol Chem* **2009**, *284* (42), 28968–28976. https://doi.org/10.1074/jbc.M109.030692.

(232) Mattay, J. Noncanonical Metabolite RNA Caps: Classification, Quantification, (de)Capping, and Function. *Wiley Interdiscip Rev RNA* **2022**, e1730. https://doi.org/10.1002/wrna.1730.

(233) Despotović, D.; Brandis, A.; Savidor, A.; Levin, Y.; Fumagalli, L.; Tawfik, D. S. Diadenosine Tetraphosphate (Ap4A) – an E. Coli Alarmone or a Damage Metabolite? *The FEBS Journal* **2017**, *284* (14), 2194–2215. https://doi.org/10.1111/febs.14113.

(234) Lee, P. C.; Bochner, B. R.; Ames, B. N. AppppA, Heat-Shock Stress, and Cell Oxidation. *Proc Natl Acad Sci U S A* **1983**, *80* (24), 7496–7500. https://doi.org/10.1073/pnas.80.24.7496.

(235) Baker, J. C.; Jacobson, M. K. Alteration of Adenyl Dinucleotide Metabolism by Environmental Stress. *Proc Natl Acad Sci U S A* **1986**, *83* (8), 2350–2352. https://doi.org/10.1073/pnas.83.8.2350.

(236) Shu, S. L.; Paruchuru, L. B.; Tay, N. Q.; Chua, Y. L.; Foo, A. S. Y.; Yang, C. M.; Liong, K. H.; Koh, E. G. L.; Lee, A.; Nechushtan, H.; Razin, E.; Kemeny, D. M. Ap4A Regulates Directional Mobility and Antigen Presentation in Dendritic Cells. *iScience* **2019**, *16*, 524–534. https://doi.org/10.1016/j.isci.2019.05.045.

(237) Luciano, D. J.; Levenson-Palmer, R.; Belasco, J. G. Stresses That Raise Np4A Levels Induce Protective Nucleoside Tetraphosphate Capping of Bacterial RNA. *Mol Cell* **2019**, *75* (5), 957-966.e8. https://doi.org/10.1016/j.molcel.2019.05.031.

(238) Hudeček, O.; Benoni, R.; Reyes-Gutierrez, P. E.; Culka, M.; Šanderová, H.; Hubálek, M.; Rulíšek, L.; Cvačka, J.; Krásný, L.; Cahová, H. Dinucleoside Polyphosphates Act as 5′-RNA Caps in Bacteria. *Nat Commun* **2020**, *11* (1), 1052. https://doi.org/10.1038/s41467-020-14896-8.

(239) Oka, K.; Suzuki, T.; Onodera, Y.; Miki, Y.; Takagi, K.; Nagasaki, S.; Akahira, J.-I.; Ishida, T.; Watanabe, M.; Hirakawa, H.; Ohuchi, N.; Sasano, H. Nudix-Type Motif 2 in Human Breast Carcinoma: A Potent Prognostic Factor Associated with Cell Proliferation. *Int J Cancer* **2011**, *128* (8), 1770–1782. https://doi.org/10.1002/ijc.25505.

(240) Marriott, A. S.; Vasieva, O.; Fang, Y.; Copeland, N. A.; McLennan, A. G.; Jones, N. J. NUDT2 Disruption Elevates Diadenosine Tetraphosphate (Ap4A) and Down-Regulates Immune Response and Cancer Promotion Genes. *PLoS One* **2016**, *11* (5), e0154674. https://doi.org/10.1371/journal.pone.0154674.

(241) Diaz, F.; Khosa, S.; Niyazov, D.; Lee, H.; Person, R.; Morrow, M. M.; Signer, R.; Dorrani, N.; Zheng, A.; Herzog, M.; Freundlich, R.; Undiagnosed Diseases Network; Birath, J. B.; Cervantes-Manzo, Y.; Martinez-Agosto, J. A.; Palmer, C.; Nelson, S. F.; Fogel, B. L.; Mishra, S. K. Novel NUDT2 Variant Causes Intellectual Disability and Polyneuropathy. *Ann Clin Transl Neurol* **2020**, *7* (11), 2320–2325. https://doi.org/10.1002/acn3.51209.

(242) Yavuz, H.; Bertoli-Avella, A. M.; Alfadhel, M.; Al-Sannaa, N.; Kandaswamy, K. K.; Al-Tuwaijri, W.; Rolfs, A.; Brandau, O.; Bauer, P. A Founder Nonsense Variant in NUDT2 Causes a Recessive

Neurodevelopmental Disorder in Saudi Arab Children. *Clin Genet* **2018**, *94* (3–4), 393–395. https://doi.org/10.1111/cge.13386.

(243) Vallejo, C. G.; Leon, P. Diadenosine 5",5"'P1,P4-Tetraphosphatase in Drosophila Embryos: Developmental Regulation and Characterization. *Int J Biochem* **1989**, *21* (11), 1223–1228. https://doi.org/10.1016/0020-711x(89)90007-4.

(244) Fisher, D. I.; Safrany, S. T.; Strike, P.; McLennan, A. G.; Cartwright, J. L. Nudix Hydrolases That Degrade Dinucleoside and Diphosphoinositol Polyphosphates Also Have 5-Phosphoribosyl 1-Pyrophosphate (PRPP) Pyrophosphatase Activity That Generates the Glycolytic Activator Ribose 1,5-Bisphosphate. *J Biol Chem* **2002**, *277* (49), 47313–47317. https://doi.org/10.1074/jbc.M209795200.

(245) *ModelSEED*. https://modelseed.org/biochem/compounds/cpd08361 (accessed 2022-11-07).

(246) Ye, W.; Ali, N.; Bembenek, M. E.; Shears, S. B.; Lafer, E. M. Inhibition of Clathrin Assembly by High Affinity Binding of Specific Inositol Polyphosphates to the Synapse-Specific Clathrin Assembly Protein AP-3. *J Biol Chem* **1995**, *270* (4), 1564–1568.

(247) Yamaguchi, Y.; Ikenaka, K.; Niinobe, M.; Yamada, H.; Mikoshiba, K. Myelin Proteolipid Protein (PLP), but Not DM-20, Is an Inositol Hexakisphosphate-Binding Protein. *J Biol Chem* **1996**, *271* (44), 27838–27846. https://doi.org/10.1074/jbc.271.44.27838.

(248) Yang, X.; Safrany, S. T.; Shears, S. B. Site-Directed Mutagenesis of Diphosphoinositol Polyphosphate Phosphohydrolase, a Dual Specificity NUDT Enzyme That Attacks Diadenosine Polyphosphates and Diphosphoinositol Polyphosphates. *J Biol Chem* **1999**, *274* (50), 35434–35440. https://doi.org/10.1074/jbc.274.50.35434.

(249) Grudzien-Nogalska, E.; Jiao, X.; Song, M.-G.; Hart, R. P.; Kiledjian, M. Nudt3 Is an MRNA Decapping Enzyme That Modulates Cell Migration. *RNA* **2016**, *22* (5), 773–781. https://doi.org/10.1261/rna.055699.115.

(250) Song, M.-G.; Bail, S.; Kiledjian, M. Multiple Nudix Family Proteins Possess MRNA Decapping Activity. *RNA* **2013**, *19* (3), 390–399. https://doi.org/10.1261/rna.037309.112.

(251) Sahu, S.; Wang, Z.; Jiao, X.; Gu, C.; Jork, N.; Wittwer, C.; Li, X.; Hostachy, S.; Fiedler, D.; Wang, H.; Jessen, H. J.; Kiledjian, M.; Shears, S. B. InsP7 Is a Small-Molecule Regulator of NUDT3-Mediated MRNA Decapping and Processing-Body Dynamics. *Proceedings of the National Academy of Sciences* **2020**, *117* (32), 19245–19253. https://doi.org/10.1073/pnas.1922284117.

(252) HUA, L. V.; HIDAKA, K.; PESESSE, X.; BARNES, L. D.; SHEARS, S. B. Paralogous Murine Nudt10 and Nudt11 Genes Have Differential Expression Patterns but Encode Identical Proteins That Are Physiologically Competent Diphosphoinositol Polyphosphate Phosphohydrolases. *Biochemical Journal* **2003**, *373* (1), 81–89. https://doi.org/10.1042/bj20030142.

(253) Chen, D.; Zhang, R.; Xie, A.; Yuan, J.; Zhang, J.; Huang, Y.; Zhang, H.; Zhang, F. Clinical Correlations and Prognostic Value of Nudix Hydroxylase 10 in Patients with Gastric Cancer. *Bioengineered* **2021**, *12* (2), 9779–9789. https://doi.org/10.1080/21655979.2021.1995104.

(254) Grisanzio, C.; Werner, L.; Takeda, D.; Awoyemi, B. C.; Pomerantz, M. M.; Yamada, H.; Sooriakumaran, P.; Robinson, B. D.; Leung, R.; Schinzel, A. C.; Mills, I.; Ross-Adams, H.; Neal, D. E.; Kido, M.; Yamamoto, T.; Petrozziello, G.; Stack, E. C.; Lis, R.; Kantoff, P. W.; Loda, M.; Sartor, O.; Egawa, S.; Tewari, A. K.; Hahn, W. C.; Freedman, M. L. Genetic and Functional Analyses Implicate the NUDT11, HNF1B, and SLC22A3 Genes in Prostate Cancer Pathogenesis. *Proc Natl Acad Sci U S A* **2012**, *109* (28), 11252–11257. https://doi.org/10.1073/pnas.1200853109.

(255) Hazelett, D. J.; Rhie, S. K.; Gaddis, M.; Yan, C.; Lakeland, D. L.; Coetzee, S. G.; Ellipse/GAME-ON consortium; Practical consortium; Henderson, B. E.; Noushmehr, H.; Cozen, W.; Kote-Jarai, Z.; Eeles, R. A.; Easton, D. F.; Haiman, C. A.; Lu, W.; Farnham, P. J.; Coetzee, G. A. Comprehensive Functional Annotation of 77 Prostate Cancer Risk Loci. *PLoS Genet* **2014**, *10* (1), e1004102. https://doi.org/10.1371/journal.pgen.1004102.

(256) Gasmi, L.; Cartwright, J. L.; McLennan, A. G. Cloning, Expression and Characterization of YSA1H, a Human Adenosine 5'-Diphosphosugar Pyrophosphatase Possessing a MutT Motif. *Biochem J* **1999**, *344 Pt 2*, 331–337.

(257) Wright, R. H. G.; Lioutas, A.; Le Dily, F.; Soronellas, D.; Pohl, A.; Bonet, J.; Nacht, A. S.; Samino, S.; Font-Mateu, J.; Vicent, G. P.; Wierer, M.; Trabado, M. A.; Schelhorn, C.; Carolis, C.; Macias, M. J.; Yanes, O.; Oliva, B.; Beato, M. ADP-Ribose–Derived Nuclear ATP Synthesis by NUDIX5 Is Required for Chromatin Remodeling. *Science* **2016**, *352* (6290), 1221–1225. https://doi.org/10.1126/science.aad9335.

(258) Zhang, H.; Zhang, L.-Q.; Yang, C.-C.; Li, J.; Tian, X.-Y.; Li, D.-N.; Cui, J.; Cai, J.-P. The High Expression of NUDT5 Indicates Poor Prognosis of Breast Cancer by Modulating AKT / Cyclin D Signaling. *PLOS ONE* **2021**, *16* (2), e0245876. https://doi.org/10.1371/journal.pone.0245876.

(259) Page, B. D. G.; Valerie, N. C. K.; Wright, R. H. G.; Wallner, O.; Isaksson, R.; Carter, M.; Rudd, S. G.; Loseva, O.; Jemth, A.-S.; Almlöf, I.; Font-Mateu, J.; Llona-Minguez, S.; Baranczewski, P.; Jeppsson, F.; Homan, E.; Almqvist, H.; Axelsson, H.; Regmi, S.; Gustavsson, A.-L.; Lundbäck, T.; Scobie, M.; Strömberg, K.; Stenmark, P.; Beato, M.; Helleday, T. Targeted NUDT5 Inhibitors Block Hormone Signaling in Breast Cancer Cells. *Nat Commun* **2018**, *9* (1), 250. https://doi.org/10.1038/s41467-017-02293-7.

(260) Ruswanto, R.; Nofianti, T.; Mardianingrum, R.; Kesuma, D.; Siswandono, null. Design, Molecular Docking, and Molecular Dynamics of Thiourea-Iron (III) Metal Complexes as NUDT5 Inhibitors for Breast Cancer Treatment. *Heliyon* **2022**, *8* (9), e10694. https://doi.org/10.1016/j.heliyon.2022.e10694.

(261) Basha, S. B.; Charles, I. D.; Raju, N.; Manokaran, S.; Kuzhandaivel, H. An Efficient 2-Aminothiazolesalicylaldehyde Fluorescent Chemosensor for Fe2+ Ion Detection and a Potential Inhibitor of NUDT5 Signaling Hormone for Breast Cancer Cell and Molecular Keypad Lock Application. *Chem Zvesti* **2022**, *76* (11), 7061–7073. https://doi.org/10.1007/s11696-022-02373-z.

(262) Shumar, S. A.; Kerr, E. W.; Geldenhuys, W. J.; Montgomery, G. E.; Fagone, P.; Thirawatananond, P.; Saavedra, H.; Gabelli, S. B.; Leonardi, R. Nudt19 Is a Renal CoA Diphosphohydrolase with Biochemical and Regulatory Properties That Are Distinct from the Hepatic Nudt7 Isoform. *J Biol Chem* **2018**, *293* (11), 4134–4148. https://doi.org/10.1074/jbc.RA117.001358.

(263) Sabari, B. R.; Zhang, D.; Allis, C. D.; Zhao, Y. Metabolic Regulation of Gene Expression through Histone Acylations. *Nat Rev Mol Cell Biol* **2017**, *18* (2), 90–101. https://doi.org/10.1038/nrm.2016.140.

(264) Song, J.; Baek, I.-J.; Park, S.; Oh, J.; Kim, D.; Song, K.; Kim, M. K.; Lee, H. W.; Jang, B. K.; Jin, E.-J. Deficiency of Peroxisomal NUDT7 Stimulates de Novo Lipogenesis in Hepatocytes. *iScience* **2022**, *25* (10), 105135. https://doi.org/10.1016/j.isci.2022.105135.

(265) Younossi, Z. M.; Koenig, A. B.; Abdelatif, D.; Fazel, Y.; Henry, L.; Wymer, M. Global Epidemiology of Nonalcoholic Fatty Liver Disease—Meta-Analytic Assessment of Prevalence, Incidence, and Outcomes. *Hepatology* **2016**, *64* (1), 73–84. https://doi.org/10.1002/hep.28431.

(266) Kowtoniuk, W. E.; Shen, Y.; Heemstra, J. M.; Agarwal, I.; Liu, D. R. A Chemical Screen for Biological Small Molecule-RNA Conjugates Reveals CoA-Linked RNA. *Proc Natl Acad Sci U S A* **2009**, *106* (19), 7768–7773. https://doi.org/10.1073/pnas.0900528106.

(267) Yu, D.; Song, W.; Tan, E. Y. J.; Liu, L.; Cao, Y.; Jirschitzka, J.; Li, E.; Logemann, E.; Xu, C.; Huang, S.; Jia, A.; Chang, X.; Han, Z.; Wu, B.; Schulze-Lefert, P.; Chai, J. TIR Domains of Plant Immune Receptors Are 2',3'-CAMP/CGMP Synthetases Mediating Cell Death. *Cell* **2022**, *185* (13), 2370-2386.e18. https://doi.org/10.1016/j.cell.2022.04.032.

(268) Dong, S.; Wang, Y. Nudix Effectors: A Common Weapon in the Arsenal of Plant Pathogens. *PLoS Pathog* **2016**, *12* (8), e1005704. https://doi.org/10.1371/journal.ppat.1005704.

(269) Kerr, E. W.; Shumar, S. A.; Leonardi, R. Nudt8 Is a Novel CoA Diphosphohydrolase That Resides in the Mitochondria. *FEBS Letters* **2019**, *593* (11), 1133–1143. https://doi.org/10.1002/1873-3468.13392.

(270) Ofman, R.; Speijer, D.; Leen, R.; Wanders, R. J. A. Proteomic Analysis of Mouse Kidney Peroxisomes: Identification of RP2p as a Peroxisomal Nudix Hydrolase with Acyl-CoA Diphosphatase Activity. *Biochem J* **2006**, *393* (Pt 2), 537–543. https://doi.org/10.1042/BJ20050893.

(271) Görigk, S.; Ouwens, D. M.; Kuhn, T.; Altenhofen, D.; Binsch, C.; Damen, M.; Khuong, J. M.-A.; Kaiser, K.; Knebel, B.; Vogel, H.; Schürmann, A.; Chadt, A.; Al-Hasani, H. Nudix Hydrolase NUDT19 Regulates Mitochondrial Function and ATP Production in Murine Hepatocytes. *Biochim Biophys Acta Mol Cell Biol Lipids* **2022**, *1867* (6), 159153. https://doi.org/10.1016/j.bbalip.2022.159153.

(272) Pettersen, I. K. N.; Tusubira, D.; Ashrafi, H.; Dyrstad, S. E.; Hansen, L.; Liu, X.-Z.; Nilsson, L. I. H.; Løvsletten, N. G.; Berge, K.; Wergedahl, H.; Bjørndal, B.; Fluge, Ø.; Bruland, O.; Rustan, A. C.; Halberg, N.; Røsland, G. V.; Berge, R. K.; Tronstad, K. J. Upregulated PDK4 Expression Is a Sensitive Marker of Increased Fatty Acid Oxidation. *Mitochondrion* **2019**, *49*, 97–110. https://doi.org/10.1016/j.mito.2019.07.009.

(273) Maruta, T.; Ogawa, T.; Tsujimura, M.; Ikemoto, K.; Yoshida, T.; Takahashi, H.; Yoshimura, K.; Shigeoka, S. Loss-of-Function of an Arabidopsis NADPH Pyrophosphohydrolase, AtNUDX19, Impacts on the Pyridine Nucleotides Status and Confers Photooxidative Stress Tolerance. *Sci Rep* **2016**, *6*, 37432. https://doi.org/10.1038/srep37432.

(274) Abdelraheim, S. R.; Spiller, D. G.; McLennan, A. G. Mammalian NADH Diphosphatases of the Nudix Family: Cloning and Characterization of the Human Peroxisomal NUDT12 Protein. *Biochem J* **2003**, *374* (Pt 2), 329–335. https://doi.org/10.1042/BJ20030441.

(275) Wu, H.; Li, L.; Chen, K.-M.; Homolka, D.; Gos, P.; Fleury-Olela, F.; McCarthy, A. A.; Pillai, R. S. Decapping Enzyme NUDT12 Partners with BLMH for Cytoplasmic Surveillance of NAD-Capped RNAs. *Cell Rep* **2019**, *29* (13), 4422-4434.e13. https://doi.org/10.1016/j.celrep.2019.11.108.

(276) Zapata-Pérez, R.; Wanders, R. J. A.; van Karnebeek, C. D. M.; Houtkooper, R. H. NAD+ Homeostasis in Human Health and Disease. *EMBO Molecular Medicine* **2021**, *13* (7), e13943. https://doi.org/10.15252/emmm.202113943.

(277) Schöndorf, D. C.; Ivanyuk, D.; Baden, P.; Sanchez-Martinez, A.; De Cicco, S.; Yu, C.; Giunta, I.; Schwarz, L. K.; Di Napoli, G.; Panagiotakopoulou, V.; Nestel, S.; Keatinge, M.; Pruszak, J.; Bandmann, O.; Heimrich, B.; Gasser, T.; Whitworth, A. J.; Deleidi, M. The NAD+ Precursor Nicotinamide Riboside Rescues Mitochondrial Defects and Neuronal Loss in IPSC and Fly Models of Parkinson's Disease. *Cell Rep* **2018**, *23* (10), 2976–2988. https://doi.org/10.1016/j.celrep.2018.05.009.

(278) Liu, D.; Pitta, M.; Jiang, H.; Lee, J.-H.; Zhang, G.; Chen, X.; Kawamoto, E. M.; Mattson, M. P. Nicotinamide Forestalls Pathology and Cognitive Decline in Alzheimer Mice: Evidence for Improved Neuronal Bioenergetics and Autophagy Procession. *Neurobiol Aging* **2013**, *34* (6), 1564–1580. https://doi.org/10.1016/j.neurobiolaging.2012.11.020.

(279) Cantó, C.; Houtkooper, R. H.; Pirinen, E.; Youn, D. Y.; Oosterveer, M. H.; Cen, Y.; Fernandez-Marcos, P. J.; Yamamoto, H.; Andreux, P. A.; Cettour-Rose, P.; Gademann, K.; Rinsch, C.; Schoonjans, K.; Sauve, A. A.; Auwerx, J. The NAD(+) Precursor Nicotinamide Riboside Enhances Oxidative Metabolism and Protects against High-Fat Diet-Induced Obesity. *Cell Metab* **2012**, *15* (6), 838–847. https://doi.org/10.1016/j.cmet.2012.04.022.

(280) Lee, H. J.; Hong, Y.-S.; Jun, W.; Yang, S. J. Nicotinamide Riboside Ameliorates Hepatic Metaflammation by Modulating NLRP3 Inflammasome in a Rodent Model of Type 2 Diabetes. *J Med Food* **2015**, *18* (11), 1207–1213. https://doi.org/10.1089/jmf.2015.3439.

(281) Chen, Y. G.; Kowtoniuk, W. E.; Agarwal, I.; Shen, Y.; Liu, D. R. LC/MS Analysis of Cellular RNA Reveals NAD-Linked RNA. *Nat Chem Biol* **2009**, *5* (12), 879–881. https://doi.org/10.1038/nchembio.235.

(282) Walters, R. W.; Matheny, T.; Mizoue, L. S.; Rao, B. S.; Muhlrad, D.; Parker, R. Identification of NAD+ Capped MRNAs in Saccharomyces Cerevisiae. *Proc Natl Acad Sci U S A* **2017**, *114* (3), 480–485. https://doi.org/10.1073/pnas.1619369114.

(283) Jiao, X.; Doamekpor, S. K.; Bird, J. G.; Nickels, B. E.; Tong, L.; Hart, R. P.; Kiledjian, M. 5' End Nicotinamide Adenine Dinucleotide Cap in Human Cells Promotes RNA Decay through DXO-Mediated DeNADding. *Cell* **2017**, *168* (6), 1015-1027.e10. https://doi.org/10.1016/j.cell.2017.02.019.

(284) Wang, Y.; Li, S.; Zhao, Y.; You, C.; Le, B.; Gong, Z.; Mo, B.; Xia, Y.; Chen, X. NAD+-Capped RNAs Are Widespread in the Arabidopsis Transcriptome and Can Probably Be Translated. *Proc Natl Acad Sci U S A* **2019**, *116* (24), 12094–12102. https://doi.org/10.1073/pnas.1903682116.

(285) Vvedenskaya, I. O.; Bird, J. G.; Zhang, Y.; Zhang, Y.; Jiao, X.; Barvík, I.; Krásný, L.; Kiledjian, M.; Taylor, D. M.; Ebright, R. H.; Nickels, B. E. CapZyme-Seq Comprehensively Defines Promoter-Sequence Determinants for RNA 5' Capping with NAD<sup/>. *Mol Cell* **2018**, *70* (3), 553-564.e9. https://doi.org/10.1016/j.molcel.2018.03.014.

(286) Zhang, Y.; Kuster, D.; Schmidt, T.; Kirrmaier, D.; Nübel, G.; Ibberson, D.; Benes, V.; Hombauer, H.; Knop, M.; Jäschke, A. Extensive 5′-Surveillance Guards against Non-Canonical NAD-Caps of Nuclear MRNAs in Yeast. *Nat Commun* **2020**, *11* (1), 5508. https://doi.org/10.1038/s41467-020-19326-3.

(287) Cahová, H.; Winz, M.-L.; Höfer, K.; Nübel, G.; Jäschke, A. NAD CaptureSeq Indicates NAD as a Bacterial Cap for a Subset of Regulatory RNAs. *Nature* **2015**, *519* (7543), 374–377. https://doi.org/10.1038/nature14020.

(288) Grudzien-Nogalska, E.; Wu, Y.; Jiao, X.; Cui, H.; Mateyak, M. K.; Hart, R. P.; Tong, L.; Kiledjian, M. Structural and Mechanistic Basis of Mammalian Nudt12 RNA DeNADding. *Nat Chem Biol* **2019**, *15* (6), 575–582. https://doi.org/10.1038/s41589-019-0293-7.

(289) Song, M.-G.; Li, Y.; Kiledjian, M. Multiple MRNA Decapping Enzymes in Mammalian Cells. *Molecular Cell* **2010**, *40* (3), 423–432. https://doi.org/10.1016/j.molcel.2010.10.010.

(290) Ghosh, T.; Peterson, B.; Tomasevic, N.; Peculis, B. A. Xenopus U8 SnoRNA Binding Protein Is a Conserved Nuclear Decapping Enzyme. *Molecular Cell* **2004**, *13* (6), 817–828. https://doi.org/10.1016/S1097-2765(04)00127-3.

(291) Lu, G.; Zhang, J.; Li, Y.; Li, Z.; Zhang, N.; Xu, X.; Wang, T.; Guan, Z.; Gao, G. F.; Yan, J. HNUDT16: A Universal Decapping Enzyme for Small Nucleolar RNA and Cytoplasmic MRNA. *Protein Cell* **2011**, *2* (1), 64–73. https://doi.org/10.1007/s13238-011-1009-2.

(292) Iyama, T.; Abolhassani, N.; Tsuchimoto, D.; Nonaka, M.; Nakabeppu, Y. NUDT16 Is a (Deoxy)Inosine Diphosphatase, and Its Deficiency Induces Accumulation of Single-Strand Breaks in Nuclear DNA and Growth Arrest. *Nucleic Acids Res* **2010**, *38* (14), 4834–4843. https://doi.org/10.1093/nar/gkq249.

(293) Wang, J.; Alvin Chew, B. L.; Lai, Y.; Dong, H.; Xu, L.; Balamkundu, S.; Cai, W. M.; Cui, L.; Liu, C. F.; Fu, X.-Y.; Lin, Z.; Shi, P.-Y.; Lu, T. K.; Luo, D.; Jaffrey, S. R.; Dedon, P. C. Quantifying the RNA Cap Epitranscriptome Reveals Novel Caps in Cellular and Viral RNA. *Nucleic Acids Research* **2019**, *47* (20), e130. https://doi.org/10.1093/nar/gkz751.

(294) Doamekpor, S. K.; Grudzien-Nogalska, E.; Mlynarska-Cieslak, A.; Kowalska, J.; Kiledjian, M.; Tong, L. DXO/Rai1 Enzymes Remove 5′-End FAD and Dephospho-CoA Caps on RNAs. *Nucleic Acids Research* **2020**, *48* (11), 6136–6148. https://doi.org/10.1093/nar/gkaa297.

(295) Auclair, O.; Han, Y.; Burgos, S. A. Consumption of Milk and Alternatives and Their Contribution to Nutrient Intakes among Canadian Adults: Evidence from the 2015 Canadian Community Health Survey-Nutrition. *Nutrients* **2019**, *11* (8), E1948. https://doi.org/10.3390/nu11081948.

(296) Augustin, P.; Toplak, M.; Fuchs, K.; Gerstmann, E. C.; Prassl, R.; Winkler, A.; Macheroux, P. Oxidation of the FAD Cofactor to the 8-Formyl-Derivative in Human Electron-Transferring Flavoprotein. *J Biol Chem* **2018**, *293* (8), 2829–2840. https://doi.org/10.1074/jbc.RA117.000846.

(297) Terai, Y.; Sato, R.; Matsumura, R.; Iwai, S.; Yamamoto, J. Enhanced DNA Repair by DNA Photolyase Bearing an Artificial Light-Harvesting Chromophore. *Nucleic Acids Res* **2020**, *48* (18), 10076–10086. https://doi.org/10.1093/nar/gkaa719.

(298) Mansoorabadi, S. O.; Thibodeaux, C. J.; Liu, H. The Diverse Roles of Flavin Coenzymes--Nature's Most Versatile Thespians. *J Org Chem* **2007**, *72* (17), 6329–6342. https://doi.org/10.1021/jo0703092.

(299) Bronfman, M.; Inestrosa, N. C.; Leighton, F. Fatty Acid Oxidation by Human Liver Peroxisomes. *Biochemical and Biophysical Research Communications* **1979**, *88* (3), 1030–1036. https://doi.org/10.1016/0006-291X(79)91512-2.

(300) Pinto, J. T.; Cooper, A. J. L. From Cholesterogenesis to Steroidogenesis: Role of Riboflavin and Flavoenzymes in the Biosynthesis of Vitamin D. *Adv Nutr* **2014**, *5* (2), 144–163. https://doi.org/10.3945/an.113.005181.

(301) Wang, J.; de Montellano, P. R. O. The Binding Sites on Human Heme Oxygenase-1 for Cytochrome P450 Reductase and Biliverdin Reductase*. *Journal of Biological Chemistry* **2003**, *278* (22), 20069–20076. https://doi.org/10.1074/jbc.M300989200.

(302) *Correcting a marginal riboflavin deficiency improves hematologic status in young women in the United Kingdom (RIBOFEM) | The American Journal of Clinical Nutrition | Oxford Academic*. https://academic.oup.com/ajcn/article/93/6/1274/4597770 (accessed 2022-11-18).

(303) Mikkelsen, K.; Stojanovska, L.; Tangalakis, K.; Bosevski, M.; Apostolopoulos, V. Cognitive Decline: A Vitamin B Perspective. *Maturitas* **2016**, *93*, 108–113. https://doi.org/10.1016/j.maturitas.2016.08.001.

(304) Asa, S. L.; Ramyar, L.; Murphy, P. R.; Li, A. W.; Ezzat, S. The Endogenous Fibroblast Growth Factor-2 Antisense Gene Product Regulates Pituitary Cell Growth and Hormone Production. *Mol Endocrinol* **2001**, *15* (4), 589–599. https://doi.org/10.1210/mend.15.4.0626.

(305) Li, A. W.; Too, C. K.; Knee, R.; Wilkinson, M.; Murphy, P. R. FGF-2 Antisense RNA Encodes a Nuclear Protein with MutT-like Antimutator Activity. *Mol Cell Endocrinol* **1997**, *133* (2), 177–182. https://doi.org/10.1016/s0303-7207(97)00148-2.

(306) Perraud, A.-L.; Shen, B.; Dunn, C. A.; Rippe, K.; Smith, M. K.; Bessman, M. J.; Stoddard, B. L.; Scharenberg, A. M. NUDT9, a Member of the Nudix Hydrolase Family, Is an Evolutionarily Conserved Mitochondrial ADP-Ribose Pyrophosphatase. *J Biol Chem* **2003**, *278* (3), 1794–1801. https://doi.org/10.1074/jbc.M205601200.

(307) Dölle, C.; Rack, J. G. M.; Ziegler, M. NAD and ADP-Ribose Metabolism in Mitochondria. *FEBS J* **2013**, *280* (15), 3530–3541. https://doi.org/10.1111/febs.12304.

(308) Liu, Q.; Florea, B. I.; Filippov, D. V. ADP-Ribosylation Goes Normal: Serine as the Major Site of the Modification. *Cell Chem Biol* **2017**, *24* (4), 431–432. https://doi.org/10.1016/j.chembiol.2017.04.003.

(309) Bouchard, V. J.; Rouleau, M.; Poirier, G. G. PARP-1, a Determinant of Cell Survival in Response to DNA Damage. *Experimental Hematology* **2003**, *31* (6), 446–454. https://doi.org/10.1016/S0301-472X(03)00083-3.

(310) Scovassi, A. I.; Denegri, M.; Donzelli, M.; Rossi, L.; Bernardi, R.; Mandarino, A.; Frouin, I.; Negri, C. Poly(ADP-Ribose) Synthesis in Cells Undergoing Apoptosis: An Attempt to Face Death before PARP Degradation. *Eur J Histochem* **1998**, *42* (4), 251–258.

(311) Aredia, F.; Scovassi, A. I. Involvement of PARPs in Cell Death. *Front Biosci (Elite Ed)* **2014**, *6* (2), 308–317. https://doi.org/10.2741/707.

(312) Ryu, K. W.; Kim, D.-S.; Kraus, W. L. New Facets in the Regulation of Gene Expression by ADP-Ribosylation and Poly(ADP-Ribose) Polymerases. *Chem. Rev.* **2015**, *115* (6), 2453–2481. https://doi.org/10.1021/cr5004248.

(313) Abdelraheim, S. R.; Spiller, D. G.; McLennan, A. G. Mouse Nudt13 Is a Mitochondrial Nudix Hydrolase with NAD(P)H Pyrophosphohydrolase Activity. *Protein J* **2017**, *36* (5), 425–432. https://doi.org/10.1007/s10930-017-9734-x.

(314) Kanda, M.; Shimizu, D.; Tanaka, H.; Shibata, M.; Iwata, N.; Hayashi, M.; Kobayashi, D.; Tanaka, C.; Yamada, S.; Fujii, T.; Nakayama, G.; Sugimoto, H.; Koike, M.; Fujiwara, M.; Kodera, Y. Metastatic Pathway-Specific Transcriptome Analysis Identifies MFSD4 as a Putative Tumor Suppressor and Biomarker for Hepatic Metastasis in Patients with Gastric Cancer. *Oncotarget* **2016**, *7* (12), 13667–13679. https://doi.org/10.18632/oncotarget.7269.

(315) Heyen, C. A.; Tagliabracci, V. S.; Zhai, L.; Roach, P. J. Characterization of Mouse UDP-Glucose Pyrophosphatase, a Nudix Hydrolase Encoded by the Nudt14 Gene. *Biochemical and Biophysical Research Communications* **2009**, *390* (4), 1414–1418. https://doi.org/10.1016/j.bbrc.2009.11.007.

(316) Muñoz, F. J.; Baroja-Fernández, E.; Morán-Zorzano, M. T.; Alonso-Casajús, N.; Pozueta-Romero, J. Cloning, Expression and Characterization of a Nudix Hydrolase That Catalyzes the Hydrolytic Breakdown of ADP-Glucose Linked to Starch Biosynthesis in Arabidopsis Thaliana. *Plant Cell Physiol* **2006**, *47* (7), 926–934. https://doi.org/10.1093/pcp/pcj065.

(317) Moreno-Bruna, B.; Baroja-Fernández, E.; Muñoz, F. J.; Bastarrica-Berasategui, A.; Zandueta-Criado, A.; Rodriguez-López, M.; Lasa, I.; Akazawa, T.; Pozueta-Romero, J. Adenosine Diphosphate Sugar Pyrophosphatase Prevents Glycogen Biosynthesis in Escherichia Coli. *Proc Natl Acad Sci U S A* **2001**, *98* (14), 8128–8132. https://doi.org/10.1073/pnas.131214098.

(318) Carter, M.; Jemth, A.-S.; Carreras-Puigvert, J.; Herr, P.; Martínez Carranza, M.; Vallin, K. S. A.; Throup, A.; Helleday, T.; Stenmark, P. Human NUDT22 Is a UDP-Glucose/Galactose Hydrolase Exhibiting a Unique Structural Fold. *Structure* **2018**, *26* (2), 295-303.e6. https://doi.org/10.1016/j.str.2018.01.004.

(319) Tanaka, H.; Maruta, T.; Ogawa, T.; Tanabe, N.; Tamoi, M.; Yoshimura, K.; Shigeoka, S. Identification and Characterization of Arabidopsis AtNUDX9 as a GDP-d-Mannose Pyrophosphohydrolase: Its Involvement in Root Growth Inhibition in Response to Ammonium. *J Exp Bot* **2015**, *66* (19), 5797–5808. https://doi.org/10.1093/jxb/erv281.

(320) Natalizio, B. J.; Wente, S. R. Postage for the Messenger: Designating Routes for Nuclear MRNA Export. *Trends Cell Biol* **2013**, *23* (8), 365–373. https://doi.org/10.1016/j.tcb.2013.03.006.

(321) *Post-transcriptional processing*. https://www.mun.ca/biology/scarr/iGen3_05-11.html (accessed 2022-11-02).

(322) Tudek, A.; Lloret-Llinares, M.; Jensen, T. H. The Multitasking PolyA Tail: Nuclear RNA Maturation, Degradation and Export. *Philos Trans R Soc Lond B Biol Sci* **2018**, *373* (1762), 20180169. https://doi.org/10.1098/rstb.2018.0169.

(323) Kühn, U.; Wahle, E. Structure and Function of Poly(A) Binding Proteins. *Biochim Biophys Acta* **2004**, *1678* (2–3), 67–84. https://doi.org/10.1016/j.bbaexp.2004.03.008.

(324) Nicholson, A. L.; Pasquinelli, A. E. Tales of Detailed Poly(A) Tails. *Trends Cell Biol* **2019**, *29* (3), 191–200. https://doi.org/10.1016/j.tcb.2018.11.002.

(325) Rüegsegger, U.; Blank, D.; Keller, W. Human Pre-MRNA Cleavage Factor Im Is Related to Spliceosomal SR Proteins and Can Be Reconstituted in Vitro from Recombinant Subunits. *Mol Cell* **1998**, *1* (2), 243–253. https://doi.org/10.1016/s1097-2765(00)80025-8.

(326) Brown, K. M.; Gilmartin, G. M. A Mechanism for the Regulation of Pre-MRNA 3' Processing by Human Cleavage Factor Im. *Mol Cell* **2003**, *12* (6), 1467–1476. https://doi.org/10.1016/s1097-2765(03)00453-2.

(327) MacDonald, C. C.; Wilusz, J.; Shenk, T. The 64-Kilodalton Subunit of the CstF Polyadenylation Factor Binds to Pre- MRNAs Downstream of the Cleavage Site and Influences Cleavage Site Location. *Molecular and Cellular Biology* **1994**, *14* (10), 6647–6654. https://doi.org/10.1128/MCB.14.10.6647.

(328) Bienroth, S.; Keller, W.; Wahle, E. Assembly of a Processive Messenger RNA Polyadenylation Complex. *EMBO J* **1993**, *12* (2), 585–594. https://doi.org/10.1002/j.1460-2075.1993.tb05690.x.

(329) Dettwiler, S.; Aringhieri, C.; Cardinale, S.; Keller, W.; Barabino, S. M. L. Distinct Sequence Motifs within the 68-KDa Subunit of Cleavage Factor Im Mediate RNA Binding, Protein-Protein Interactions, and Subcellular Localization*. *Journal of Biological Chemistry* **2004**, *279* (34), 35788–35797. https://doi.org/10.1074/jbc.M403927200.

(330) Ruepp, M.-D.; Aringhieri, C.; Vivarelli, S.; Cardinale, S.; Paro, S.; Schümperli, D.; Barabino, S. M. L. Mammalian Pre-MRNA 3′ End Processing Factor CF Im68 Functions in MRNA Export. *MBoC* **2009**, *20* (24), 5211–5223. https://doi.org/10.1091/mbc.e09-05-0389.

(331) Iseli, C.; Stevenson, B. J.; de Souza, S. J.; Samaia, H. B.; Camargo, A. A.; Buetow, K. H.; Strausberg, R. L.; Simpson, A. J. G.; Bucher, P.; Jongeneel, C. V. Long-Range Heterogeneity at the 3' Ends of Human MRNAs. *Genome Res* **2002**, *12* (7), 1068–1074. https://doi.org/10.1101/gr.62002.

(332) Zhu, Y.; Wang, X.; Forouzmand, E.; Jeong, J.; Qiao, F.; Sowd, G. A.; Engelman, A. N.; Xie, X.; Hertel, K. J.; Shi, Y. Molecular Mechanisms for CFIm-Mediated Regulation of MRNA Alternative Polyadenylation. *Mol Cell* **2018**, *69* (1), 62-74.e4. https://doi.org/10.1016/j.molcel.2017.11.031.

(333) Phillips, C.; Jung, S.; Gunderson, S. I. Regulation of Nuclear Poly(A) Addition Controls the Expression of Immunoglobulin M Secretory MRNA. *EMBO J* **2001**, *20* (22), 6443–6452. https://doi.org/10.1093/emboj/20.22.6443.

(334) Yang, Q.; Coseno, M.; Gilmartin, G. M.; Doublié, S. Crystal Structure of a Human Cleavage Factor CFI(m)25/CFI(m)68/RNA Complex Provides an Insight into Poly(A) Site Recognition and RNA Looping. *Structure* **2011**, *19* (3), 368–377. https://doi.org/10.1016/j.str.2010.12.021.

(335) Kubo, T.; Wada, T.; Yamaguchi, Y.; Shimizu, A.; Handa, H. Knock-down of 25 KDa Subunit of Cleavage Factor Im in Hela Cells Alters Alternative Polyadenylation within 3′-UTRs. *Nucleic Acids Research* **2006**, *34* (21), 6264–6271. https://doi.org/10.1093/nar/gkl794.

(336) Vallejos Baier, R.; Picao-Osorio, J.; Alonso, C. R. Molecular Regulation of Alternative Polyadenylation (APA) within the Drosophila Nervous System. *J Mol Biol* **2017**, *429* (21), 3290–3300. https://doi.org/10.1016/j.jmb.2017.03.028.

(337) Masamha, C. P.; Xia, Z.; Yang, J.; Albrecht, T. R.; Li, M.; Shyu, A.-B.; Li, W.; Wagner, E. J. CFIm25 Links Alternative Polyadenylation to Glioblastoma Tumour Suppression. *Nature* **2014**, *510* (7505), 412–416. https://doi.org/10.1038/nature13261.

(338) Sugita, A.; Kuruma, S.; Yanagisawa, N.; Ishiguro, H.; Kano, R.; Ohkuma, Y.; Hirose, Y. The Cap-Specific M6A Methyltransferase, PCIF1/CAPAM, Is Dynamically Recruited to the Gene Promoter in a Transcription-Dependent Manner. *The Journal of Biochemistry* **2021**, *170* (2), 203–213. https://doi.org/10.1093/jb/mvab032.

(339) Wang, Z.-H.; Liu, Y.; Chaitankar, V.; Pirooznia, M.; Xu, H. Electron Transport Chain Biogenesis Activated by a JNK-Insulin-Myc Relay Primes Mitochondrial Inheritance in Drosophila. *Elife* **2019**, *8*, e49309. https://doi.org/10.7554/eLife.49309.

(340) Reuter, G.; Wolff, I. Isolation of Dominant Suppressor Mutations for Position-Effect Variegation in Drosophila Melanogaster. *Mol Gen Genet* **1981**, *182* (3), 516–519. https://doi.org/10.1007/BF00293947.

(341) Wang, S. H.; Nan, R.; Accardo, M. C.; Sentmanat, M.; Dimitri, P.; Elgin, S. C. R. A Distinct Type of Heterochromatin at the Telomeric Region of the Drosophila Melanogaster Y Chromosome. *PLoS One* **2014**, *9* (1), e86451. https://doi.org/10.1371/journal.pone.0086451.

(342) Mteirek, R.; Gueguen, N.; Jensen, S.; Brasset, E.; Vaury, C. Drosophila Heterochromatin: Structure and Function. *Curr Opin Insect Sci* **2014**, *1*, 19–24. https://doi.org/10.1016/j.cois.2014.04.003.

(343) Nichols, C. D.; Becnel, J.; Pandey, U. B. Methods to Assay Drosophila Behavior. *J Vis Exp* **2012**, No. 61, 3795. https://doi.org/10.3791/3795.

(344) Kunar, R.; Roy, J. K. The MRNA Decapping Protein 2 (DCP2) Is a Major Regulator of Developmental Events in Drosophila-Insights from Expression Paradigms. *Cell Tissue Res* **2021**, *386* (2), 261–280. https://doi.org/10.1007/s00441-021-03503-x.

(345) Lee, Y.-M.; Chiang, P.-H.; Cheng, J.-H.; Shen, W.-H.; Chen, C.-H.; Wu, M.-L.; Tian, Y.-L.; Ni, C.-H.; Wang, T.-F.; Lin, M.-D.; Chou, T.-B. Drosophila Decapping Protein 2 Modulates the Formation of Cortical F-Actin for Germ Plasm Assembly. *Dev Biol* **2020**, *461* (1), 96–106. https://doi.org/10.1016/j.ydbio.2020.01.013.

(346) Shumar, S. A.; Kerr, E. W.; Geldenhuys, W. J.; Montgomery, G. E.; Fagone, P.; Thirawatananond, P.; Saavedra, H.; Gabelli, S. B.; Leonardi, R. Nudt19 Is a Renal CoA Diphosphohydrolase with Biochemical and Regulatory Properties That Are Distinct from the Hepatic Nudt7 Isoform. *J Biol Chem* **2018**, *293* (11), 4134–4148. https://doi.org/10.1074/jbc.RA117.001358.

(347) Brown, J. A.; Kinzig, C. G.; DeGregorio, S. J.; Steitz, J. A. Methyltransferase-like Protein 16 Binds the 3'-Terminal Triple Helix of MALAT1 Long Noncoding RNA. *Proc Natl Acad Sci U S A* **2016**, *113* (49), 14013–14018. https://doi.org/10.1073/pnas.1614759113.

(348) Su, R.; Dong, L.; Li, Y.; Gao, M.; He, P. C.; Liu, W.; Wei, J.; Zhao, Z.; Gao, L.; Han, L.; Deng, X.; Li, C.; Prince, E.; Tan, B.; Qing, Y.; Qin, X.; Shen, C.; Xue, M.; Zhou, K.; Chen, Z.; Xue, J.; Li, W.; Qin, H.; Wu, X.; Sun, M.; Nam, Y.; Chen, C.-W.; Huang, W.; Horne, D.; Rosen, S. T.; He, C.; Chen, J. METTL16 Exerts an M6A-Independent Function to Facilitate Translation and Tumorigenesis. *Nat Cell Biol* **2022**, *24* (2), 205–216. https://doi.org/10.1038/s41556-021-00835-2.

(349) Xu, Y.-X.; Hirose, Y.; Zhou, X. Z.; Lu, K. P.; Manley, J. L. Pin1 Modulates the Structure and Function of Human RNA Polymerase II. *Genes Dev* **2003**, *17* (22), 2765–2776. https://doi.org/10.1101/gad.1135503.

(350) Yunokuchi, I.; Fan, H.; Iwamoto, Y.; Araki, C.; Yuda, M.; Umemura, H.; Harada, F.; Ohkuma, Y.; Hirose, Y. Prolyl Isomerase Pin1 Shares Functional Similarity with Phosphorylated CTD Interacting Factor PCIF1 in Vertebrate Cells. *Genes to Cells* **2009**, *14* (9), 1105–1118. https://doi.org/10.1111/j.1365-2443.2009.01339.x.

(351) Churchman, L. S.; Weissman, J. S. Nascent Transcript Sequencing Visualizes Transcription at Nucleotide Resolution. *Nature* **2011**, *469* (7330), 368–373. https://doi.org/10.1038/nature09652.

(352) Lopes, R.; Agami, R.; Korkmaz, G. GRO-Seq, A Tool for Identification of Transcripts Regulating Gene Expression. *Methods Mol Biol* **2017**, *1543*, 45–55. https://doi.org/10.1007/978-1-4939-6716-2_3.

(353) Liu, A.; Oliver-Krasinski, J.; Stoffers, D. A. Two Conserved Domains in PCIF1 Mediate Interaction with Pancreatic Transcription Factor PDX-1. *FEBS Letters* **2006**, *580* (28), 6701–6706. https://doi.org/10.1016/j.febslet.2006.11.021.

(354) Claiborn, K. C.; Sachdeva, M. M.; Cannon, C. E.; Groff, D. N.; Singer, J. D.; Stoffers, D. A. Pcif1 Modulates Pdx1 Protein Stability and Pancreatic β Cell Function and Survival in Mice. *J Clin Invest* **2010**, *120* (10), 3713–3721. https://doi.org/10.1172/JCI40440.

(355) Wang, L.; Zhou, Y.; Chen, D.; Lee, T. H. Peptidyl-Prolyl Cis/Trans Isomerase Pin1 and Alzheimer's Disease. *Front Cell Dev Biol* **2020**, *8*, 355. https://doi.org/10.3389/fcell.2020.00355.

(356) Mazina, M. Y.; Ziganshin, R. H.; Magnitov, M. D.; Golovnin, A. K.; Vorobyeva, N. E. Proximity-Dependent Biotin Labelling Reveals CP190 as an EcR/Usp Molecular Partner. *Sci Rep* **2020**, *10* (1), 4793. https://doi.org/10.1038/s41598-020-61514-0.

(357) Santin, Y. G.; Doan, T.; Lebrun, R.; Espinosa, L.; Journet, L.; Cascales, E. In Vivo TssA Proximity Labelling during Type VI Secretion Biogenesis Reveals TagA as a Protein That Stops and Holds the Sheath. *Nat Microbiol* **2018**, *3* (11), 1304–1313. https://doi.org/10.1038/s41564-018-0234-3.

(358) Opitz, N.; Schmitt, K.; Hofer-Pretz, V.; Neumann, B.; Krebber, H.; Braus, G. H.; Valerius, O. Capturing the Asc1p/Receptor for Activated C Kinase 1 (RACK1) Microenvironment at the Head Region of the 40S Ribosome with Quantitative BioID in Yeast. *Mol Cell Proteomics* **2017**, *16* (12), 2199–2218. https://doi.org/10.1074/mcp.M116.066654.

(359) Uezu, A.; Kanak, D. J.; Bradshaw, T. W. A.; Soderblom, E. J.; Catavero, C. M.; Burette, A. C.; Weinberg, R. J.; Soderling, S. H. Identification of an Elaborate Complex Mediating Postsynaptic Inhibition. *Science* **2016**, *353* (6304), 1123–1129. https://doi.org/10.1126/science.aag0821.

(360) Velentzas, P. D.; Velentzas, A. D.; Pantazi, A. D.; Mpakou, V. E.; Zervas, C. G.; Papassideri, I. S.; Stravopodis, D. J. Proteasome, but Not Autophagy, Disruption Results in Severe Eye and Wing Dysmorphia: A Subunit- and Regulator-Dependent Process in Drosophila. *PLOS ONE* **2013**, *8* (11), e80530. https://doi.org/10.1371/journal.pone.0080530.

(361) Rousakis, A.; Vlanti, A.; Borbolis, F.; Roumelioti, F.; Kapetanou, M.; Syntichaki, P. Diverse Functions of MRNA Metabolism Factors in Stress Defense and Aging of Caenorhabditis Elegans. *PLOS ONE* **2014**, *9* (7), e103365. https://doi.org/10.1371/journal.pone.0103365.

(362) Tang, N. H.; Kim, K. W.; Xu, S.; Blazie, S. M.; Yee, B. A.; Yeo, G. W.; Jin, Y.; Chisholm, A. D. The MRNA Decay Factor CAR-1/LSM14 Regulates Axon Regeneration via Mitochondrial Calcium Dynamics. *Curr Biol* **2020**, *30* (5), 865-876.e7. https://doi.org/10.1016/j.cub.2019.12.061.

(363) Jäschke, A.; Höfer, K.; Nübel, G.; Frindert, J. Cap-like Structures in Bacterial RNA and Epitranscriptomic Modification. *Curr Opin Microbiol* **2016**, *30*, 44–49. https://doi.org/10.1016/j.mib.2015.12.009.

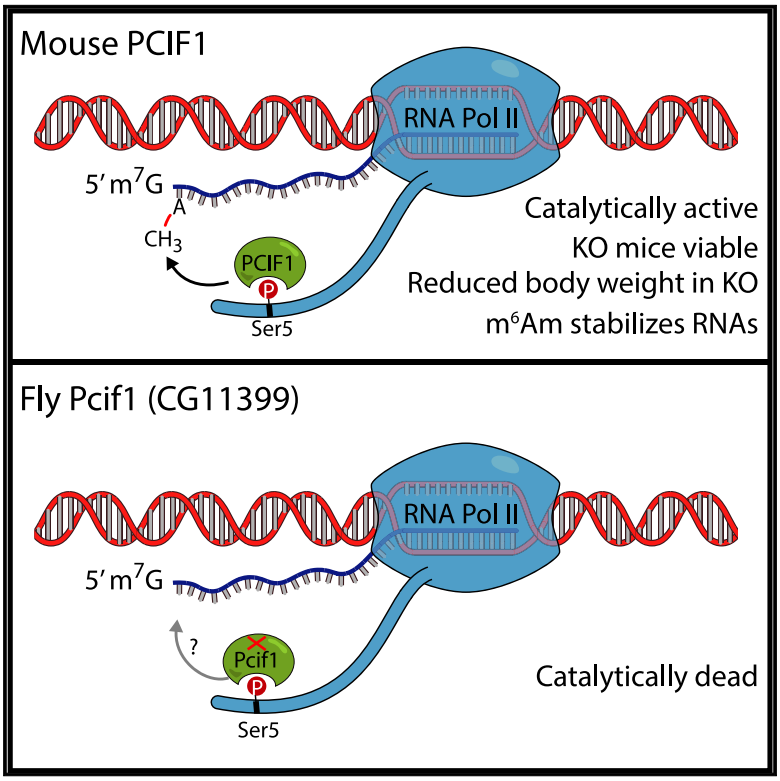
ANNEX

The Mammalian Cap-Specific m^6Am RNA Methyltransferase PCIF1 Regulates Transcript Levels in Mouse Tissues (Pandey et al., 2020 Cell Reports)

Cell Reports

The Mammalian Cap-Specific m^{6}Am RNA Methyltransferase PCIF1 Regulates Transcript Levels in Mouse Tissues

Graphical Abstract



Highlights

- Mouse mutants of the m^{6}Am methylase *Pcif1* display reduced body weight
- Transcripts with a TSS adenosine are destabilized in the *Pcif1* mutant mice
- Catalytically dead *Drosophila* Pcif1 binds Ser5-phospho CTD
- *Trypanosoma* Pcif1 is an m^{6}Am methylase that creates the cap4 structure

Authors

Radha Raman Pandey, Elena Delfino, David Homolka, ..., Emmanuel Taillebourg, Marie-Odile Fauvarque, Ramesh S. Pillai

Correspondence

raman.pandey@unige.ch (R.R.P.), ramesh.pillai@unige.ch (R.S.P.)

In Brief

Pandey et al. demonstrate that a loss of the cap-specific m^{6}Am RNA methylation in mice destabilizes transcripts and that such *Pcif1* mutant mice have reduced body weight. In contrast, *Drosophila* Pcif1 is inactive as a methylase but, like its mammalian counterpart, still binds to Ser5-phosphorylated RNA Pol II CTD.

Pandey et al., 2020, Cell Reports *32*, 108038
August 18, 2020
https://doi.org/10.1016/j.celrep.2020.108038

Cell Reports

CellPress
OPEN ACCESS

Report

The Mammalian Cap-Specific m^{6}Am RNA Methyltransferase PCIF1 Regulates Transcript Levels in Mouse Tissues

Radha Raman Pandey,[1,4,*] Elena Delfino,[1,4] David Homolka,[1,4] Adriana Roithova,[1] Kuan-Ming Chen,[1] Lingyun Li,[1] Giulia Franco,[2] Cathrine Broberg Vågbø,[3] Emmanuel Taillebourg,[2] Marie-Odile Fauvarque,[2] and Ramesh S. Pillai[1,5,*]

[1]Department of Molecular Biology, Science III, University of Geneva, 30 Quai Ernest-Ansermet, 1211 Geneva, Switzerland
[2]University Grenoble Alpes, CEA, INSERM, BGE, 38000 Grenoble, France
[3]Proteomics and Modomics Experimental Core (PROMEC), Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology (NTNU) and St. Olavs Hospital Central Staff, Trondheim, Norway
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: raman.pandey@unige.ch (R.R.P.), ramesh.pillai@unige.ch (R.S.P.)
https://doi.org/10.1016/j.celrep.2020.108038

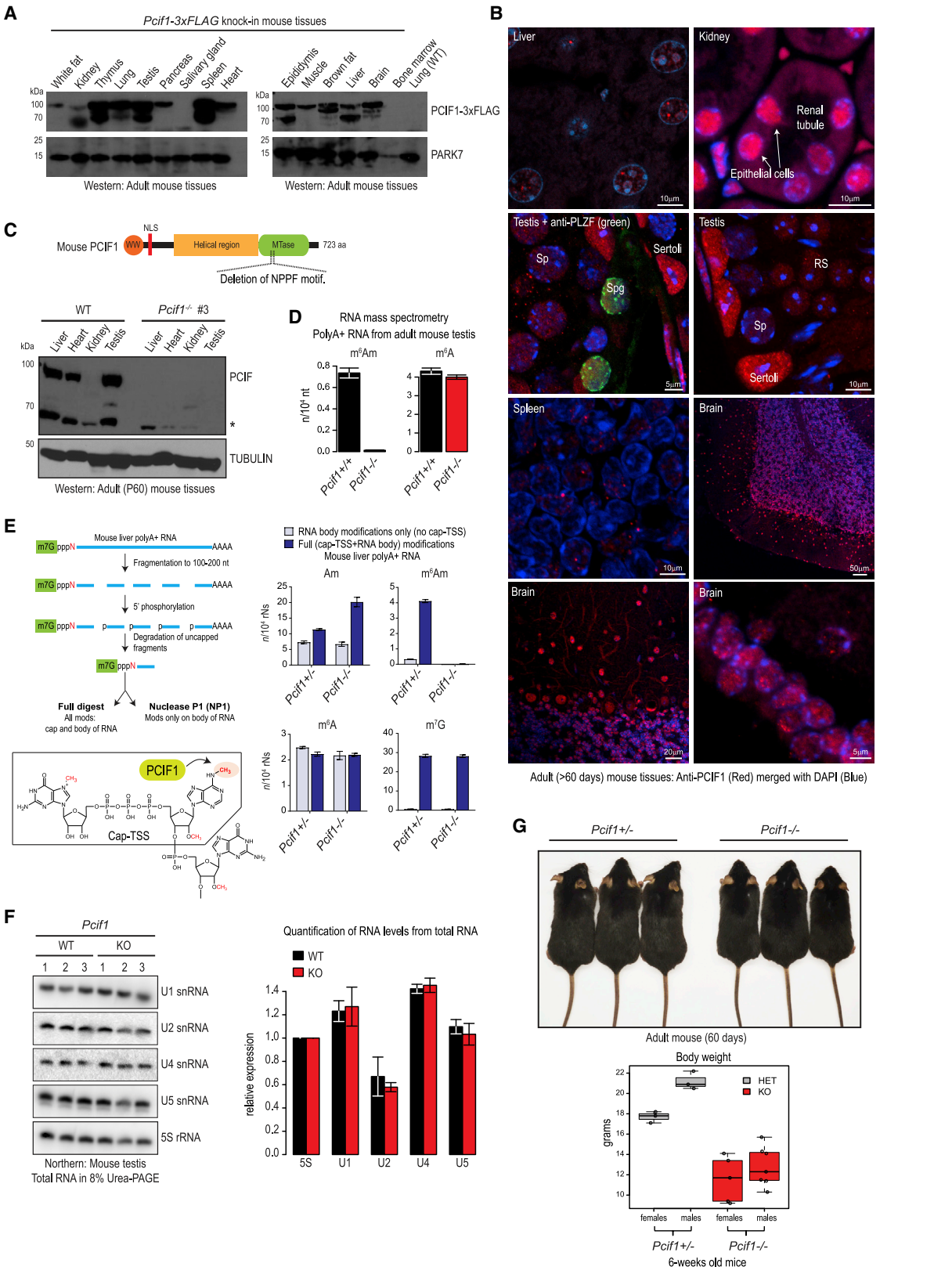
SUMMARY

The 5′ end of eukaryotic mRNAs is protected by the m^{7}G-cap structure. The transcription start site nucleotide is ribose methylated (Nm) in many eukaryotes, whereas an adenosine at this position is further methylated at the N^6 position (m^{6}A) by the mammalian Phosphorylated C-terminal domain (CTD)-interacting Factor 1 (PCIF1) to generate m^{6}Am. Here, we show that although the loss of cap-specific m^{6}Am in mice does not affect viability or fertility, the *Pcif1* mutants display reduced body weight. Transcriptome analyses of mutant mouse tissues support a role for the cap-specific m^{6}Am modification in stabilizing transcripts. In contrast, the *Drosophila* Pcif1 is catalytically dead, but like its mammalian counterpart, it retains the ability to associate with the Ser5-phosphorylated CTD of RNA polymerase II (RNA Pol II). Finally, we show that the *Trypanosoma* Pcif1 is an m^{6}Am methylase that contributes to the N^6,N^6,2′-*O*-trimethyladenosine (m$^6{}_2$Am) in the hypermethylated cap4 structure of trypanosomatids. Thus, PCIF1 has evolved to function in catalytic and non-catalytic roles.

INTRODUCTION

The 5′ end of eukaryotic RNA polymerase II (RNA Pol II) transcripts is protected with a N^7-methylguanosine (m^{7}G) cap structure (Furuichi et al., 1975a; Shatkin, 1976). This cap0 structure is added to the nascent RNA (~20–25 nucleotide [nt]) by an inverted 5′-5′ triphosphate linkage to the 5′ terminal transcription start site (TSS) nucleotide (Inesta-Vaquera and Cowling, 2017; Mao et al., 1995; Muthukrishnan et al., 1975; Shafer et al., 2005; Shatkin and Manley, 2000). The ribose of the TSS nucleotide in most eukaryotes is 2′-*O*-methylated (Am, Um, Cm, or Gm) by the action of CMTR1 to become the cap1 structure (Bélanger et al., 2010; Byszewska et al., 2014; Haline-Vaz et al., 2008; Wei et al., 1975b). CMTR1 uses its C-terminal WW domains to interact with the RNA Pol II C-terminal domain (CTD) (Haline-Vaz et al., 2008), enabling it to access the nascent RNAs. Some of the RNA Pol II transcripts also have a ribose methylation on the second transcribed nucleotide catalyzed by CMTR2, creating the cap2 structure (Furuichi et al., 1975b; Werner et al., 2011). It was recognized early on that when the TSS nucleotide is an adenosine, it can be further methylated at the N^6 position to create the dimethylated adenosine referred to as m^{6}Am (Wei et al., 1975a).

Apart from these cap-specific modifications, RNA Pol II transcripts contain abundant internal N^6-methyladenosine (m^{6}A) marks (Perry et al., 1975; Schibler et al., 1977; Wei et al., 1975b) that are catalyzed by the RNA methyltransferase heterodimer composed of METTL3-METTL14 (Liu et al., 2014; Śledź and Jinek, 2016; Wang et al., 2016). Mapping experiments with anti-m^{6}A antibodies reveal its presence at thousands of sites across the transcriptome, with multiple marks distributed along the length of the transcript body (Dominissini et al., 2012; Meyer et al., 2012; Schwartz et al., 2013). Additionally, the METTL16 writer catalyzes the internal m^{6}A mark on specific structured RNA contexts within a select set of RNAs that include the U6 small nuclear RNA (snRNA) and the S-Adenosyl methionine (SAM) synthetase mRNA (Doxtader et al., 2018; Mendel et al., 2018; Pendleton et al., 2017; Warda et al., 2017). These internal marks alter the fate of the RNA in terms of splicing, choice of polyadenylation site, RNA export, and RNA stability and translation. These consequences are mediated by reader proteins, especially those from the YTH family that can recognize the internal m^{6}A marks (Fu et al., 2014; Patil et al., 2018; Roignant and Soller, 2017). The essential role of the internal m^{6}A marks is highlighted by the embryonic lethality exhibited by mouse mutants of METTL3 (Batista et al., 2014; Geula et al., 2015), METTL16



(legend on next page)

(Mendel et al., 2018), and YTHDC1 (Kasowitz et al., 2018), whereas mutants of YTHDC2 (Bailey et al., 2017; Hsu et al., 2017; Jain et al., 2018; Wojtas et al., 2017) and YTHDF2 (Ivanova et al., 2017) lead to infertility in mice.

Unlike the internal m^6A marks that have been shown to mediate RNA decay by the action of YTH proteins (Ke et al., 2017; Wang et al., 2014), the cap-specific m^6Am on the TSS adenosine is demonstrated to promote RNA stability by providing resistance to the action of the mRNA decapping enzyme DCP2 (Mauer et al., 2017). An activity responsible for cap-specific m^6Am methylation was partially purified as a ~65-kDa protein from human HeLa cell extracts (Keith et al., 1978), and this was recently revealed to be Phosphorylated CTD-Interacting Factor 1 (PCIF1) (Akichika et al., 2019; Boulias et al., 2019; Sendinc et al., 2019; Sun et al., 2019). PCIF1 was originally identified as a factor interacting with the phosphorylated CTD of RNA Pol II by its N-terminal WW domain (Fan et al., 2003; Hirose et al., 2008) and was later shown to have an affinity for the Ser5-phosphorylated CTD (Akichika et al., 2019). Methylation by PCIF1 seems to be efficient in human HEK293T cells, as up to 92% of m^7G-capped RNAs with a TSS adenosine carry the m^6Am mark, with the rest being Am (Akichika et al., 2019).

Early studies with *Pcif1* knockout (KO) chicken B cell DT40 lines did not identify any impact on cell cycle or viability (Yunokuchi et al., 2009). Removal of *PCIF1* in human cell lines also did not affect cell viability or growth (Akichika et al., 2019; Boulias et al., 2019; Sendinc et al., 2019). Investigations with *PCIF1* mutant human cells have supported the RNA stabilization role for m^6Am (Boulias et al., 2019) or implicated it in either enhancing (Akichika et al., 2019) or inhibiting (Sendinc et al., 2019) cap-dependent translation. Thus, although PCIF1-mediated m^6Am methylation is shown to impact diverse aspects of gene expression, its physiological role is currently not known. Here, we show that the cap-specific m^6Am modification catalyzed by PCIF1 is not essential for viability and fertility in mice. However, the *Pcif1* mutant mice display reduced body weight. Confirming an RNA-stabilizing role for m^6Am, transcripts starting with m^6Am are downregulated in the *Pcif1* mutant mouse tissues.

RESULTS

Loss of Mouse *Pcif1* Is Tolerated but Mutants Display Reduced Body Weight

To study the *in vivo* tissue distribution of mammalian PCIF1, we created a knockin mouse expressing a C-terminal FLAG-tagged version of the protein (Figure S1A and S1B; STAR Methods). The western blot analysis shows a near-ubiquitous expression of the protein at ~90 kDa, which is abundantly seen in the thymus, testis, and spleen. Expression in the kidney is very low, and the protein is not detectable in the salivary gland (Figure 1A). Similar results were obtained when detecting the endogenous protein in wild-type mouse tissues (Figure S1C). We consistently observed a second band at ~60 kDa in many of the tissues from both the wild-type and knockin mice (Figure 1A; Figure S1C), which we believe is an N-terminal truncated version of the protein (see STAR Methods). Immunofluorescence analysis using the anti-PCIF1 antibodies reveals a predominantly nuclear accumulation of PCIF1 in all tissues examined (Figure 1B).

To determine its physiological role, we created a *Pcif1* KO mouse model (Figures S1A and S1B; STAR Methods). Both heterozygous (HET) $Pcif1^{+/-}$ and homozygous $Pcif1^{-/-}$ KO animals of both sexes are viable, fertile, and born in the expected Mendelian ratio, indicating the absence of embryonic lethality (Figure S1D). Our western analyses of tissue lysates from the $Pcif1^{-/-}$ KO animals reveal a complete absence of the full-length protein, indicating that they are null mutants (Figures 1C and S1E). Consistent with the loss of the PCIF1 protein, the m^6Am modification is absent in polyA+ RNA from the $Pcif1^{-/-}$ KO testis (Figure 1D). We prepared ~200-nt m^7G-capped fragments of polyA+ RNAs from mouse liver (see STAR Methods) and analyzed separately the modifications on the full fragment (cap-TSS+ RNA body) versus that present only internally within the RNA body (no cap-TSS). The m^6Am modification is only detected on the TSS nucleotide and is lost in the $Pcif1^{-/-}$ mutant (Figure 1E). Expectedly, there is a slight increase of Am modification in the mutant, as it would normally be converted to m^6Am. Importantly, the overall levels of m^6A are not affected in the $Pcif1^{-/-}$ mutant (Figures 1D and 1E). Thus, mouse PCIF1 is the

Figure 1. *Pcif1* Mutant Mice Display Reduced Body Weight

(A) Multiple-tissue western analysis of *Pcif1-3xFLAG* knockin mouse showing presence of PCIF1-3xFLAG in most tissues. See also Figure S1C. Lung tissue from a wild-type (WT) animal is used to control signal from the anti-FLAG antibody. Signal from the PARK7 protein is used as a loading control.

(B) Subcellular localization of endogenous PCIF1 in mouse tissues. The protein is nuclear with some punctae observed (see liver and spleen). In testis sections, the protein is detected in the nucleus of PLZF-marked (green) undifferentiated spermatogonia (Spg), meiotic spermatocytes (Sp), post-meiotic round spermatids (RS), and somatic Sertoli cells.

(C) Domain architecture of mouse PCIF1. WW, tryptophan-rich domain; NLS, nuclear localization signal; helical region; MTase, methyltransferase domain. Creation of *Pcif1* knockout (KO) mouse by deletion of the region encoding the catalytic residues. See also Figures S1A and S1B. Western analysis with multiple tissues shows the absence of the protein in the KO ($Pcif1^{-/-}$) animals. See also Figure S1E for other biological replicates.

(D) RNA mass spectrometry confirms absence of m^6Am in polyA+ transcripts from the *Pcif1* KO mutant testis.

(E) Scheme showing preparation of m^7G-capped RNA fragments for RNA mass spectrometry to detect different modified nucleotides. Such fragments were treated to identify internal RNA modifications within the body of the transcript (excluding the cap and TSS nucleotide) and those from the entire fragment. RNA from mouse liver of indicated *Pcif1* genotypes was analyzed. Chemical structure of the eukaryotic cap structures indicating location of the m^6Am modification catalyzed by PCIF1 is shown.

(F) Northern analyses of total RNA from WT and *Pcif1* KO mutant mouse testis to detect U snRNAs. Signal from 5S rRNA is used as a loading control. The numbers indicate the biological triplicates tested. Quantification of signal intensities is shown on the right. Error bars show SD. See also Figure S1F for northern analyses with other tissues.

(G) Representative picture of adult $Pcif1^{-/-}$ KO mutant mice and control $Pcif1^{+/-}$ heterozygous (HET) littermates. The KO animals (6-weeks old) show reduced body weight. See also Figure S1G.

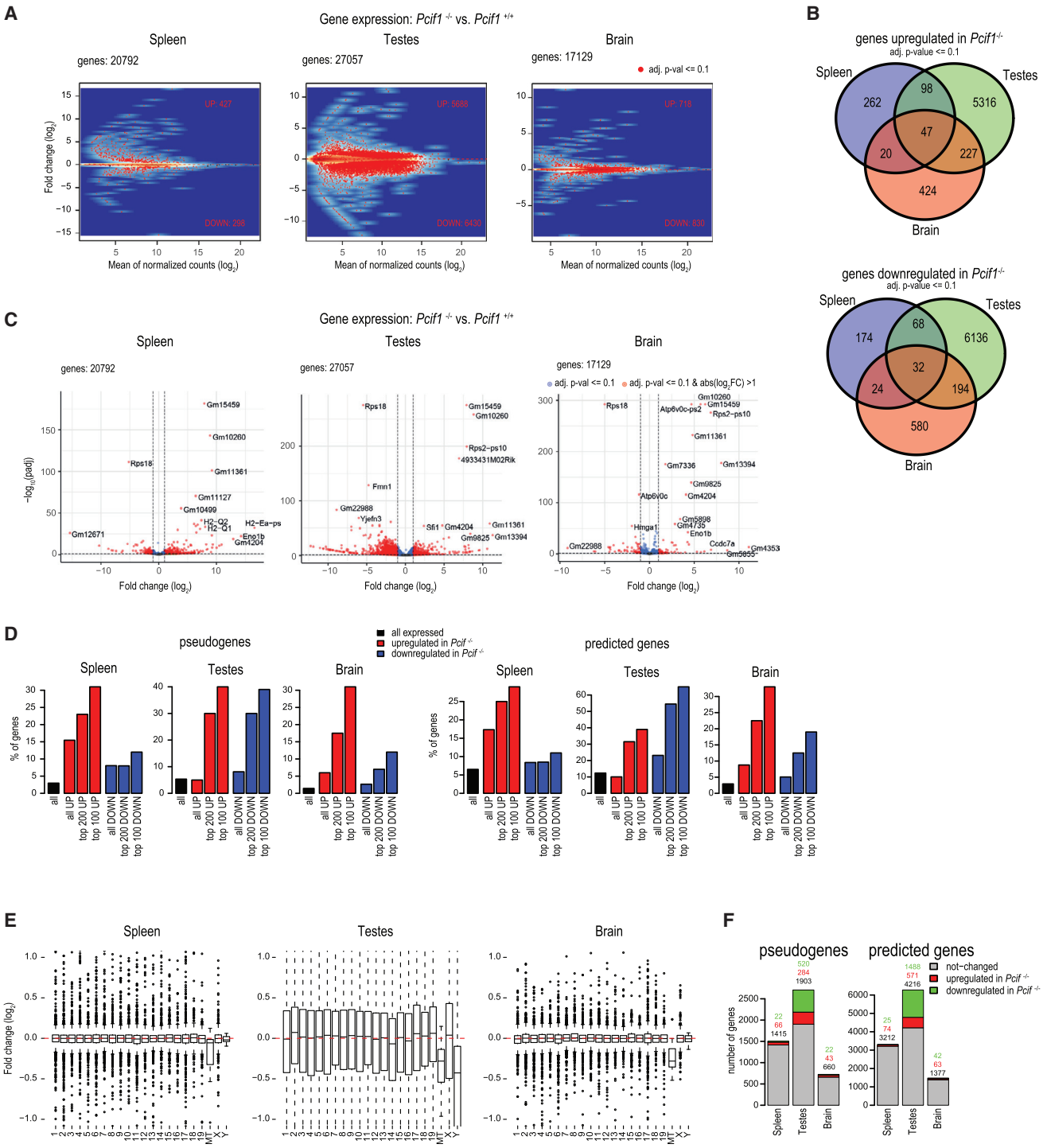


Figure 2. Pseudogene and Predicted Gene Transcript Levels Altered in the *Pcif1* Mutant Mouse Tissues

(A) MA plot showing transcript changes in the *Pcif1*$^{-/-}$ mutant mouse spleen, testes, and brain. Testicular transcriptome shows alteration of thousands of transcripts. See also Figure S2A.

(B) Venn diagram showing overlap in sets of upregulated and downregulated genes in different *Pcif1*$^{-/-}$ mutant mouse tissues.

(C) Volcano plot showing significant changes in levels for pseudogenes and predicted gene model (Gm) transcripts in the *Pcif1*$^{-/-}$ mutant tissues.

(D) Bar plots showing the enrichment of pseudogenes and predicted genes in the transcripts that are dysregulated in the *Pcif1*$^{-/-}$ mutant.

(legend continued on next page)

sole enzyme catalyzing the m^6Am modification on the TSS adenosine of m^7G-capped polyA+ transcripts.

Apart from messenger RNAs, RNA-Pol-II-transcribed splicing snRNAs like U1, U2, U4, and U5 also have the m^6Am modification on the TSS adenosine (Mauer et al., 2019; Wei et al., 2018). To examine the impact of the loss of PCIF1 on U snRNAs, we directly assessed their levels by northern blotting. In biological triplicates of total RNA from liver, spleen, and testis, we did not see any dramatic change in snRNA levels in the *Pcif1*$^{-/-}$ mutant (Figures 1F and S1F). Interestingly, despite the lack of any viability or fertility defects, we noticed that the *Pcif1*$^{-/-}$ mutant animals of both sexes consistently display reduced body weight compared to their HET littermates (Figure 1G). The overall body length of the *Pcif1*$^{-/-}$ mutants is not changed (Figure S1G), indicating that weight reduction is not due to a difference in growth during development. Taken together, our findings show that although a loss of mouse *Pcif1* is tolerated, the mutants have reduced body weight.

Dysregulation of Pseudogene and Predicted Gene Transcripts in *Pcif1* Mutant Mouse Tissues

To examine the global gene expression changes in the *Pcif1*$^{-/-}$ mutant, we carried out RNA sequencing (RNA-seq) analysis with total RNA from three different mouse tissues (Table S3). All the tissues show altered gene expression, with the testicular transcriptome being the most impacted (Figures 2A and S2A). Hundreds (in spleen and brain) to thousands (in testis) of transcripts are either up- or downregulated. There is no considerable overlap in terms of the altered transcripts across the tissues (Figure 2B), supporting the idea that PCIF1 regulates transcripts that are largely specific to each tissue. Surprisingly, among the top altered sequences in all the tissues are several pseudogenes and poorly characterized predicted protein-coding Gene model (Gm) transcripts (Figure 2C). Because some of the pseudogenes are annotated as predicted genes and vice versa, this is a largely overlapping set of transcription units.

Pseudogenes are inactive copies of functional genes that arose during evolution, either by tandem duplication or activity of retrotransposons (Pink et al., 2011). Compared to their parental genes, they have acquired mutations that render them non-functional in terms of protein-coding potential. Although duplicated pseudogenes still maintain their intron-exon organization, the processed pseudogenes arise from mature cDNAs due to retrotransposition activity of endogenous transposons. Pseudogenes are as numerous as protein-coding genes (up to 20,000 in the human genome), with the bulk of them being processed as pseudogenes (Zhang and Gerstein, 2004). Although many of the pseudogenes are not transcribed due to a loss of promoter sequences or due to an insertion of processed pseudogenes into regions lacking promoter activity, about 2%–20% of the human pseudogenes are transcribed (Harrison et al., 2005; Zheng et al., 2005). In our mouse tissue RNA-seq datasets, pseudogenes constitute only a small part (3% of spleen, 5% of testis, and 2% of brain) of the expressed genes in these tissues. Strikingly, despite being only a small part of the transcriptome, the pseudogenes constitute a significant chunk (30%–40%) of the top 100 upregulated genes (Figure 2D). Pseudogenes also form a substantial part (10%–40%) of the top 100 downregulated genes (Figure 2D). Similarly, predicted Gm transcripts constitute up to 25%–40% of the top 100 upregulated and 10%–60% of the top downregulated genes in the *Pcif1* mutant (Figure 2D).

The altered transcripts in the *Pcif1*$^{-/-}$ tissues do not show any particular genomic origins (Figure S2B). Interestingly, most of the testicular transcripts arising from the Y chromosome are largely downregulated (Figure 2E), with many being pseudogenes (Figure S2B). Likewise, transcripts from the mitochondrial genome are consistently downregulated in all the three tissues (Figure 2E). Because PCIF1 likely acts only in the nuclear compartment, we expect the change in mitochondrial transcriptome to be an indirect consequence of other gene expression changes. Although changes in the expression of pseudogene and predicted genes in the *Pcif1*$^{-/-}$ mutant are striking, they constitute only a small subset of such expressed sequences (Figure 2F). Taken together, we identify a set of pseudogenes and predicted Gm transcripts as being regulated by mouse PCIF1, with Y-chromosome-derived pseudogenes being heavily downregulated in the mouse testis.

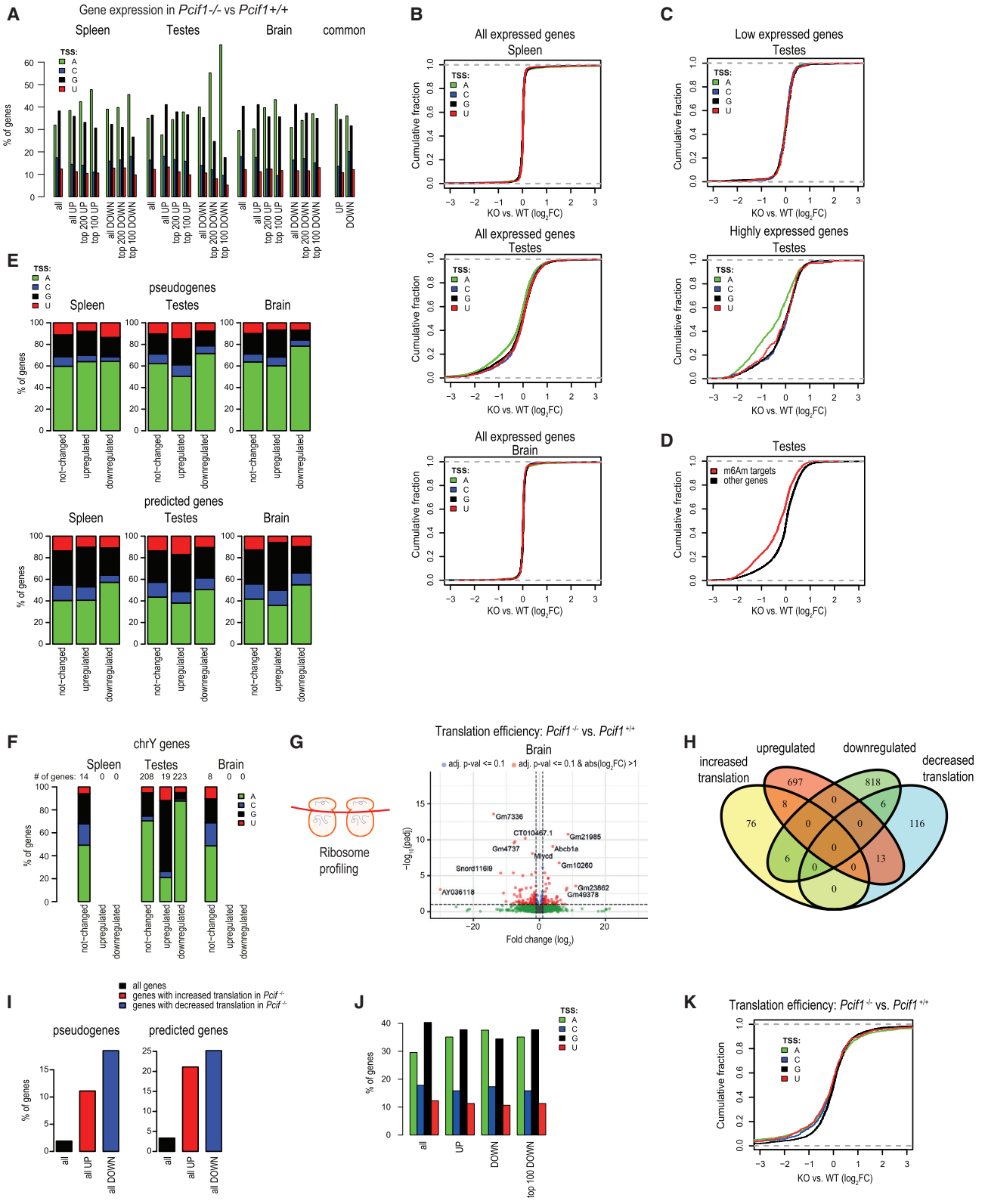
Transcripts with a TSS Adenosine Are Downregulated in *Pcif1* Mutant Mouse Tissues

To identify a link between the observed transcriptome changes and the activity of PCIF1, we first sorted the genes based on their annotated TSS nucleotide (STAR Methods). As expected (Carninci et al., 2006; Yamashita et al., 2006), most genes produce transcripts that start with either an adenosine or guanosine. By comparing the gene expression in the *Pcif1*$^{-/-}$ mutant with the control wild type (*Pcif1*$^{+/+}$), we calculated the proportion of dysregulated genes with a specific TSS nucleotide. Strikingly, in the testis datasets, we see an enrichment of transcripts starting with an adenosine among the top 100 and top 200 downregulated genes in the *Pcif1* mutant (Figure 3A). Likewise, when all the expressed genes are considered, those with an adenosine as the TSS nucleotide generally show lower expression in the *Pcif1*$^{-/-}$ mutant testes (Figure 3B). When the genes in the respective tissues were further sorted as high or low expressed (STAR Methods), we found that mainly highly expressed genes with a TSS adenosine are downregulated in the *Pcif1*$^{-/-}$ mutant testes (Figure 3C). Such a global relationship was not observed in the spleen and brain samples (Figures 3B and S3A) but was seen for specific transcript sets (see below).

To further link the change in testicular RNA levels to a loss of the m^6Am methylation, we experimentally defined a high-confidence set of m^6Am-target genes by using a previously described protocol (Sendinc et al., 2019). Briefly, ~200-nt capped RNA fragments from mouse testes were prepared and used for m^6A-IPseq, and enrichments were calculated (STAR Methods).

(E) Boxplots showing gene expression changes from individual chromosomes, as observed in the different *Pcif1*$^{-/-}$ mutant mouse tissues. The view is limited to only changes below 2-fold to better see the median differences. See also Figure S2B.

(F) Bar plots showing the total number of pseudogenes and predicted genes detected in the indicated mouse tissues and the fraction of those that are up- and downregulated in the *Pcif1*$^{-/-}$ mutant.



(legend on next page)

The anti-m^{6}A antibody which does not differentiate between m^{6}A and m^{6}Am, enriches the capped RNA fragments (Figure S3B). Supporting the fact that some of this enrichment is due to cap-specific m^{6}Am catalyzed by PCIF1, there is a clear decrease in the distribution of reads over the TSS in the *Pcif1*$^{-/-}$ mutant (Figure S3B). Furthermore, a clear and significant decrease (Figure S3C) in the peak is observed mainly at those TSSs with an annotated adenosine, indicating they are likely m^{6}Am targets *in vivo* (Table S4). Notably, the annotated TSS is the most dominant documented one from a collection of start sites at a transcription unit, so it is possible that the decrease seen in the *Pcif1*$^{-/-}$ mutant at non-adenosine TSS is just a reflection of this heterogeneity (Figure S3C). Importantly, when the expression levels of such high-confidence m^{6}Am-target genes were plotted, they show a clear decrease in the *Pcif1*$^{-/-}$ mutant testes (Figure 3D). When the TSS nucleotide was specifically examined for pseudogenes, we found those with an adenosine as the TSS to be downregulated in the testis and brain datasets (Figure 3E). Similarly, predicted genes from all three tissues that are downregulated tend to have an adenosine as the TSS nucleotide (Figure 3E). As noted previously, the Y chromosome genes tend to be downregulated in the *Pcif1*$^{-/-}$ mutant (Figure 2E), and ~90% of these downregulated genes use an adenosine as the TSS nucleotide (Figure 3F). These findings are consistent with the previous observation that the m^{6}Am modification has a transcript-stabilizing role (Mauer et al., 2017).

Finally, we examined if there is any impact on translation by carrying out ribosome profiling experiments with a *Pcif1* mutant brain (Table S5). An analysis of ribosome footprints indicates a significant alteration in the translation of several transcripts (Figure 3G), without any changes in transcript levels for most of them (Figure 3H). We find some predicted genes and pseudogenes to be among those that show changes in translation (Figure 3I). Note that although pseudogenes have mutations that prevent full-length protein production, this does not per se preclude their entry into ribosomes. However, the genes that show translation changes do not have any particular preference for the TSS nucleotide (Figures 3J and 3K). Considered together, these data indicate that cap-specific m^{6}Am methylation of RNA Pol II transcripts by mouse PCIF1 has an RNA-stabilizing role in mouse tissues, with a subset of pseudogenes and poorly characterized predicted genes being a common set of regulated transcripts.

Drosophila Pcif1 Is Catalytically Dead but Binds Ser5-Phosphorylated RNA Pol II CTD

The m^{6}Am modification is so far detected only in vertebrates (Sendinc et al., 2019). It is absent in worms and insects (Figures 4A and S4A). Although Pcif1 is absent in yeasts and worms, a clear ortholog with a high level of conservation is detected in several insect genomes (Figure S4B). It is represented in the fruit fly *Drosophila* by the uncharacterized gene CG11399 (dPcif1). Rabbit polyclonal antibodies raised against dPcif1 detects the endogenous protein in *Drosophila* ovaries and in the *Drosophila* S2 cell line (Figure 4B). Nevertheless, RNA mass spectrometry did not detect the m^{6}Am modification in adult flies, early fly embryos (2 h post-fertilization), and in the S2 cell line (Figure 4C). Transfecting S2 cells with a plasmid expressing human PCIF1 (hPCIF1) shows the ectopic installation of m^{6}Am on fly transcripts (Figure 4C), validating our detection strategy. The fly protein is nuclear in transfected *Drosophila* S2 cells (Figure S4C). Similarly, the endogenous dPcif1 protein is also nuclear in fly embryos (Figure S4D) and adult fly ovaries, in which it is excluded from the nucleoli that are positively stained by fibrillarin (Figure 4D). The protein is detected in both the somatic follicle cells and germline nurse cells of individual egg chambers in the fly ovaries (Figure 4D).

Examination of a protein sequence alignment shows that compared to mammalian PCIF1, the insect orthologs have mutations in the putative catalytic residues that can potentially render them inactive (Figures 4E and S4B). To directly verify this finding, we expressed a recombinant version of the full-length *Drosophila* Pcif1 protein (Figure S5A) and incubated it with m^{7}G-capped RNAs, together with radioactive ^{14}C-SAM as the methyl-donor. Recombinant hPCIF1 was used as a positive control (Figure S5A). As expected, the human protein shows methylation activity with an m^{7}G-capped RNA having an adenosine as the TSS nucleotide (m^{7}GpppA-RNA). However, consistent with mutation of the putative catalytic residues, the fly protein was

Figure 3. Transcripts Starting with an Adenosine Are Downregulated in the Mouse *Pcif1* Mutant

(A) Percentage of genes with a specific nucleotide as a transcription start site (TSS) is compared between all expressed genes, and the genes found to be up- and downregulated in the *Pcif1*$^{-/-}$ KO mutant mouse tissues compared to the WT, *Pcif1*$^{+/+}$) control.

(B) Plots showing the cumulative distributions of gene expression changes in the *Pcif1*$^{-/-}$ KO mutant mouse tissues compared to the WT control. The distributions of genes with different TSS nucleotides are shown separately.

(C) Plots showing the cumulative distributions of gene expression changes in the *Pcif1*$^{-/-}$ mutant mouse testes separately for low- and high-expressed genes. See also Figure S3A.

(D) Plot showing the cumulative distributions of gene expression changes in the *Pcif1*$^{-/-}$ mutant testes for those genes experimentally determined to express transcripts with m^{6}Am (m^{6}Am targets) and those without m^{6}Am (other genes). See also Figures S3B and S3C.

(E) Bar plots showing the percentage of pseudogenes and predicted genes with specific TSS nucleotides among the genes up- and downregulated in the *Pcif1*$^{-/-}$ mutant mouse tissues.

(F) Bar plots showing percentage of chromosome Y pseudogenes and predicted genes with specific TSS nucleotides among the genes with same or differential expression in the *Pcif1*$^{-/-}$ mutant mouse tissues compared to the WT control.

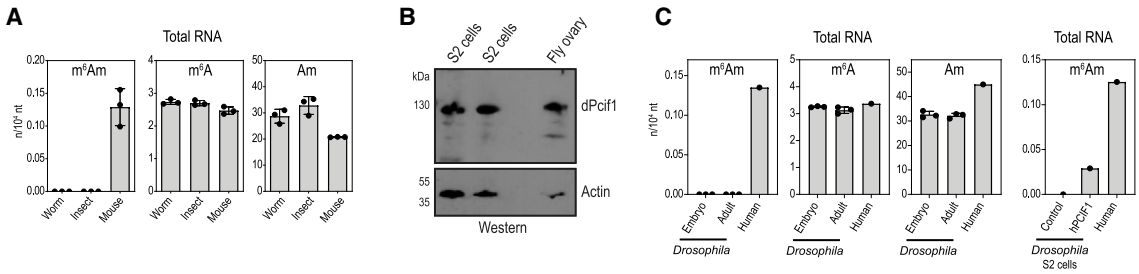
(G) Ribosome profiling of the *Pcif1*$^{-/-}$ mutant brain. Volcano plot shows changes in translation efficiency in the mutant compared to that of WT control. Several predicted genes are dramatically altered in their translation.

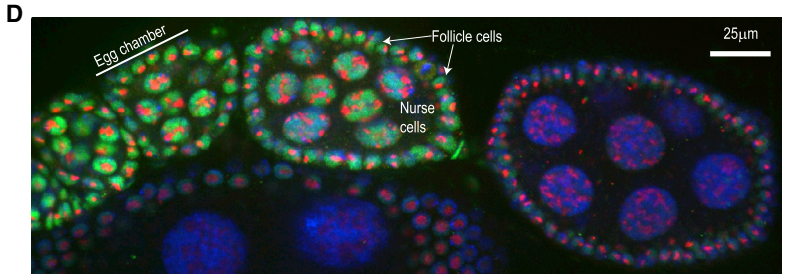
(H) Venn diagram comparing changes (upregulated and downregulated) in transcript levels and translation in the *Pcif1*$^{-/-}$ mutant brain.

(I) Bar plots showing the enrichment of predicted genes and pseudogenes among the genes with altered translation efficiency in *Pcif1*$^{-/-}$ mutant brain.

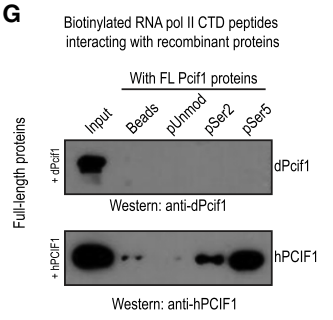
(J) Percentage of genes with specific TSS nucleotides is compared between genes with increased and decreased translation efficiency in the *Pcif1*$^{-/-}$ mutant brain.

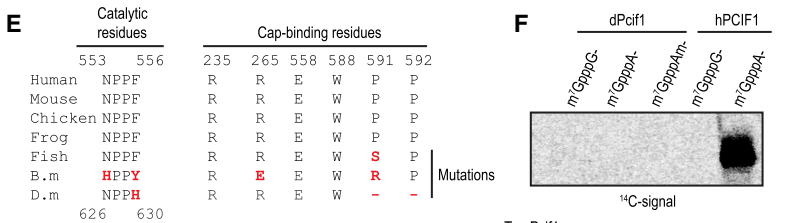
(K) Plot showing the cumulative distribution of changes in translation efficiency in the *Pcif1*$^{-/-}$ mutant mouse brain compared to the WT control.

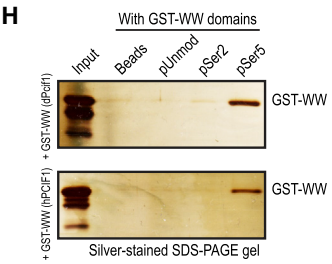


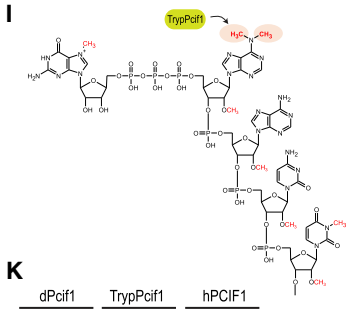


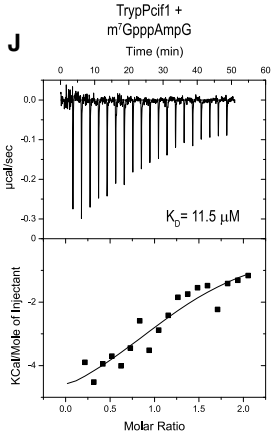
Immunofluorescence with adult fly ovary (5-days old): anti-dPcif1 (green), anti-Fibrillarin (red), DAPI (blue)

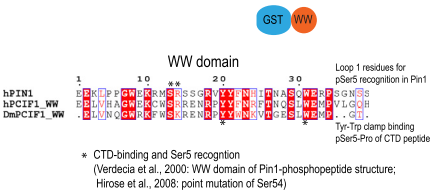


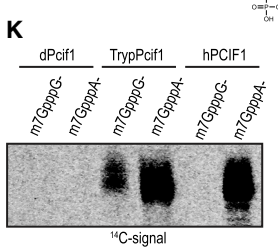


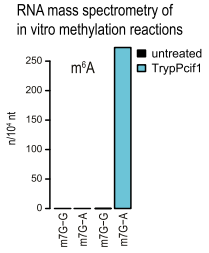


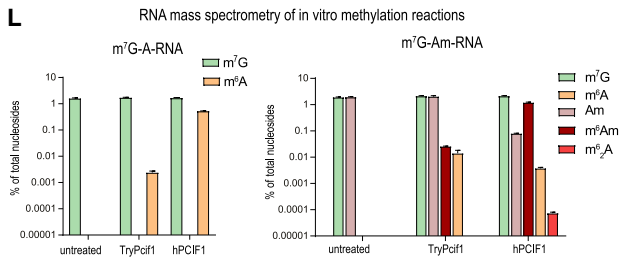












(legend on next page)

inactive as an RNA methylase on all tested RNAs (Figure 4F). Next, we examined whether the fly Pcif1 can bind the m^7G cap structure using isothermal calorimetry (ITC) experiments. Such direct binding experiments reveal that the recombinant full-length fly Pcif1 protein does not associate with the cap analogs m^7GpppA or m^7GpppG (Figure S5B). This is consistent with the fact that some of the cap-binding residues conserved in vertebrate PCIF1 proteins are absent in the fly protein (Figures 4E and S4B).

Next, we examined the CTD-interaction activity of fly Pcif1. The RNA Pol II CTD is composed of multiple repeats of a heptad sequence (YSPTSPS) that can be phosphorylated at different Serine (Ser) residues (Buratowski, 2009). We directly tested this interaction and confirm that recombinant full-length hPCIF1 preferentially associates with the Ser5-phosphorylated CTD peptide (Figures 4G and S5C). However, the full-length fly Pcif1 failed to bind any of the tested CTD peptides (Figures 4G and S5C). This finding was surprising, as the overall WW domain of PCIF1 is highly conserved from insects to human (Figure 4H) and is similar to that found in the prolyl isomerase hPIN1. The crystal structure of the hPIN1 WW domain in complex with the phospho CTD peptide is solved (Verdecia et al., 2000). This reveals key WW domain residues like the Tyr-Trp clamp (Y61-W72 in hPCIF1) that holds the Ser5-Pro motif of the CTD peptide [YSPT(pS)PS], and the Ser-Arg/Lys (S54-R55 in hPCIF1) of loop1 that recognizes the phosphorylated Ser5. Mutation of the S54 in the hPCIF1 WW domain abolishes the interaction with the CTD (Hirose et al., 2008). All these key residues are also conserved in the fly Pcif1 (Figure 4H). Importantly, when the binding assay was performed with the isolated WW domains from both proteins, fly Pcif1 also specifically binds the Ser5-phosphorylated CTD peptide (Figures 4H and S5D).

Taken together, our experiments reveal that the fly Pcif1 is biochemically inert as an RNA methylase and does not bind the m^7G cap structure. However, it retains an active WW domain that is able to bind Ser5-phosphorylated CTD of the RNA Pol II. We propose that fly Pcif1 might have evolved to function as a nuclear factor that transcriptionally regulates gene expression.

Trypanosoma Pcif1 Catalyzes m^6Am That Is Part of the cap4 Structure

The 5′ end of RNA Pol II transcripts in most eukaryotes is defined by transcription initiation. In contrast, nematodes receive a trimethylated guanosine cap ($m^{2,2,7}G$) by *trans*-splicing of a spliced leader (SL) sequence (Lasda et al., 2011), but they lack m^6Am due to the absence of a PCIF1 ortholog. The situation with trypanosomatid protozoans is partly similar to that of nematodes in that they also acquire an extensively modified cap4 structure (m^7Gpppm$^6{}_2$Am-p-Am-p-Cm-p-m^3Um) by *trans*-splicing. The structure consists of the first four nucleotides being ribose methylated (Nm), in addition to having a N^6,N^6-dimethyl modification on the first adenosine ($m^6{}_2$A), and N^3-methyl modification on the uridine (m^3U) at the fourth position (Figure 4I). The cap4 structure is first formed on the specialized SL RNA and then transferred to the pre-mRNAs by a *trans*-splicing reaction (Mair et al., 2000). Because the activity involved in the formation of the $m^6{}_2$A in the cap4 structure is currently not known, we wondered whether a Pcif1 ortholog could play a role.

We identified the Pcif1 ortholog in *Trypanosoma cruzi*, the trypanosomatid responsible for South American trypanosomiasis (Chagas disease). The *Trypanosoma* Pcif1 (TrypPcif1) lacks a WW domain but contains the NPPF catalytic residues as present in the vertebrate PCIF1 (Figure S5E). To test its biochemical properties, we expressed the TrypPcif1 as a recombinant protein (Figure S5F). TrypPcif1 has a strong affinity (dissociation constant, K_D = 11.5–29.9 μM) for the cap1 structure represented by the m^7GpppAmpG cap analog (Figures 4J and S5F), whereas binding to m^7GpppA and m^7GpppG are

Figure 4. *Drosophila* Pcif1 Is Catalytically Dead but Interacts with the Ser5-Phosphorylated RNA Polymerase II (RNA Pol II) C-Terminal Domain (CTD)

(A) RNA mass spectrometry analysis of total RNA from worm (*Caenorhabditis elegans*), insect (*Bombyx* BmN4 cell line), and adult mouse testis RNA. Worms and insects lack the m^6Am modification. See also Figure S4A.

(B) Western analysis of endogenous dPcif1 in fly ovaries and *Drosophila* S2 cell line.

(C) RNA mass spectrometry analysis of adult *Drosophila*, early (2-h) fly embryos, and human HAP1 cell line as a control. Also used was RNA from *Drosophila* S2 cells with or without (control) transfection of a plasmid expressing human PCIF1. See also Figure S4A.

(D) dPcif1 protein is detected in adult fly ovaries. Egg chambers show the protein being nuclear in the single layer of somatic follicle cells and the germline nurse cells contained within each chamber. See also Figures S4C and S4D.

(E) Insect Pcif1 proteins (B.m., *Bombyx mori*; D.m., *Drosophila melanogaster*) have mutations of the catalytic residues in the MTase and in the critical residues that are required for m^7G cap binding. See also Figure S4B.

(F) *In vitro* methylation assays with full-length recombinant *Drosophila* or human PCIF1 proteins. The reactions used m^7G-capped RNAs with either an adenosine or a guanosine or Am as the TSS nucleotide and radioactive ^{14}C-SAM as the methyl donor. See also Figure S5A.

(G) *In vitro* binding assay between recombinant full-length fly or human PCIF1 and different biotinylated peptides with heptad repeats from the CTD of RNA Pol II. Interaction was assayed by western blotting with specific antibodies. Phosphorylation at Ser2 or Ser5 of the CTD heptad sequence is indicated. Unmod, non-phosphorylated peptide; beads, beads alone without any bound peptide. See also Figure S5C.

(H) Alignment of the WW domains from human and *Drosophila* Pcif1 and human PIN1. CTD peptide binding assay with recombinant GST-WW domain fusions from human and fly Pcif1. Bound GST-WW proteins are detected by silver staining. See also Figure S5D.

(I) The cap4 structure of trypanosomatids showing the presence of a dimethyladenine ($m^6{}_2$A) at the TSS nucleotide.

(J) Isothermal calorimetry (ITC) experiment showing binding between recombinant full-length *Trypanosoma* Pcif1 (TrypPcif1) and the indicated cap1 analog. See also Figure S5F.

(K) *In vitro* methylation assays with recombinant full-length PCIF1 proteins from *Drosophila*, *T. cruzi*, and human. The m^7G-capped RNAs have either an adenosine or a guanosine as the TSS nucleotide, and radioactive ^{14}C-SAM was used as methyl donor. Reactions are resolved by urea-PAGE. The size of the capped RNA used is 40 nt. RNA mass spectrometry analysis of *in vitro* methylation reactions carried out with TrypPcif1, revealing the catalysis of m^6A.

(L) RNA mass spectrometry analysis of *in vitro* methylation reactions carried out with human PCIF1 or TrypPcif1 using m^7G-capped RNAs with either A or Am as the TSS nucleotide.

highly reduced (Figure S5F). To test its RNA methylation activity, TrypPcif1 was incubated with an m^7G-capped RNA and radioactive ^{14}C-SAM as a methyl donor. This revealed a robust methylation activity on the m^7GpppA-RNA (Figure 4K). We note that the TrypPcif1 shows an unexplained weak activity with the m^7GpppG-RNA, but this is not seen with the hPCIF1. RNA mass spectrometry analysis of reactions conducted in parallel with non-radioactive SAM shows the clear presence of m^6A in the reactions treated with TrypPcif1 (Figure 4K). When TrypPcif1 is incubated with the m^7GpppAmpG-RNA carrying the cap1 modification, we detect a signal for m^6Am, but the $m^6{}_2$A dimethyladenosine modification was never reliably detected (Figure 4L). Considered together, our findings show that the protozoan Pcif1 is a cap-dependent m^6Am RNA methylase that contributes to the cap4 structure of trypanosomatids.

DISCUSSION

Despite being a ubiquitous mark on RNA Pol II transcripts starting with an adenosine (Akichika et al., 2019), we were surprised to find that a loss of cap-specific m^6Am did not result in dramatic phenotypes in mutant mice (Figure 1). However, a clear contribution to optimal growth is evident, as the *Pcif1* KO mice display a reduced body weight phenotype (Figure 1G). RNA-seq analysis shows that PCIF1 has effects on the abundance of hundreds to thousands of transcripts in the mutant tissues (Figure 2A). One common set of dysregulated transcripts is a subset of pseudogenes and other poorly characterized predicted genes (Figure 2D). Given their sequence homology, pseudogenes that have lost their protein-coding capacity are proposed to regulate their parental genes by acting as non-coding RNAs (Yano et al., 2004). We explored this possibility by first identifying expressed pseudogene/parental gene pairs, and then we compared their levels in the *Pcif1* mutant. Only in the mutant brain we saw a negative correlation between expression levels of pseudogenes and that of their parental genes (Figure S2C). During spermatogenesis, mitotic spermatogonia enter meiosis to become spermatocytes, which later complete meiosis to form haploid spermatids, and further mature into sperm (Oakberg, 1956). We find that the top upregulated transcripts in the *Pcif1*$^{-/-}$ mutant testis are those normally present in mitotic spermatogonia, whereas the top downregulated transcripts are normally those highly expressed in the meiotic spermatocytes and haploid round spermatids (Figure S3D). This finding could suggest an apparent failure to proceed along spermatogenesis. However, we do not see any impact on fertility in the young *Pcif1*$^{-/-}$ mutants, although long-term impact in aged animals needs to be evaluated.

Several molecular roles are attributed to the m^6Am modification. The presence of m^6Am on the cap structure is shown to provide stability to transcripts in human HEK293T cell cultures (Boulias et al., 2019; Mauer et al., 2017). Our own study of the *Pcif1* KO mouse tissues confirms this impact on RNA levels (Figure 3). However, the impact on RNA stability is not an observation that is broadly shared by other studies that examined *Pcif1* KO human cell cultures (Akichika et al., 2019; Sendinc et al., 2019). Instead, they reveal a role for m^6Am in regulating translation, although opposing effects are proposed: promoting (Akichika et al., 2019) or repressing cap-dependent translation (Sendinc et al., 2019). Our own analysis of ribosome footprints in the *Pcif1* mutant mouse brain identified several transcripts that show either up- or downregulated translation (Figure 3G), but we do not find any correlation with the identity of the TSS nucleotide (Figures 3J and 3K). Given the results of earlier studies, a tissue-specific effect on translation should be left open as a possibility.

How the catalytically dead fly Pcif1 finds its target genes and what molecular consequences it has on them are not clear. One possibility is that it interacts with initiating RNA Pol II by its Ser5-phosphorylated CTD (Figure 4H) to transcriptionally regulate gene expression. Because only the isolated WW domain (Figures 4H and S5D), but not the full-length fly Pcif1 (Figures 4G and S5C), is able to interact with the CTD, we speculate that conformational changes might be needed to expose the WW domain. Perhaps interaction with some partner protein might allow this structural change. Interestingly, it was previously noted that due to its affinity for the CTD, the WW domain of hPCIF1 or full-length PCIF1 can interfere with dephosphorylation of the CTD *in vitro* (Hirose et al., 2008). Given that a change in CTD phosphorylation status from Ser5 to Ser2 is critical for transcription elongation (Buratowski, 2009), the observed CTD interaction of fly Pcif1 could have a regulatory role at the transcriptional level.

Finally, our analysis of the *Trypanosoma* Pcif1 identifies it as the activity that catalyzes N^6 methylation of the TSS adenosine within the cap4 structure of the protozoan mRNAs (Figure 4K). The TSS adenosine of the cap4 structure is dimethylated at the N^6 position ($m^6{}_2$A), but we did not find such an activity in the recombinant TrypPcif1 protein. Perhaps the sequential co-transcriptional installation of the other modifications in a hierarchical manner (Mair et al., 2000) is a prerequisite for dimethylation. Indeed, reduced ribose methylation on some of the positions can affect methylation at other sites of the cap4 structure in *Trypanosoma brucei* (Zamudio et al., 2006). Alternatively, TrypPcif1 might install m^6A, with yet another unidentified methylase completing the reaction. Collectively, our combined analyses of PCIF1 in three model systems shed light on its evolutionary functional plasticity in catalytic and non-catalytic roles.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animal Work
  - Pcif1 knockout and Flag-Pcif1 knockin mice
  - Synthesis and annealing of gRNA
  - Preparation of injection mix
  - Genotyping
- METHOD DETAILS

- Clones and constructs
- Constructs for *Drosophila* S2 cell expression
- Constructs for recombinant protein production in prokaryotic expression systems
- Constructs for recombinant protein production in insect expression systems
- Recombinant protein production
- Purification of human, fly and Trypanosoma PCIF1
- Purification of GST-fused human and fly PCIF1 WW domain

- ANTIBODIES
  - Commercial antibodies
  - Antibodies generated for this study
  - Antibody purification
  - *Drosophila* S2 cell culture and transfections
  - Mouse multiple tissue western blot
  - Western Blot
  - Immunofluorescence analysis
  - *Drosophila* ovaries
  - *Drosophila* embryos
  - *Drosophila* S2 cells
  - Northern blot of U snRNAs
  - Quantification of RNA modifications using LC-MS/MS
  - Preparation of m^7G-cap enriched RNA fragment for RNA mass spectrometry
  - Preparation of m^7G-capped RNA by *in vitro* transcription
  - *In vitro* methylation assays with PCIF1
  - Phosphorylated CTD-binding assays with recombinant PCIF1
  - Isothermal calorimetry (ITC) experiments
  - Preparation of RNA libraries
  - Ribosome profiling
  - m^6Am-Exo-Seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Analysis of RNaseq data
  - Analysis of ribosome footprinting data
  - Analysis of m^6Am-Exo-Seq data

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.108038.

ACKNOWLEDGMENTS

We thank François Karch for fly embryos and help with staining; Pascal Gos and Fabienne Fleury-Olela for assistance with experiments; Johanna Mattay and Michaela Dohnalkova for critical reading of the manuscript; and Nicolas Roggli for scientific illustration. We thank Andrew McCarthy, European Molecular Biology Laboratory (EMBL) Grenoble, France, for facilitating the ITC experiments. We thank the following University of Geneva core facilities: Institute of Genetics and Genomics of Geneva (iGE3) Genomics Platform and the Transgenic Mouse Facility. We also thank the EMBL Genomics core facility for deep sequencing. A.R. receives a Swiss Government Excellence Postdoctoral Scholarship, L.L. is supported by the iGE3 PhD Fellowship, and G.F. receives a PhD Fellowship from the Initiative d'excellence (IDEX) program of the University of Grenoble Alpes funded by the French l'Agence Nationale de la Recherche (ANR-15-IDEX-02). This work was supported by grants to R.S.P. from the Swiss National Science Foundation: European Research Council (ERC) transfer grant (CRETP3_166923), project grant (310030B_185386), and Sinergia grant (CRSII5_183524), and by funding from the Swiss National Center of Competence in Research (NCCR) RNA & Disease (51NF40_182880). Work in the Pillai lab is supported by the Republic and Canton of Geneva.

AUTHOR CONTRIBUTIONS

R.R.P. generated the Pcif1 mouse lines and prepared deep sequencing libraries and ribosome profiling experiments; E.D. carried out the biochemical analyses of Pcif1; D.H. performed computational analyses; K.-M.C. performed ITC experiments; A.R. and G.F. carried out immunofluorescence analyses and Northern analysis of snRNAs; L.L. produced recombinant fly Pcif1; G.F., E.T., and M.-O.F. carried out analysis of the fly *Pcif1* mutant not included in this study; C.B.V. analyzed RNA modifications; and manuscript preparation and writing were performed by R.S.P. and R.R.P. with input from everyone.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 23, 2020
Revised: July 10, 2020
Accepted: July 23, 2020
Published: August 18, 2020

REFERENCES

Akichika, S., Hirano, S., Shichino, Y., Suzuki, T., Nishimasu, H., Ishitani, R., Sugita, A., Hirose, Y., Iwasaki, S., Nureki, O., and Suzuki, T. (2019). Cap-specific terminal N^6-methylation of RNA by an RNA polymerase II-associated methyltransferase. Science *363*, eaav0080.

Bailey, A.S., Batista, P.J., Gold, R.S., Chen, Y.G., de Rooij, D.G., Chang, H.Y., and Fuller, M.T. (2017). The conserved RNA helicase YTHDC2 regulates the transition from proliferation to differentiation in the germline. eLife *6*, e26116.

Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K., et al. (2014). m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell *15*, 707–719.

Bélanger, F., Stepinski, J., Darzynkiewicz, E., and Pelletier, J. (2010). Characterization of hMTr1, a human Cap1 2′-O-ribose methyltransferase. J. Biol. Chem. *285*, 33037–33044.

Bieniossek, C., Imasaki, T., Takagi, Y., and Berger, I. (2012). MultiBac: expanding the research toolbox for multiprotein complexes. Trends Biochem. Sci. *37*, 49–57.

Blighe, K. (2019). EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling (Bioconductor). https://bioconductor.org/packages/release/bioc/html/EnhancedVolcano.html.

Boulias, K., Toczydlowska-Socha, D., Hawley, B.R., Liberman, N., Takashima, K., Zaccara, S., Guez, T., Vasseur, J.J., Debart, F., Aravind, L., et al. (2019). Identification of the m(6)Am Methyltransferase PCIF1 Reveals the Location and Functions of m(6)Am in the Transcriptome. Mol. Cell *75*, 631–643.e638.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol. Cell *36*, 541–546.

Byszewska, M., Śmietański, M., Purta, E., and Bujnicki, J.M. (2014). RNA methyltransferases involved in 5′ cap biosynthesis. RNA Biol. *11*, 1597–1607.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engström, P.G., Frith, M.C., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet. *38*, 626–635.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Coleman, T.M., Wang, G., and Huang, F. (2004). Superior 5′ homogeneity of RNA from ATP-initiated transcription under the T7 ϕ 2.5 promoter. Nucleic Acids Res. *32*, e14.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201–206.

Doxtader, K.A., Wang, P., Scarborough, A.M., Seo, D., Conrad, N.K., and Nam, Y. (2018). Structural Basis for Regulation of METTL16, an S-Adenosylmethionine Homeostasis Factor. Mol. Cell *71*, 1001–1011.e1004.

Fan, H., Sakuraba, K., Komuro, A., Kato, S., Harada, F., and Hirose, Y. (2003). PCIF1, a novel human WW domain-containing protein, interacts with the phosphorylated RNA polymerase II. Biochem. Biophys. Res. Commun. *301*, 378–385.

Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene expression regulation mediated through reversible m^6A RNA methylation. Nat. Rev. Genet. *15*, 293–306.

Furuichi, Y., Morgan, M., Muthukrishnan, S., and Shatkin, A.J. (1975a). Reovirus messenger RNA contains a methylated, blocked 5′-terminal structure: m-7G(5′)ppp(5′)G-MpCp-. Proc. Natl. Acad. Sci. USA *72*, 362–366.

Furuichi, Y., Morgan, M., Shatkin, A.J., Jelinek, W., Salditt-Georgieff, M., and Darnell, J.E. (1975b). Methylated, blocked 5 termini in HeLa cell mRNA. Proc. Natl. Acad. Sci. USA *72*, 1904–1908.

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A.A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y.S., et al. (2015). Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. Science *347*, 1002–1006.

Haline-Vaz, T., Silva, T.C., and Zanchin, N.I. (2008). The human interferon-regulated ISG95 protein interacts with RNA polymerase II and shows methyltransferase activity. Biochem. Biophys. Res. Commun. *372*, 719–724.

Harrison, P.M., Zheng, D., Zhang, Z., Carriero, N., and Gerstein, M. (2005). Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. Nucleic Acids Res. *33*, 2374–2383.

Hirose, Y., Iwamoto, Y., Sakuraba, K., Yunokuchi, I., Harada, F., and Ohkuma, Y. (2008). Human phosphorylated CTD-interacting protein, PCIF1, negatively modulates gene expression by RNA polymerase II. Biochem. Biophys. Res. Commun. *369*, 449–455.

Hsu, P.J., Zhu, Y., Ma, H., Guo, Y., Shi, X., Liu, Y., Qi, M., Lu, Z., Shi, H., Wang, J., et al. (2017). Ythdc2 is an N^6-methyladenosine binding protein that regulates mammalian spermatogenesis. Cell Res. *27*, 1115–1127.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods *12*, 115–121.

Inesta-Vaquera, F., and Cowling, V.H. (2017). Regulation and function of CMTR1-dependent mRNA cap methylation. Wiley Interdiscip. Rev. RNA *8*, e1450.

Ivanova, I., Much, C., Di Giacomo, M., Azzi, C., Morgan, M., Moreira, P.N., Monahan, J., Carrieri, C., Enright, A.J., and O'Carroll, D. (2017). The RNA m(6)A Reader YTHDF2 Is Essential for the Post-transcriptional Regulation of the Maternal Transcriptome and Oocyte Competence. Mol. Cell *67*, 1059–1067.e1054.

Jain, D., Puno, M.R., Meydan, C., Lailler, N., Mason, C.E., Lima, C.D., Anderson, K.V., and Keeney, S. (2018). *ketu* mutant mice uncover an essential meiotic function for the ancient RNA helicase YTHDC2. eLife *7*, e30919.

Jiao, X., Doamekpor, S.K., Bird, J.G., Nickels, B.E., Tong, L., Hart, R.P., and Kiledjian, M. (2017). 5′ End Nicotinamide Adenine Dinucleotide Cap in Human Cells Promotes RNA Decay through DXO-Mediated deNADding. Cell *168*, 1015–1027.e1010.

Kasowitz, S.D., Ma, J., Anderson, S.J., Leu, N.A., Xu, Y., Gregory, B.D., Schultz, R.M., and Wang, P.J. (2018). Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. PLoS Genet. *14*, e1007412.

Ke, S., Pandya-Jones, A., Saito, Y., Fak, J.J., Vågbø, C.B., Geula, S., Hanna, J.H., Black, D.L., Darnell, J.E., Jr., and Darnell, R.B. (2017). m^6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. Genes Dev. *31*, 990–1006.

Keith, J.M., Ensinger, M.J., and Moss, B. (1978). HeLa cell RNA (2′-O-methyladenosine-N6-)-methyltransferase specific for the capped 5′-end of messenger RNA. J. Biol. Chem. *253*, 5033–5039.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97.

Lasda, E.L., Kuersten, S., and Blumenthal, T. (2011). SL trans-splicing in a Caenorhabditis elegans in vitro extract. Cold Spring Harb Protoc. *2011*, pdb.prot5574.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat. Chem. Biol. *10*, 93–95.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Mair, G., Ullu, E., and Tschudi, C. (2000). Cotranscriptional cap 4 formation on the Trypanosoma brucei spliced leader RNA. J. Biol. Chem. *275*, 28994–28999.

Mao, X., Schwer, B., and Shuman, S. (1995). Yeast mRNA cap methyltransferase is a 50-kilodalton protein encoded by an essential gene. Mol. Cell. Biol. *15*, 4167–4174.

Mauer, J., Luo, X., Blanjoie, A., Jiao, X., Grozhik, A.V., Patil, D.P., Linder, B., Pickering, B.F., Vasseur, J.J., Chen, Q., et al. (2017). Reversible methylation of m^6A_m in the 5′ cap controls mRNA stability. Nature *541*, 371–375.

Mauer, J., Sindelar, M., Despic, V., Guez, T., Hawley, B.R., Vasseur, J.J., Rentmeister, A., Gross, S.S., Pellizzoni, L., Debart, F., et al. (2019). FTO controls reversible m^6Am RNA methylation during snRNA biogenesis. Nat. Chem. Biol. *15*, 340–347.

Mendel, M., Chen, K.M., Homolka, D., Gos, P., Pandey, R.R., McCarthy, A.A., and Pillai, R.S. (2018). Methylation of Structured RNA by the m(6)A Writer METTL16 Is Essential for Mouse Embryonic Development. Mol. Cell *71*, 986–1000.e1011.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell *149*, 1635–1646.

Muthukrishnan, S., Filipowicz, W., Sierra, J.M., Both, G.W., Shatkin, A.J., and Ochoa, S. (1975). mRNA methylation and protein synthesis in extracts from embryos of brine shrimp, Artemia salina. J. Biol. Chem. *250*, 9336–9341.

Oakberg, E.F. (1956). Duration of spermatogenesis in the mouse and timing of stages of the cycle of the seminiferous epithelium. Am. J. Anat. *99*, 507–516.

Patil, D.P., Pickering, B.F., and Jaffrey, S.R. (2018). Reading m^6A in the Transcriptome: m^6A-Binding Proteins. Trends Cell Biol. *28*, 113–127.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.

Pendleton, K.E., Chen, B., Liu, K., Hunter, O.V., Xie, Y., Tu, B.P., and Conrad, N.K. (2017). The U6 snRNA m(6)A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. Cell *169*, 824–835.e814.

Perry, R.P., Kelley, D.E., Friderici, K., and Rottman, F. (1975). The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5′ terminus. Cell *4*, 387–394.

Pink, R.C., Wicks, K., Caley, D.P., Punch, E.K., Jacobs, L., and Carter, D.R. (2011). Pseudogenes: pseudo-functional or key regulators in health and disease? RNA *17*, 792–798.

Port, F., Chen, H.M., Lee, T., and Bullock, S.L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proc. Natl. Acad. Sci. USA *111*, E2967–E2976.

R Core Team (2017). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

Roignant, J.Y., and Soller, M. (2017). m^{6}A in mRNA: An Ancient Mechanism for Fine-Tuning Gene Expression. Trends Genet. *33*, 380–390.

Schibler, U., Kelley, D.E., and Perry, R.P. (1977). Comparison of methylated sequences in messenger RNA and heterogeneous nuclear RNA from mouse L cells. J. Mol. Biol. *115*, 695–714.

Schwartz, S., Agarwala, S.D., Mumbach, M.R., Jovanovic, M., Mertins, P., Shishkin, A., Tabach, Y., Mikkelsen, T.S., Satija, R., Ruvkun, G., et al. (2013). High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. Cell *155*, 1409–1421.

Sendinc, E., Valle-Garcia, D., Dhall, A., Chen, H., Henriques, T., Navarrete-Perea, J., Sheng, W., Gygi, S.P., Adelman, K., and Shi, Y. (2019). PCIF1 Catalyzes m6Am mRNA Methylation to Regulate Gene Expression. Mol. Cell *75*, 620–630.e629.

Shafer, B., Chu, C., and Shatkin, A.J. (2005). Human mRNA cap methyltransferase: alternative nuclear localization signal motifs ensure nuclear localization required for viability. Mol. Cell. Biol. *25*, 2644–2649.

Shatkin, A.J. (1976). Capping of eucaryotic mRNAs. Cell *9*, 645–653.

Shatkin, A.J., and Manley, J.L. (2000). The ends of the affair: capping and polyadenylation. Nat. Struct. Biol. *7*, 838–842.

Śledź, P., and Jinek, M. (2016). Structural insights into the molecular mechanism of the m(6)A writer complex. eLife *5*, e18434.

Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. *4*, 1521.

Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthès, P., Kokkinaki, M., Nef, S., Gnirke, A., et al. (2013). Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep. *3*, 2179–2190.

Sun, H., Zhang, M., Li, K., Bai, D., and Yi, C. (2019). Cap-specific, terminal N^6-methylation by a mammalian m^{6}Am methyltransferase. Cell Res. *29*, 80–82.

Verdecia, M.A., Bowman, M.E., Lu, K.P., Hunter, T., and Noel, J.P. (2000). Structural basis for phosphoserine-proline recognition by group IV WW domains. Nat. Struct. Biol. *7*, 639–643.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120.

Wang, P., Doxtader, K.A., and Nam, Y. (2016). Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases. Mol. Cell *63*, 306–317.

Warda, A.S., Kretschmer, J., Hackert, P., Lenz, C., Urlaub, H., Höbartner, C., Sloan, K.E., and Bohnsack, M.T. (2017). Human METTL16 is a N^6-methyladenosine (m^{6}A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. EMBO Rep. *18*, 2004–2014.

Wei, C., Gershowitz, A., and Moss, B. (1975a). N6, O2′-dimethyladenosine a novel methylated ribonucleoside next to the 5′ terminal of animal cell and virus mRNAs. Nature *257*, 251–253.

Wei, C.M., Gershowitz, A., and Moss, B. (1975b). Methylated nucleotides block 5′ terminus of HeLa cell messenger RNA. Cell *4*, 379–386.

Wei, J., Liu, F., Lu, Z., Fei, Q., Ai, Y., He, P.C., Shi, H., Cui, X., Su, R., Klungland, A., et al. (2018). Differential m(6)A, m(6)Am, and m(1)A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. Mol. Cell *71*, 973–985.e975.

Werner, M., Purta, E., Kaminska, K.H., Cymerman, I.A., Campbell, D.A., Mittra, B., Zamudio, J.R., Sturm, N.R., Jaworski, J., and Bujnicki, J.M. (2011). 2′-O-ribose methylation of cap2 in human: function and evolution in a horizontally mobile family. Nucleic Acids Res. *39*, 4756–4768.

Wojtas, M.N., Pandey, R.R., Mendel, M., Homolka, D., Sachidanandam, R., and Pillai, R.S. (2017). Regulation of m6A Transcripts by the 3→5′ RNA Helicase YTHDC2 Is Essential for a Successful Meiotic Program in the Mammalian Germline. Mol. Cell *68*, 374–387.e312.

Wu, H., Li, L., Chen, K.M., Homolka, D., Gos, P., Fleury-Olela, F., McCarthy, A.A., and Pillai, R.S. (2019). Decapping Enzyme NUDT12 Partners with BLMH for Cytoplasmic Surveillance of NAD-Capped RNAs. Cell. Rep. *29*, 4422–4434.e4413.

Yamashita, R., Suzuki, Y., Wakaguri, H., Tsuritani, K., Nakai, K., and Sugano, S. (2006). DBTSS: DataBase of Human Transcription Start Sites, progress report 2006. Nucleic Acids Res. *34*, D86–D89.

Yano, Y., Saito, R., Yoshida, N., Yoshiki, A., Wynshaw-Boris, A., Tomita, M., and Hirotsune, S. (2004). A new role for expressed pseudogenes as ncRNA: regulation of mRNA stability of its homologous coding gene. J. Mol. Med. (Berl.) *82*, 414–422.

Yunokuchi, I., Fan, H., Iwamoto, Y., Araki, C., Yuda, M., Umemura, H., Harada, F., Ohkuma, Y., and Hirose, Y. (2009). Prolyl isomerase Pin1 shares functional similarity with phosphorylated CTD interacting factor PCIF1 in vertebrate cells. Genes Cells *14*, 1105–1118.

Zamudio, J.R., Mittra, B., Zeiner, G.M., Feder, M., Bujnicki, J.M., Sturm, N.R., and Campbell, D.A. (2006). Complete cap 4 formation is not required for viability in Trypanosoma brucei. Eukaryot. Cell *5*, 905–915.

Zhang, Z., and Gerstein, M. (2004). Large-scale analysis of pseudogenes in the human genome. Curr. Opin. Genet. Dev. *14*, 328–335.

Zheng, D., Zhang, Z., Harrison, P.M., Karro, J., Carriero, N., and Gerstein, M. (2005). Integrated pseudogene annotation for human chromosome 22: evidence for transcription. J. Mol. Biol. *349*, 27–45.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
For detecting mouse proteins		
Rabbit anti-PCIF1 (for mouse and human)	Bethyl Laboratories	Cat. No. A304-711A; RRID: AB_2620906
rabbit anti-PCIF (for mouse)	Sigma-Aldrich	Cat. No. HPA049517; RRID: AB_2680801
mouse anti-PLZF	Santa Cruz	Cat. No. SC-28319; RRID: AB_2218941
Rabbit anti-PARK7	ThermoFischer Scientitic	Cat. No. PA5-13404; RRID: AB_2160112
Rabbit anti-TUBULIN	Cell Signaling Technology	Cat. No. 2148; RRID: AB_2288042
Anti-m^{6}A antibody	Synaptic Systems	Cat. no. 202003; RRID: AB_2279214
For detecting fly proteins		
Rabbit anti-dPcif1 (CG11399)	This study	N/A
Rabbit anti-Actin	Santa Cruz Biotech	Cat. No. 1616-R
mouse anti-Fibrillarin	AbCam	Cat. No. ab4566; RRID: AB_304523
mouse anti-FLAG	Sigma	Cat. No. F3165; RRID: AB_259529
anti-rabbit IgG HRP-linked	GE Healthcare	Cat. No. NA934; RRID: AB_772206
anti-mouse IgG HRP-linked	GE Healthcare	Cat. No. NA931; RRID: AB_772210
Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Invitrogen	Cat. No. A11034; RRID: AB_2576217
Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 594	Invitrogen	Cat. No. A11037; RRID: AB_2534095
Goat anti-Mouse IgG (H+L), Superclonal Recombinant Secondary Antibody, Alexa Fluor 488	Invitrogen	Cat. No. A28175; RRID: AB_2536161
Goat anti-Mouse IgG (H+L) Cross-Adsorbed ReadyProbes Secondary Antibody, Alexa Fluor 594	Invitrogen	Cat. No. R37121; RRID: AB_2556549
anti-mouse conjugated to Cyanine 3	Jackson ImmunoResearch	Cat. No. 111165144; RRID: AB_2338006
Bacterial and Virus Strains		
DH10EMBacY bacterial strain	Bieniossek et al., 2012	N/A
Top10	This study	
BL21(DE3)	This study	
Chemicals, Peptides, and Recombinant Proteins		
Sodium deoxycholate	Sigma	Cat. No. 30968
Complete EDTA-free protease inhibitor	Roche	Cat. No. 11 873 580 001
m7G(5′)ppp(5′)G RNA Cap Structure Analog	NEB	Cat. No. S1404S
m7G(5′)ppp(5′)A RNA Cap Structure Analog	NEB	Cat. No. S1405S
m7G(5′)ppp(5′)AmpG RNA Cap Structure Analog	Trilink	Cat. No. N-7413
Ponceau S	Sigma	P3504
Schneider's *Drosophila* Medium	GIBCO	Cat. No. 21720001
fetal calf serum	BioConcept	Cat. No. 2-01F30-I
30% acrylamide (37.5:1)	National Diagnostic	Cat. No. EC-890
N,N,N',N'-Tetramethylethylendiamin	Merck	Cat. No. 1107320100
Tween20	SIGMA	Cat. No. P7949
Amersham Prime Western Blotting Detection Reagent	GE Healthcare	Cat. No. RPN2232
SuperSignal West Femto Maximum Sensitivity Substrate	ThermoFisher	Cat. No. 34095
Pierce ECL 2 Substrate	ThermoFisher	Cat. No. 1896433A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays/Kits		
MinElute Gel Extraction Kit	QIAGEN	Cat. No. 28604
MEGAshortscript T7 Transcription Kit	Life technologies	Cat. No. AM1354
Pierce Detergent Compatible Bradford Assay Kit	ThermoFisher	Cat. No. 23246
SuperScript III Reverse Transcriptase kit	ThermoFisher	Cat. No. 18080085
The TruSeq Stranded Total RNA kit	Illumina	Cat. No. RS-122-2301
Deposited Data		
Deep sequencing datasets	This study	GEO: GSE151229
All raw gel data are deposited at Mendeley Data.	This study	https://doi.org/10.17632/jvfp8z5n8x.1
Experimental Models: Cell Lines		
Sf9 insect cells for protein production	Eukaryotic Expression Facility, EMBL Grenoble, France	N/A
High Five insect cells for protein production	Eukaryotic Expression Facility, EMBL Grenoble, France	N/A
Experimental Models: Organisms/Strains		
Mouse: *Pcif1* knockout	This study	Available from Lead Contact
Mouse: *Pcif1-3xFLAG* knockin	This study	Available from Lead Contact
Oligonucleotides		
DNA and RNA oligos		See Table S1
Recombinant DNA		
pACEBac2	Bieniossek et al., 2012	N/A
Human *Pcif1* cDNA	This study	GenBank: BAC45238
Drosophila Pcif1 cDNA	This study	Fly Base: CG11399-PB
Trypanosoma Pcif1	This study	GenBank: RNF19590.1
pIDK	Bieniossek et al., 2012	N/A
pU6-BbsI-chiRNA		
pAC5.1-EGFP		
pCFD4 vector	Port et al., 2014	
Software and Algorithms		
ENRICHR	Chen et al., 2013; Kuleshov et al., 2016	http://amp.pharm.mssm.edu/Enrichr/
R	R Core Team, 2017	https://www.r-project.org
DESeq2	Love et al., 2014	N/A
Bioconductor	Huber et al., 2015	https://www.bioconductor.org/
Salmon	Patro et al., 2017	N/A
tximport	Soneson et al., 2015	N/A
EnhancedVolcano	Blighe, 2019	https://github.com/kevinblighe/EnhancedVolcano
STAR	Dobin et al., 2013	https://github.com/alexdobin/STAR
targetFinder		http://targetfinder.flycrispr.neuro.brown.edu/
Other		
Chelating Sepharose Fast Flow beads	GE Healthcare	Cat. No. 17-0575-01
StrepTrap HP	GE Healthcare	Cat. No. 28-9075-46
Superdex S75 10/300 GL	GE Healthcare	Cat. No. 17-5174-01
Superdex 200 10/300 GL	GE Healthcare	Cat. No. 17-5175-01
MethaPhor agarose	Lonza	Cat. No. 50180
Amersham Hyperfilm ECL	GE Healthcare	Cat. No. 28906837
Amersham Protran 0.45 μm nitrocellulose membrane	GE Healthcare	Cat. No. 10600002

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Ramesh S. Pillai (ramesh.pillai@unige.ch).

Materials Availability

All unique reagents (plasmids, antibodies, and animal mutants) generated in this study are available from the Lead Contact without any restriction. This includes the *Pcif1* mouse knockout and *Pcif1-3xFLAG* knockin lines.

Data and Code Availability

Deep sequencing data generated in this study are deposited with Gene Expression Omnibus (GEO) under accession number GEO: GSE151229. Code used in the current study is available from the Lead Contact upon reasonable request. Other raw data associated with this study are deposited with Mendeley Data (https://doi.org/10.17632/jvfp8z5n8x.1).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animal Work

Mutant mice were generated at the Transgenic Mouse Facility of University of Geneva. The mice were bred in the Animal Facility of Sciences III, University of Geneva. The use of animals in research at the University of Geneva is regulated by the Animal Welfare Federal Law (LPA 2005), the Animal Welfare Ordinance (OPAn 2008) and the Animal Experimentation Ordinance (OEXA 2010). The Swiss legislation respects the Directive 2010/63/EU of the European Union. Any project involving animals has to be approved by the Direction Générale de la Santé and the official ethics committee of the Canton of Geneva, performing a harm-benefit analysis of the project. Animals are treated with respect based on the 3Rs principle in the animal care facility of the University of Geneva. We use the lowest number of animals needed to conduct our specific research project. Discomfort, distress, pain, and injury is limited to what is indispensable and anesthesia and analgesia is provided when necessary. Daily care and maintenance are ensured by fully trained and certified staff. This work was approved by the Canton of Geneva (GE/16/19).

Pcif1 knockout and Flag-Pcif1 knockin mice

This study concerns the *Pcif1* gene (MGI: 2443858; NCBI Gene: 228866) encoding for Phosphorylated CTD-interacting factor 1 (PCIF1) (Fan et al., 2003; Hirose et al., 2008), which was later shown to be a cap-dependent RNA methyltransferase (and termed as CAPAM) (Akichika et al., 2019). The *Pcif1* gene studied here is located on mouse chromosome 2 (Chr2: 164879368-164894454 bp, + strand) and consists of 17 exons (Figure S1A). It encodes for a protein with 723 aa (NCBI: NP_001361055).

We targeted the *Pcif1* locus in mouse embryos of the B6D2F1/J hybrid line (also called B6D2; The Jackson Laboratory, stock no. 100006). It is a cross between C57BL/6J (B6) and DBA/2J (D2), and heterozygous for all B6 and D2 alleles. Single-cell mouse embryos were injected with two different guide RNAs (gRNAs) that direct the DNA endonuclease Cas9 to delete a region between exon 15 and intron 16 (Figure S1A). Founder mice were identified by genotyping PCR (Figure S1B) and crossed with wild-type C57BL/6J (Janvier) partners to obtain germline transmission. We identified a mouse line with 89 bp deletion in the exon 15/intron 16 that removes the critical catalytic residues (NPPF) within the methyltransferase domain and also creates a premature termination codon. Indeed, Western analysis confirms the homozygous *Pcif1* mutant mice lack the full-length PCIF1 protein (Figures 1C and S1E). The antibody used for the western blot detects the C terminus of the PCIF1 protein. Heterozygous $Pcif1^{+/-}$ and homozygous $Pcif1^{-/-}$ mice of both sexes were viable and fertile. We did not find any obvious phenotypic alterations in the homozygous $Pcif1^{-/-}$ mice under the tested normal conditions. However, $Pcif1^{-/-}$ mice display reduced body-weight compared to heterozygous littermates (Figure 1G).

To create an allele expressing C-terminal FLAG-tagged PCIF1, we inserted a sequence corresponding to a four-amino acid flexible linker (SGGG) followed by the 3xFLAG tag (MDYKDHDGDYKDHDIDYKDDD) upstream of the translation stop codon (TAA) of the *Pcif1* open-reading frame (ORF). To this end, we used a single guide RNA to target the Cas9 endonuclease to the *Pcif1* gene locus in mouse embryos of the B6D2F1/J hybrid line (also called B6D2; The Jackson Laboratory, stock no. 100006) (Figures S1A and S1B). We used an ssDNA oligonucleotide with the required 3xFLAG sequence as the repair template (Table S1). Heterozygous and homozygous *Pcif1-3xFLAG* knockin mice were viable and fertile, without any impact on the PCIF1 protein. Based on the western blot analysis, PCIF1-3xFLAG expression (Figure 1A) was identical to that detected for endogenous PCIF1 (Figure S1C) in multiple mouse tissues.

Synthesis and annealing of gRNA

The gRNAs were commercially synthesized [Integrated DNA Technologies (IDT)] in two parts (Table S1). A tracrRNA (67 bases long; IDT, Cat. No. 1072533) which is the common part of the gRNA that binds to the Cas9 protein, and crRNA (IDT) that has a

complementarity of 20 bases with the genomic DNA sequence of interest plus common additional bases for pairing with tracrRNA at the 3′ end (GUUUUAGAGCUAUGCU). The complete gRNA was obtained by annealing 200 pmol each of tracrRNA and crRNA in 1xTE buffer (IDT 1X TE buffer, pH 7.5; Cat. No. 11-05-01-05) to a final volume of 10 uL. The gRNAs mix was heated at 95°C for 5 min and then allowed to cool at room temperature for 30 min to get annealed functional gRNA. The annealed gRNA was stored at −20°C until used for injection into mouse embryos.

Preparation of injection mix

Just before injection into mouse embryos, the annealed gRNAs (final concentration 0.6 pmol/μL) were mixed with the Cas9 protein at 30 ng/μL final concentration (Cat. No. 1081058, IDT) in a final volume of 9 μL. The mix was incubated at RT for 10 min for complex formation. For knockin mouse generation a ssDNA repair template (IDT; 20 ng/μL final concentration) was also added to the injection mix and volume was adjusted with 1x TE buffer (pH 7.5) to 100 μL. The injection mix was centrifuged at 13000 rpm for 5 min at 4°C, and 50 μL of supernatant was collected and stored on ice.

Injection of mouse embryos of the hybrid background B6D2F1/J (black coat color) was carried out at the Transgenic Mouse Core Facility, University Medical Centre (CMU), University of Geneva. The NMRI (Naval Medical Research Institute) mice, which have a white coat color were used as foster mothers.

crRNA: /Alt1/NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAUGCU /Alt2/

N represent the gene-specific sequence. The sequence of crRNAs used are provided (Table S1).

The sequence of the ssDNA repair template for 3xFLAG sequence insertion is as follows:

CAAGGACCGGGACTCAGGCCGGGAACAGGGCCCTAGTAGAGAGCCTCACCCCACT*TCAGGCGGTGGA***ATG**GACTACAAAGA
CCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGAT**TAA**CATATCCTGCGGGGAGGAGGAGCTCCAGGGGTG
CTGTCACAGACTGCTAGGACTCAGCCTC**TAA** is the natural STOP sequence in the mouse PCIF1 coding sequence. The sequence (underlined) is the coding region for 3xFLAG sequence. The sequence in italics is the linker between the PCIF1 protein and the C-terminal 3xFLAG tag.

Genotyping

Ear-punches of the weaned animals (at 21 days after birth) were digested in 100 μL of buffer containing 10 mM NaOH, 0.1 mM EDTA for 120 min at 95°C. The digested ear-punch was centrifuged at 3000 rpm for 10 min and 50 μL of the supernatant was transferred to a new tube containing 50 μL of TE buffer (20 mM Tris-HCl, pH 8.0 and 0.1 mM EDTA). Primers (RRoligo1129 and RRoligo1130) flanking the targeted deletion region in the mouse *Pcif1* gene were used to genotype the knockout allele (Table S1). The expected sizes of PCR products are: $Pcif1^{+/+}$ (362 bp, WT) and $Pcif1^{-/-}$ (273 bp, knockout). For genotyping *Pcif1-3xFlag* knockin allele the primers RRoligo1132 and RRoligo1133 were used (Table S1). The expected sizes of PCR products are: 399 bp for the wild-type allele and 474 bp for the *Pcif1-3xFlag* allele. Identity of the bands was confirmed by Sanger sequencing.

Reaction mix for 20 μL PCR reactions: 10 μL of Phire Green Hot Start II PCR Master Mix (Thermo Scientific, Cat. No. F126L), 1.0 μL primer mix (stock 10 nM each), 2.0 μL ear-punch DNA (50-200 ng), and 7.0 μL water to make 20 μL final volume. Reactions were run using the following conditions: 98°C for 30 s, 35 cycles of [98°C for 5 s, 63°C for 5 s and 72°C for 10 s], 72°C for 1 min, and finally at 4°C to hold the reaction. PCR products were directly analyzed by electrophoresis on 2% agarose gel in 1xTBE buffer (Figure S1B). Primer sequences are given in Table S1.

METHOD DETAILS

Clones and constructs

The mouse PCIF1 protein depicted in the cartoons is based on the sequence from NCBI (NCBI: NP_001361055). Coding sequences for full-length (FL) human PCIF1 (704 aa; GenBank: BAC45238) was amplified from human HeLa cell total RNA by reverse transcription-PCR (RT-PCR). The *Drosophila melanogaster* Pcif1 (920 aa; Fly Base: CG11399-PB) was amplified from fly ovary total RNA by reverse transcription-PCR (RT-PCR). Coding sequence for full-length *Trypanosoma cruzi* Pcif1 (TrypPcif1; 532 aa; GenBank: RNF19590.1) was compared to protein sequences in the Kinetoplastid Genome Resource database TriTRypDB. Unlike the GenBank sequence (which has Proline at position 527), eleven strains of *T. cruzi* showed a serine at position 527. So for our study, we used the TrypPcif1 sequence from (GenBank: RNF19590.1), but carrying the P527S change in the coding sequence. The sequence was synthesized (ShineGene Bio-Technologies, Inc, Shanghai) after codon-optimization for insect cell expression.

Constructs for *Drosophila* S2 cell expression

For expression in *Drosophila* S2 cells, the required coding sequences were cloned into the expression vector pAC5.1-3xFLAG-HA vector at the XhoI-NotI sites. The pAC5.1-3xFLAG vector was prepared in-house by insertion of a multiple cloning site (MCS) cassette followed by a 3xFLAG tag at the site KpnI-XbaI of the pAC5.1vector, to allow expression of C-terminal 3xFLAG-HA fusions.

Sequence of the C-terminal 3xFLAG-HA tag is as follows:

ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGATTACAAGGATGACGATGACAAG*GGCGGCAGCGGC*TA
CCCATATGATGTTCCAGATTACGCTGGT**TGA**

The linker between the 3xFLAG and the HA sequence is in italics. **TGA** is the stop codon.

Constructs for recombinant protein production in prokaryotic expression systems

For the production of full-length human PCIF1 protein in *E.coli*, the coding sequence (UniProtKB: Q9H4Z3) was cloned into pETM30 vector to express the N-terminal 6xHis-GST-PCIF1 fusion protein. For expression of isolated WW domains, sequences coding for human PCIF1 WW domain (42-80 aa of NCBI ID: NP_071387.1) and fly WW domain (138-172 aa of FlyBase ID: FBgn0037021) were cloned in-frame downstream of 6xHis-GST tag in the pETM-30 vector (EMBL Protein Expression Facility, Heidelberg), in order to express the 6xHis-GST-WW fusion protein.

Constructs for recombinant protein production in insect expression systems

For insect cell expression of fly Pcif1 (dPcif1), the coding sequence (UniProtKB: Q9VPB7) was cloned into pACEBac2S vector for the expression as N-terminal His-Strep-SUMO fusion protein. The constructs used for recombinant protein production were verified by restriction digest, as well as by Sanger sequencing. The coding sequence for *Trypanosoma cruzi* Pcif1 (GenBank: RNF19590.1) was codon-optimized for insect cell expression and synthesized (ShineGene Bio-Technologies, Inc, Shanghai) with a P527S mutation (see note above), and was cloned into pACEBac2S vector for protein expression in insect cell.

Recombinant protein production

Production of full-length fly and *Trypanosoma* Pcif1 was carried out in insect cell lines using the baculovirus expression system. The insect ovary-derived cell lines used are: High Five (Hi5) insect cell line originating from the cabbage looper (*Trichoplusia ni*) and the *Sf*9 cells derived from the fall army worm *Spodoptera frugiperda*. Briefly, the full-length fly Pcif1 coding sequence was cloned into pACEBac2S vector. Plasmids carrying target gene were transformed into DH10EMBacY competent cells for recombination with the baculovirus genomic DNA (bacmid). The bacmid DNA was extracted and transfected with FuGENE HD (Promega, cat. no. E231A) into the *Sf*9 insect cells for virus production. The supernatant (V_0) containing the recombinant baculovirus was collected after 48 to 72 h post-transfection. To expand the virus pool, 3.0 mL of the V_0 virus stock was added into 25 mL of *Sf*9 (0.5×10^6/mL) cells. The resulting cell culture supernatant (V_1) was collected 24 h post-proliferation arrest. For large-scale expression of the protein, Hi5 cells were infected with virus (V_1) and cells were harvested 72 h after infection.

For human PCIF1 expression pETM30-hPCIF1 plasmids were transformed into the *E. coli* BL21(DE3) strain and grown in LB media. Protein expression was induced by addition of 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) when the culture density reached around 0.8 (OD_{600}). The proteins were then expressed overnight at 18°C following induction. Similarly, the expression of GST-fusions of WW domains (from human and fly PCIF1) was also performed in the prokaryotic system.

Purification of human, fly and Trypanosoma PCIF1

The cells (bacteria or insect cells) were collected by centrifugation and lysed by sonication in lysis buffer: 50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5% Glycerol, 5 mM 2-mercaptoethanol, 40 mM Imidazole and EDTA- free protease inhibitor (Thermo Scientific, Cat. No. PI88666). The lysate was centrifuged at 16000 rpm for 40 min at 4°C. Supernatant was collected in a new tube and incubated with with Ni^{2+} chelating Sepharose FF beads for 2 h in cold room. After 2 h of incubation, the beads were washed by imidazole gradient washing buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5% Glycerol, 5 mM 2-mercaptoethanol, 40 mM Imidazole) and finally bound protein was eluted with 250 mM imidazole in lysis buffer. Subsequently, the N-terminal tag was cleaved by the TEV protease at 4°C overnight in the dialysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM 2-mercaptoethanol). The cleaved tag was removed by a second purification on Nickel beads. The protein was further purified by gel-filtration chromatography with Superdex 200 columns (GE Healthcare, Cat. No. 28-9909-44), in the buffer (25 mM Tris-HCl pH 8.0, 150 mM KCl, 5% glycerol and 1 mM DTT). The pure fractions were verified by SDS-PAGE electrophoresis (Figures S5A and S5F) and used for biochemical assays.

Purification of GST-fused human and fly PCIF1 WW domain

The bacterial cells [*E. coli* BL21(DE3)] expressing GST-fused WW domain from human and fly PCIF1 were collected by centrifugation and lysed by sonication in buffer: 50 mM Tris-HCl, pH 8.0, 400 mM NaCl, 5% Glycerol, 5 mM 2-mercaptoethanol, 0.1% Triton X-100, 5 mM $MgCl_2$, 5μM $ZnCl_2$ and EDTA-free protease inhibitor. Following the pre-wash (three times) of GST conjugated beads (Mecherey-Nagel, Cat. No. 745500.10) with lysis buffer, the lysate is incubated for 3 h at 4°C with shaking. Then the beads were washed five times: alternating with wash buffer (50 mM Tris-HCl, pH 8.0, 400 mM NaCl, 5 mM 2-mercaptoethanol, 5% Glycerol) and high salt wash buffer (50 mM Tris-HCl, pH 8.0, 1000 mM NaCl, 5 mM 2-mercaptoethanol, 5% Glycerol, 0.1% Triton X-100). Finally, the bound protein was eluted with elution buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5 mM 2-mercaptoethanol, 5% Glycerol, 15 mM Glutathione). The protein was further purified by gel-filtration chromatography with Superdex S75 columns (GE Healthcare, Cat. No. 29148721), in the buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM 2-mercaptoethanol, 5% Glycerol, 0.1% Triton X-100, 2mM MgCl2). The pure fractions were verified by SDS-PAGE electrophoresis (Figure S5D) and used for biochemical interaction assays.

ANTIBODIES

Commercial antibodies

Primary antibodies

rabbit anti-PCIF1 (Bethyl Laboratories, Cat. No. A304-711A) recognizing a C-terminal epitope (654-704 aa) was used for Western analysis; rabbit anti-PCIF (Sigma-Aldrich, Cat. No. HPA049517) recognizing an N-terminal epitope (98-188 aa) were used for immunofluorescence; mouse anti-PLZF (Santa Cruz, D-9, Cat. No. SC-28319); mouse anti-FLAG (Sigma, Cat. No. F3165), rabbit anti-PARK7 (ThermoFischer Scientific, Cat. No. PA5-13404) for detecting mouse protein, rabbit anti-TUBULIN (Cell Signaling Technology, Cat. No. 2148) for detecting mouse protein, and rabbit anti-Actin (Santa Cruz Biotech, Cat. No. 1616-R) to detect fly actin protein, were used for Western. For immunofluorescence analysis with *Drosophila* ovaries, mouse anti-Fibrillarin (AbCam, Cat. No. ab4566) were used. Anti-m^{6}A antibody (Synaptic Systems, Cat. no. 202003) were used for m^{6}A-IPseq experiments.

Secondary antibodies

For western blot analyses, the following secondary antibodies conjugated to Horse Radish Peroxidase were used: anti-rabbit IgG HRP-linked (GE Healthcare, Cat. No. NA934), anti-mouse IgG HRP-linked (GE Healthcare, Cat. No. NA931); for immunofluorescence studies, the following secondary antibodies coupled to fluorescent dyes were used: anti-rabbit (Alexa Fluor 488, Cat. No. A11034), anti-rabbit (Alexa Fluor 594, Cat. No. A11037), anti-mouse (Alexa Fluor 488, Cat. No. A28175), anti-mouse (Alexa Fluor 594, Cat. No. R37121), anti-mouse conjugated to Cyanine 3 (1:500; Jackson ImmunoResearch, Cat. No. 111165144).

Antibodies generated for this study

We generated rabbit polyclonal antibodies against fly Pcif1 protein. Two New Zealand White (NZW) rabbits were immunized with soluble antigens (Biotem, France). The antigen used was the purified untagged full-length fly Pcif1 (dPcif1) protein produced in insect cells. For each injection, 1 mg/mL protein was used. After six injections (at day 0, 14, 28, 56, 70 and 89) crude immune serum was collected (at day 39, 67, 82 and 96) and frozen. The crude sera and the affinity-purified antibodies detect a single band in fly ovary and *Drosophila* S2 cell lysates as detected by Western analysis (Figure 4B). The antibodies were also suitable for immunofluorescence studies of fly embryos and fly ovaries (Figures 4D and S4D). Antibodies were purified using recombinant full-length fly Pcif1 (dPcif1) protein.

Antibody purification

To purify antibodies against fly Pcif1 (dPcif1), large amount of recombinant untagged full-length fly Pcif1 (dPcif1) protein was resolved via SDS-PAGE and semi-dry transferred onto a nitrocellulose membrane. After reversible staining with Ponceau S (Sigma, Cat. No. P3504), part of the membrane containing the antigen was cut out and incubated overnight with 2 mL crude immune sera (diluted in 8 mL 1XPBS). After washes (1xPBS), bound antibodies were eluted with 1.0 mL (elute twice with 500 μL each) of low pH solution (0.1 M Glycine pH 2.5, 150 mM NaCl, three times) and immediately neutralized with 100 μL of neutralization buffer (0.5 M Tris-HCl pH 8.8, determine volume required by tests just before use). Antibodies were concentrated with 2.0 mL Amicon Ultra spin column (GE Health, 10 kDa cut-off) and stored in 1x PBS, 50% glycerol at −20°C.

Drosophila S2 cell culture and transfections

Drosophila melanogaster Schneider 2 (S2) cells were derived from a primary culture of late-stage (1d) *Drosophila* embryos. Based on molecular features, they likely have a macrophage-like lineage. Cells were cultured in Schneider's *Drosophila* Medium (GIBCO, Cat. No. 21720001) supplemented with 10% fetal calf serum (FCS) (BioConcept, Cat. No. 2-01F30-I) at 25°C.

For transfections, cells were plated into 60 mm culture dishes overnight and transfected with 1 μg of pAC5.1 vector expressing hPCIF1-3xFLAG-HA tagged proteins. Cells in a separate dish were transfected with pAC5.1-EGFP vector as control for transfection. To deliver the DNA into the cells 5 μL FuGENE HD (Promega, Cat. No. E231A) was used for each transfection, cells were collected after 3 days post-transfection.

For RNA mass spectrometry, S2 cells were transfected with plasmids expressing hPCIF1-3xFLAG-HA tagged proteins as described above, cells were collected after 3 days post-transfection. Total RNA from cells was extracted using TRIzol (Life technology, Cat. No. 15596-026), according to manufacturer instructions. RNA pellet was washed with 80% ethanol, briefly dried, and dissolved in 40 μL RNase-free water. To purify the polyA+ fraction from the total RNA, the samples were incubated with oligo dT Dynabeads (Thermo Fisher, Cat. No. 61006) and eluted in 20 μL of 10 mM Tris-HCl following the protocol suggested by the manufacturer. RNA mass spectrometry confirmed the absence of m^{6}Am RNA modification in total RNA from S2 cells (Figure 4C). When S2 cells were transfected with 3Xflag-HA-tagged human PCIF1 (Figure S4A), ectopic methylation of RNA with m^{6}Am was confirmed (Figure 4C).

Mouse multiple tissue western blot

Multiple tissues were isolated from an adult (> 60 d) mouse. After flash-freezing in liquid nitrogen, a piece of different tissues was homogenized in 1 mL lysis buffer [20 mM Tris pH 7.4, 150 mM NaCl, 0.5% Triton X-100, 0.5% sodium deoxycholate, 1 mM DTT, Complete Protease Inhibitor Cocktail Tablet (Roche, Cat. No. 5056489001). The lysate was transferred to a 1.5 mL Eppendorf tube, centrifuged at 14000xg for 30 min, and the supernatant collected. An aliquot was taken to measure the concentration by Pierce

Detergent Compatible Bradford Assay Kit (Thermo Fisher, Cat. No. 23246). Lysate concentrations were normalized to 1 mg/μL. The SDS loading buffer was added and the samples were boiled at 95°C for 10 min. From each tissue sample, 20 mg of protein per lane was loaded and resolved by SDS-PAGE. Western blot analyses (Figures 1A and S1C) with anti-FLAG or anti-PCIF1 rabbit polyclonal antibodies (Bethyl Laboratories, Cat. No. A304-711A; recognizes a region between 654-704 aa) were performed. After stripping, the blot was re-probed with anti-PARK7 rabbit polyclonal antibodies, to serve as loading control.

The expected size of ~90 kDa is seen for mouse PCIF1 in all tissues, while in some tissues we consistently observed a second band at ~60 kDa (Figures 1A and S1C). Since the anti-PCIF1 antibody recognizes C-terminal epitopes, and since the tag in our PCIF1-FLAG knockin mouse is C-terminal, we conclude that the ~60 kDa band is likely to be an N-terminal truncated version.

Western Blot

Whole cell lysates and suspensions of boiled beads were separated on SDS-PAGE in order to detect proteins of interest. SDS-PAGE gels were prepared using Ultra-Pure ProtoGel 30% acrylamide (37.5:1) (National Diagnostic, Cat. No. EC-890), ultra-pure water, and resolving gel buffer (0.375 M Tris, 0.1% SDS, pH 8.8) to obtain 8% resolving gel, and with stacking gel buffer (0.125 M Tris, 0.1% SDS, pH 6.8) to obtain 12% stacking gel. N,N,N',N'-Tetramethylethylendiamin (TEMED; Merck, Cat. No. 1107320100) and 10% ammonium persulfate were added to catalyze the polymerization reaction. Gel electrophoresis was performed at 90 V for 30 min. and then at 120 V for 90 min. After separation, proteins from the gel were transferred to the Amersham Protran 0.45 μm nitrocellulose membrane (GE Healthcare, Cat. No. 10600002) overnight at 5 V at room temperature using Trans-Blot SD. Semi-Dry Transfer Cell system (Bio-Rad, Cat. No. 1703940). After transfer, membranes were washed with Tris-buffered saline (TBS, 20 mM Tris, 150 mM NaCl, pH 7.6) and blocked for 1 h at room temperature with 5% fat free milk in TBS with 0.05% Tween20 (TTBS) (Sigma, Cat. No. P7949). The blocked membranes were incubated with primary antibody for 3 h at room temperature in 5% milk with TTBS. Then, membranes were washed 3 times for 10 min with TTBS and incubated with HRP-conjugated secondary antibody at 1:10 000 dilution, either with anti-rabbit IgG HRP-linked (GE Healthcare, Cat. No. NA934) or anti-mouse IgG HRP-linked (Invitrogen, Cat. No. a27025) for 1 h at room temperature in 5% milk in TTBS. The membranes were washed 3 times for 10 min with TTBS and incubated with one of detection reagents: Amersham Prime Western Blotting Detection Reagent (GE Healthcare, Cat. No. RPN2232), SuperSignal West Femto Maximum Sensitivity Substrate (ThermoFisher, Cat. No. 34095) or Pierce ECL 2 Substrate (ThermoFisher, Cat. No. 1896433A) for 5 min. at room temperature. Signal was detected using Amersham Hyperfilm ECL (GE Healthcare, Cat. No. 28906837). The processed films were scanned using Perfection 3200 Photo scanner (Epson) with XSane image scanning software (ver. 0999).

Immunofluorescence analysis

Mouse tissue sections

To prepare the paraffin sections, the mouse tissues were washed in PBS, and fixed in 4% paraformaldehyde overnight at 4°C. After washing in PBS, testes were dehydrated in 70% ethanol and stored in 70% ethanol at 4°C. Alternatively isolated tissues were fixed in Bouin's solution (Sigma, cat. No HT10132) overnight at room temperature and subsequently washed 3 times in 50% ethanol and in several changes of 70% ethanol, until no yellow dye could be extracted into solution. Samples were sent to Histology core service in University of Geneva, Switzerland where they were further dehydrated in 80%, 90%, 96% and 100% ethanol (90 min for each step), followed by incubation in xylene (3 times 30 min). Xylene was removed and replaced with paraffin and incubated at 56-58°C. Tissues were then transferred into plastic molds (Polysciences mold S-22; NC0397999) filled with melted paraffin, and paraffin was allowed to become solid at room temperature. The tissue sections (~5 μM thickness) were prepared using microtome. The sections were allowed to stretch at 42°C and then stored at room temperature. For histological analysis, the slides containing the paraffin sections were placed in a glass slide holder filled with xylene (3 × 5 min) to remove the paraffin. For rehydration, the slides were incubated in 100% ethanol, 96% ethanol, 70% ethanol, 50% ethanol (2 × 10 min for each step) and miliQ water (2x2 min for each step).

Next, antigen retrieval was performed with Heat-Induced Epitope Retrieval (HIER). Briefly, slides were submerged in 1L of 10 mM Citrate Buffer pH 6.0 (10 mM Sodium citrate, 0.05% TWEEN) and heated in pressure cooker at 1 atmosphere at 120°C for 20 min. Tissues were allowed to cool down at room temperature for at least 1 h, washed twice in PBST (PBS, 0.5% Triton X-100) at RT for 10 min and 2x2 min in miliQ water. Slides were blocked in blocking soultion (5% normal goat serum, PBST) for 1 h at room temperature in a humidified chamber. Primary antibodies were diluted in blocking solution at different concentrations (see below) and incubated overnight at 4°C. Next day, slides were washed in PBST (3 × 5 min), incubated with secondary antibody (anti-mouse or anti-rabbit conjugated to Alexa 488, or 594 fluorophore) in a humidified chamber for 1 h (dilution 1:1000 in blocking solution). Slides were washed twice in PBST and incubated with DAPI (0.5 μg/mL, Bio-Rad, cat. No. 10043282) for 5 min to counterstain the nuclei. Sections were finally washed in PBS (5 min), in miliQ water (5 min) and mounted with ProLong Gold Antifade ReagentDiamond Antifade Mountant (Life technologies, cat. No. S36942P36970). Pictures were taken using Zeiss LSM800, point-scan confocal microscope (Bioimaging Center, University of Geneva). Primary antibodies concentrations: anti-PCIF1 1:100, anti-PLZF 1:100.

Drosophila ovaries

5-10 ovary pairs were dissected from wild-type or heterozygous *dPcif1*$^{F9/+}$ mutant *Drosophila* females in PBS. They were immediately fixed in 4% PFA for 30 min at room temperature (RT). Fixed ovaries were then washed (3 × 5 min) with 1xPBS, permeabilized for 5 min with PBT (1xPBS containing 0.1% Triton X-100) at RT and blocked in blocking buffer (1xPBS, 0.1% Triton X-100, 5% horse

serum) for 1 h at RT. They were incubated with rabbit anti-dPcif1 (1:100) and mouse anti-fibrillarin (1:100; abcam ab4566) antibodies in blocking buffer overnight at 4°C. Next day, they were washed (3 × 15 min) in PBT at RT and incubated for 4 h with anti-rabbit conjugated to Alexa 488 (1:500; Invitrogen, Cat. No. A11029), anti-mouse conjugated to Cyanine 3 (1:500; Jackson ImmunoResearch, Cat. No. 111165144) antibodies and Hoechst, 1:1000 (ThermoFisher, Cat. No. 33342) in blocking buffer at RT. Ovaries were then washed in PBT (3 × 10 min) at RT and stored in PBS before mounting in Dako (Cat. No. S3022). Images were captured with Nikon TI-E Eclipse confocal spinning-disk inverted microscope (CEA, Grenoble).

Drosophila embryos

Thousands of fly embryo (2 to 15 h) were collected in 0.4 mL 1xPBS. The embryos were dechorionated in 50% chlorex for 2 min followed by washing with tap water several times. Embryos were fixed in 2.0 mL of fixing solution (4% formaldehyde, 0.1M Pipes pH 6.9, 2 mM $MgSO_4$, 1mM EGTA). Equal volume of Heptane (2.0 mL) was added to the embryos, mixed by shaking and kept on rotating wheel at room temperature for 15-20 min. After fixing, aqueous phase (lower phase) was removed carefully and equal volume of methanol (equal volume to that of heptane) was added and embryo were shaken vigorously to devitilinize them. Once the embryos were settled in the bottom, heptane, and methanol was removed. Embryos were washed twice with a small volume of methanol (2.0 mL), and once with PBT (1xPBS, 0.1% Triton -X100, and 0.1% BSA). Care should be taken to minimize the exposure of embryos to methanol. Embryos were incubated in PBT buffer for 2 h at RT, followed by incubation with primary antibody (rabbit anti-dPcif1, 1:1000 dilution) in PBT overnight at 4°C on a wheel. Next day, embryos were washed 3x in PBT at RT and finally incubated with secondary anti-rabbit antibody coupled with Alexa Fluor 594 for 2 h at RT in dark. Embryos were washed 2x with PBT followed by incubation with 1:1000 diluted DAPI in PBT for 10 min at RT. After two more washes in PBT, embryos were mounted on glass slides in ProLong Gold Antifade Reagent mounting media (Thermo Scientific, Cat. No. P36931). Images were captured on Zeiss LSM 780 confocal microscope (Bioimaging Center, University of Geneva).

Drosophila S2 cells

Glass coverslips (Sigma, Cat. No. PEZGS0416) were previously coated with Concavalin A (0.5 mg/mL; Sigma, Cat. No. L7647), in order to allow the cells to attach on it. S2 cells were grown in dishes containing such coverslips and the cells were transfected with a plasmid expressing dPcif1-3xFLAG-HA, as previously described. After two days of transfection, the immunostaining protocol was started.

The cells were washed in 1xPBS, and then fixed in 4% paraformaldehyde at room-temperature (RT) for 15 min. The cells were washed 3 times and incubated for 5 min with 0.1M Glycine (MP Biomedicals, Cat. No. 808822). After three more washes, the cells were permeabilized with 0.1% Triton X-100 for 5 min. After the cells are incubated for 15 min in blocking buffer (0.2% BSA, 1% goat serum, 0.1% Triton X-100 in PBS). The cells were incubated for 90 min with the primary antibody (monoclonal mouse anti-FLAG; Sigma, Cat. No. F3165), diluted 1:1000 in blocking buffer. After 3 washes in 1xPBS, the cells were incubated for 60 min. in the dark with the secondary antibody conjugated to Alexa Fluor 488, diluted 1:250 in blocking buffer. Afterward, cell nuclei were stained for 30 min with DAPI 0.5 μg/mL (Bio-Rad, Cat. No. 10043282) diluted in 1xPBS. Finally, the cells were washed 3 times in 1xPBS and once in double-distilled H2O to remove the salts. The coverslip was mounted with Slowfade Gold Antifade Reagent (Life technologies, Cat. No. S36942). Images (Figure S4C) were taken using Zeiss LSM710 confocal microscope (Bioimaging Center, University of Geneva).

Northern blot of U snRNAs

Total RNA was extracted from mouse testes using TRIzol RNA extraction kit (Life technology, Cat. No. 15596-026), further purified with double phenol-chloroform treatment, precipitated in ethanol for at least 20 min at −20 °C and resuspended in milli-Q water. For Northern blotting, 2 μg of total RNA was resolved by 8% urea-PAGE. The RNA was transferred by capillarity to a Nylon membrane (Hybond N+, Amersham) for at least 16 h in in 20 × SSC solution (3 M NaCl, 300 mM sodium citrate). After the transfer, the RNA was UV cross-linked to the membrane using a Stratagene "cross linker" (120 mJ/cm^2 in auto-crosslinking mode). Pre-hybridization was performed for at least 2 h in Church buffer (0.25 M sodium phosphate buffer pH 7.2, 1 mM EDTA, 1% BSA, 7% SDS) at 65°C. Single-stranded DNA oligonucleotide probes (Table S1) were labeled with [γ-^{32}P]ATP (3000 Ci/mmol, 10 mCi/mL, Perkin Elmer) using the T4 polynucleotide kinase (Life technologies, No. EK 0031). The resulting probes were purified on Micro-Spin G25 columns (GE Healthcare, cat. No. 27-5325-01) to remove unincorporated nucleotides, denatured for 5 min at 95°C and incubated with the membrane in 10 mL Church buffer at 65°C overnight. Next day, washing was performed at 50°C as follows: twice 10 min each with Wash buffer-1 (2 × SSC, 0.1% SDS) and twice 10 min each with Wash buffer-2 (0.2 × SSC, 0.1% SDS). The membrane was wrapped in Saranfilm, exposed to Storage Phosphor Screens (GE Healthcare; BAS IP MS 2025 E, Cat no. 28-9564-75) and scanned (GE Healthcare; Typhoon FLA 9500 IP, Cat no. 29-1885-90). The signal of 5S rRNA was used as a loading control. Sequences of the probes are given (Table S1).

Quantification of RNA modifications using LC-MS/MS

RNA was hydrolyzed to ribonucleosides by 20 U benzonase (Santa Cruz Biotech) and 0.2 U nuclease P1 (Sigma) in 10 mM ammonium acetate pH 6.0 and 1 mM magnesium chloride at 40°C for 1 h. After that, ammonium bicarbonate to 50 mM, 0.002 U phosphodiesterase I and 0.1 U alkaline phosphatase (Sigma) were added, and incubated further at 37°C for 1 h. The hydrolysates were mixed with

3 volumes of acetonitrile and centrifuged (16,000xg, 30 min, 4°C). The supernatants were dried and dissolved in 50 μL water for LC-MS/MS analysis of modified and unmodified ribonucleosides. Chromatographic separation was performed using an Agilent 1290 Infinity II UHPLC system with an ZORBAX RRHD Eclipse Plus C18 150 × 2.1 mm ID (1.8 μm) column protected with an ZORBAX RRHD Eclipse Plus C18 5 × 2.1 mm ID (1.8 μm) guard column (Agilent). The mobile phase consisted of water and methanol (both added 0.1% formic acid) run at 0.23 mL/min. For modifications, starting with 5% methanol for 0.5 min followed by a 2.5 min gradient of 5%–15% methanol, a 3 min gradient of 15%–95% methanol and 4 min re-equilibration with 5% methanol. A portion of each sample was diluted for the analysis of unmodified ribonucleosides which was chromatographed isocratically with 20% methanol. Mass spectrometric detection was performed using an Agilent 6495 Triple Quadrupole system operating in positive electrospray ionization mode, monitoring the mass transitions 269.1-150.1 (m^6Am), 282.1-150.1 (m^6A), 282.1-136.1 (Am), 268.1 (A), 284.1-152.1 (G), 244.1-112.1 (C), 245.1-113.1 (U), 296.1-164.1 ($m^6{}_2A$), and 298.1-166.1 (m^7G).

Preparation of m^7G-cap enriched RNA fragment for RNA mass spectrometry

Liver Poly(A)$^+$ RNA (~8 μg) from three biological replicates of both *Pcif1*$^{+/-}$ and *Pcif1*$^{-/-}$ genotypes was purified with mRNA purification kit (Thermo Scientific, Catalog 61006). The poly(A)$^+$ RNA was further purified with RNA Clean & Concentrator-5 (ZYMO Research, R1014) and eluted in 17 μL water. The RNAs were fragmented with RNA fragmentation reagent (Thermo Scientific, catalog no. AM8740) in a final volume of 20 μL at 70°C for 7.5 min in a thermocycler. Fragmentation reaction was stopped by transferring the reaction mix quickly on ice and adding 2.2 μL of STOP solution (supplied with the fragmentation reagent) and gentle mixing. The fragmented RNAs were purified with RNA Clean & Concentrator-5 and eluted in 17 μL water. RNAs were 5′ phosphorylated by adding 2 μL of 10X T4 DNA ligase buffer (NEB) and 20 units of T4 PNK (ThermoScientific) in a final reaction volume of 20 μL. The phosphorylation reaction was carried out for 1.30 h at 37°C. The m^7G-capped RNA fragments were enriched by specifically degrading the 5′-phosphorylated RNA fragments with the Terminator 5′-Phosphate-Dependent Exonuclease (Epibio). To the phosphorylation reaction, 14 μL of water, 4 μL 10x Exonuclease buffer A and 2 μL Terminator exonuclease was added and the reaction mixture was incubated at 30°C for 3 h. RNAs were purified from above reaction with RNA Clean & Concentrator-5 and eluted in 15 μL water. At this point we get 300-400 ng m^7G-capped RNAs fragments from each sample.

The m^7G-cap enriched RNAs were divided into two parts. One part was treated with nuclease P1 (cleaves only phosphodiester bonds, leaving the cap structure linked to the TSS nucleotide) and alkaline phosphatase followed by LC-MS/MS analysis to identify internal (body) RNA modifications in the RNAs, whereas second part was completely digested with a mixture of nucleases [20 U benzonase (Santa Cruz Biotech) and 0.2 U nuclease P1 (Sigma) in 10 mM ammonium acetate pH 6.0 and 1 mM magnesium chloride at 40°C for 1 h] to obtain RNA modification present on the entire fragment in the RNAs. The nuclease P1-digest will have all internal modifications, but lack that of the m^7G cap structure and that on the transcription start site (TSS) nucleotide. The complete digest will have all modifications including that of the cap, TSS nucleotide and body of the RNA.

Preparation of m^7G-capped RNA by *in vitro* transcription

A 40 nt m^7G-capped RNA initiating with an adenosine was generated by *in vitro* transcription as previously described (Jiao et al., 2017). The commonly used T7 class III promoter ϕ6.5 initiates using GTP, while the class II promoter ϕ2.5 initiates with ATP (Coleman et al., 2004). We prepared a single-stranded DNA template containing the T7 class II promoter ϕ2.5 at the 5′ end, with the transcription start site adenosine, followed by residues composed of T, C and G (there is no adenosine except at the TSS position). The template with a G as the TSS nucleotide was prepared using the commonly used T7 class III promoter ϕ6.5. The templates (Wu et al., 2019) were amplified by PCR reaction to generate the dsDNA template for *in vitro* transcription (Table S1).

Sequences of DNA template used are indicated below. Only the top-strand (sense strand) is shown.

CAGTAATACGACTCACTATT**A**GGCCTCTCGCTCTGCTGGGTGTGCGCTTGCTTGGCTTGC.

The ϕ2.5 promoter (underlined) was used for the template with TSS as adenosine (A). The transcription-start site nucleotide is in bold.

CAGTAATACGACTCACTATA**G**GGCCTCTCGCTCTGCTGGGTGTGCGCTTGCTTGGCTTGC

The T7 class III promoter ϕ6.5 was used for template with TSS as guanosine (G).

In vitro transcription reaction was carried out at 37°C for 4 h in a 20 μL reaction containing 6 pmol DNA template, 2 μL 10x reaction buffer, 2 μL of 75 mM CTP and 75 mM UTP, 0.5 μL of 75 mM GTP and 2 μL 75 mM $m^7G(5)ppp(5)G$ (NEB, Cat. No. S1404S) or 75 mM $m^7G(5)ppp(5)A$ (NEB, Cat. No. S1405S) or $m^7G(5)ppp(5)(2'OMeA)pG$ (Trilink, Cat. No. N-7413) and 2 μL T7 enzyme mix (MEGAshortscript Kit; ThermoFisher Scientific). Generated RNAs were purified by phenol-chloroform extraction.

In vitro methylation assays with PCIF1

The m^7G-capped RNAs were prepared by *in vitro* transcription as described above. The *in vitro* methylation assay was performed overnight at 37°C in a reaction mix containing 50 mM HEPES-KOH pH 7.5, 1 mM DTT, 0.02 μCi C^{14}SAM (PerkinElmer, Cat. No. NEC363010UC), 500 pmol RNA substrate and 5 μg recombinant hPCIF1, dPcif1 or TrypPcif1. The next day, the RNAs in the samples were purified by phenol-chloroform extraction and precipitation. 50% of the samples were resolved via 15% UREA-PAGE. The gel was run 1 h at 12 W. After the run, the gel was stained with methylene blue staining to check for the presence of RNA. Afterward the

gel was dried in a dryer (Bio-Rad, model 583) with gradual heating and cooling, 80°C for 1.5 h. The dried gel was exposed to exposed to Storage Phosphor Screens (GE Healthcare; BAS IP MS 2025 E, Cat no. 28-9564-75) for 48 h and scanned (GE Healthcare; Typhoon FLA 9500 IP, Cat no. 29-1885-90).

For RNA mass spectrometry, the m^7G-capped RNAs were prepared by *in vitro* transcription as described above. The *in vitro* methylation assay was performed overnight at 37°C in a reaction mix containing 50 mM HEPES-KOH pH 7.5, 1 mM DTT 2.5 mM SAM (NEB, Cat. No. B9003S), 500 pmol RNA substrate and 5 μg recombinant hPCIF1, dPcif1 or TrypPcif1. The next day, the RNAs in the samples were purified by phenol-chloroform extraction and precipitation, RNA pellet was washed with 80% ethanol, briefly dried, and dissolved in 20 μL RNase-free water.

Phosphorylated CTD-binding assays with recombinant PCIF1

The following N-term biotinylated peptides (Thermo Fisher Scientific) were purchased:

UnmodCTD: Heptad repeats of YSPTSPS, unmodified
Ser5-CTD: Heptad repeats of YSPT[(pS)]PS
Ser2-CTD: Heptad repeats of Y[(pS)]PTSPS

Dynabeads M-280 (Invitrogen, Cat. No. 11205D) were washed in binding buffer (25 mM Tris-HCl pH8.0, 50 mM NaCl, 1 mM DTT, 5% glycerol and 0.03% Triton X-100), and 1 nmol of the biotinylated CTD peptides were immobilized on 0.5 mg of beads in 50 μL of binding buffer. The beads were mixed with 3 μg of the recombinant full-length dPcif1, full-length hPCIF1 or the GST fused WW domains from the respective proteins, and incubated for 30 min on ice. After 3 washes, the beads were resuspended in 20 μL of binding buffer and boiled in 1x Laemmli buffer. The 20% of input protein, 40% of the eluate from the beads control or experimental samples were resolved via 15% SDS-PAGE. The bound full-length PCIF1 proteins were detected by Western analyses. The bound GST-fused WW domains were detected by Silver stain (Invitrogen, Cat. No. LC6070).

Isothermal calorimetry (ITC) experiments

ITC experiments were performed using a MicroCal iTC 200 (Malvern Panalytical) at 25°C. The dPcif1 protein was dialyzed in buffer (20 mM HEPES, pH 7.5, and 100 mM NaCl) overnight, and was titrated with two different RNA cap analogs, m^7G(5′)ppp(5′)G (NEB, Cat. No. S1404S) and m^7G(5′)ppp(5′)A (NEB, Cat. No. S1405S). The sample cell was filled with 13.5 μM dPcif1, and the syringe was filled with 135 μM RNA cap analog. Titrations were carried out by a preliminary 0.4 μL injection followed by constant volume injections (19 injections of 2 μL) with 150 s spacing. Data analysis was performed using Origin software. As for TrypPcif1, the protocol is same as dPcif1, except that the concentration of protein is 30 μM and the concentration of RNA cap analog is 300 μM [m^7G(5)ppp(5)G (NEB, Cat. No. S1404S) or m^7G(5)ppp(5)A (NEB, Cat. No. S1405S) or m^7G(5)ppp(5)(2′OMeA)pG (Trilink, Cat. No. N-7413)]. We thank Andrew McCarthy at EMBL Grenoble, France for access to the instrument.

Preparation of RNA libraries

Total RNA sequencing

Total RNA from biological triplicates of 2 months-old wild-type (WT) and *Pcif1* KO mice were used for the experiment with mouse brain, spleen, and testis samples. The tissues were snap-frozen in liquid nitrogen. Total RNA was isolated from tissues using TRIzol (Thermo Fisher Scientific) according to manufacturer instructions. The TruSeq Stranded Total RNA kit with Ribo-Zero Gold (Illumina, Cat. No. 20020612) was used for library preparation with 400 ng of total RNA as input. Ribodepleted RNAs were subjected to RNA-seq library preparation with NEB kit NEBNext® Ultra II Directional RNA Library Prep Kit (Illumina, Cat. No. E7760), following manufacturer's instruction. Library molarity and quality was assessed with the Qubit and Tapestation using a DNA High sensitivity chip (Agilent Technologies). Libraries were diluted at 2 nM and pooled before the clustering process on a HiSeq 4000 Single Read flow cell. Paired-end reads of 100 bases were generated using the TruSeq SBS reagents on the Illumina HiSeq 4000 sequencer (iGE3 Genomics Platform, University of Geneva).

Ribosome profiling

Brains from three adult wild-type (WT) controls and three *Pcif1* KO mice were isolated and snap-frozen in liquid nitrogen. Frozen brain (~400 mg) was homogenized in 1300 μL (3x weight of the tissue in miligrams) of ice cold polysome lysis buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 5 mM DTT, 100 μg/ml Cycloheximide, 40 U/ml RNasin plus, (Promega), Complete EDTA-free protease inhibitor (Thermo Scientific), 1% TritonX-100, 0.5% Sodium deoxycholate) using a glass homogenizer (6-8 strokes). The lysate was transferred to a 1.5 mL Eppendorf tube and incubated on ice for 10 min with intermittent mixing. The lysate was centrifuged at 3000 rpm for 3 min at 4°C. Supernatant was carefully collected in a new 1.5 mL Eppendorf tube. The OD_{260} of the lysate was measured with a Nanodrop instrument and approximately 15 OD_{260} (which correspond to 100 μL of lysate), was used for ribosome footprint preparation, and 50 μL was directly transferred to TriZol (Ambion) for total input RNA extraction. The remaining lysate was divided in aliquots of 150 μL and was snap frozen in liquid nitrogen and stored in −80°C freezer for later use. For ribosome footprint preparation, 100 μL of the supernatant was treated with 2.5 μL (5 units) of Turbo-DNase (Ambion) and 6.5 μL (650 units) of RNase I at 23°C for 45 min with rotation at 300rmp in a thermoblock. In meantime, the s-400 HR columns (GE healthcare) were washed three

times with 700 μL polysome buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 5 mM DTT, 100 μg/mL Cycloheximide, 40U/mL RNasin plus, (Promega), Complete EDTA-free protease inhibitor (Thermo Scientific)) plus 20U/mL SUPERaseIn (Ambion). During each wash columns with polysome buffer were spun at 2000rpm for 1 min in a cold centrifuge. The columns were stored on ice with the polysome buffer in it (last wash). Once the RNaseI-treatment of lysate completed, 8.7 μL of SUPERaseIn RNase inhibitor was added to the reaction, gently mixed with the help of a pipette and stored on ice. The S-400 HR columns were spun at 2000 rpm for 1 min to remove the polysome buffer. RNaseI-treated lysate was applied on the top of the column and spun at 2400 rpm for 2 min. The flow through was transferred to a fresh Eppendorf tube and RNA was extracted using 1.0 mL TriZol reagent. The RNA contain ribosome protected fragments (RPF) of mRNA as well as ribosomal RNA.

Approximately 20-25 μg of RNaseI-treated RNAs were separated on 15% UREA-PAGE to purify RPF which are in the range of 26-34 nt. Radiolabelled RNA marker of length 28 and 30 nt were used to size select the ribosome protected RNA fragment (RPF). The urea polyacrylamide gel was run at 16 W for 1 h with the samples. Once the bromophenol blue reached the bottom of the gel, the power was switched off and the gel was taken out of the running apparatus. One glass plate was removed and the gel was wrapped in saran wrap and exposed to Storage Phosphor Screens (GE Healthcare; BAS IP MS 2025 E, Cat no. 28-9564-75) for 30 min, and scanned (GE Healthcare; Typhoon FLA 9500 IP, Cat no. 29-1885-90). Based on the position of markers, the gel piece corresponding to approximately 26-34 nt was excised and transfer to an RNase-free Eppendorf tube. The gel piece was ground to a fine powder and 600 μL of RNA extraction buffer (0.5 M Ammonium Acetate, 1 mM EDTA pH 8.0, and 0.1% SDS) was added and the tubes were incubated at 25°C overnight. Next day, the slurry was transferred to filter tubes (from Art-seq kit) and spun for 5 min at 5000 rpm. The flow-through was collected in a new RNase-free Eppendorf tube and precipitated with 1.0 μL of glycogen and 900 μL isopropanol overnight in −20°C freezer. The RNA solution was centrifuged at 13000 rpm for 20 min in a cold centrifuge to pellet RNA. RNA pellet was washed with 80% ethanol once, air-dried and resuspended in 20 μL of RNase-free water.

Ribosomal RNAs were depleted from both input RNAs (400 ng) and RPF RNAs with mouse ribosomal RNAs probe from siTOOLs (mouse riboPOOLs) following their protocol. The ribosomal RNA-depleted input and RPF RNAs were used to make next generation sequencing library with TruSeq Ribo Profile (mammalian) kit from illumina. The libraries were sequenced at Genomics platform (University of Geneva) for 50 cycles.

m^6Am-Exo-Seq

m^6Am-Exo-Seq (Sendinc et al., 2019) is a variation of m^6A-IP seq. In this protocol, first, the m^7G-capped RNA fragments (~100-200 nt) from poly(A)$^+$ RNAs are enriched by multiple enzymatic treatments followed by m^6A IP-seq to get m^6A modification near the 5′ end of the mRNAs.

Approximately 8.0 μg of poly(A)$^+$ RNAs from biological triplicates from *Pcif1*$^{+/-}$ and *Pcif1*$^{-/-}$ mouse testes were fragmented with RNA fragmentation buffer (Ambion) in a total volume of 20 μL for 7.30 min at 70°C in a thermcycler to obtain RNA fragments of ~100-200nt length. Fragmentation reaction was stopped by transferring the RNAs quickly to ice and adding 2.2 μL 10x stop solution (Ambion). The fragmented RNAs were purified with RNA Clean & Concentrator-5 (Zymo Research, R1014) and eluted in 17 μL water. Fragmented RNAs were 5′ phosphorylated by adding 10x, 2 μL T4 DNA ligase buffer (NEB) and 20 units of T4 PNK (ThermoScientific) in a final reaction volume of 20 μL. The phosphorylation reaction was carried out for 1.30 h at 37°C. The 5′ phosphorylated RNA fragments were degraded by Terminator 5′-Phosphate-Dependent Exonuclease (Epibio). To the phosphorylation reaction, 14 μL of water, 4 μL 10x Exonuclease buffer A and 2 μL Terminator exonuclease was added and the reaction mixture was incubated at 30°C for 3 h. RNAs were purified from above reaction with RNA Clean & Concentrator-5 and eluted in 15 μL water. At this point we get 300-400 ng of m^7G-cap-enriched RNA fragments from each sample.

The RNA obtained from above reaction were directly used for m^6A-IP (Wojtas et al., 2017) with some modifications. First, 100 μL of protein A dynabeads (Life Technologies; 10002D) were washed twice with PXL buffer (1X PBS, 0.1% SDS, 0.5% sodium deoxycholate, 0.5% NP-40) and blocked with 1 μg/μL BSA in 400 μL PXL buffer for 45 min. Blocked beads were then incubated with 5 μg of anti-m^6A antibody (Synaptic Systems, Cat. no. 202003) in 200 μL PXL buffer supplemented with 4 μL of RNasin RNase inhibitor (Promega; N2611) for 1 h at room temperature. Antibody-conjugated beads were washed twice with PXL buffer and finally resuspended in 400 μL PXL buffer plus 4 μL RNAsin RNase inhibitor. Fragmented m^7G-cap enriched RNAs were added and beads were incubated at 4°C for 2 h with rotation. 10% of the RNA was kept separately and used to prepare input libraries. After 2 h incubation, the beads were washed twice by ice-cold Nelson low-salt buffer (15 mM Tris at pH 7.5, 5 mM EDTA), once by ice-cold Nelson high-salt buffer (15 mM Tris at pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS, 1 M NaCl), once by ice-cold Nelson stringent wash buffer (15 mM Tris at pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS, 120 mM NaCl, 25 mM KCl), and last by ice-cold NT-2 buffer (50 mM Tris at pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NP-40). Antibody-bound RNAs were eluted by adding 1mL TRIzol (Invitrogen) directly to the beads. RNAs were extracted and precipitated with Glycogen and ethanol overnight in −20°C, and used directly for the library preparation with NEBNext® Ultra II Directional RNA Library Prep Kit for Illumina® (E7765) according to the protocol 3 of the kit (Protocol for use with FFPE RNA, step 3.5 onward). The final PCR was performed for 15 cycles (98°C – 30 s, 12 cycle of 98°C – 10 s, 65°C – 75 s and final extension at 65°C – 5 min). The libraries were sequenced on HiSeq4000 machine at genomics platform (University of Geneva), for single end 100 cycles.

All sequencing libraries prepared are listed in Table S2.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis of RNaseq data

Analysis of gene expression data

Reads were sorted into individual libraries based on the barcodes and mapped to the mouse genome (Ensembl release 95) using salmon v0.12.0 (salmon quant with options -l A –validateMappings–gcBias) (Patro et al., 2017). Further analysis was performed using R version 3.6.2 (R Core Team, 2017) and Bioconductor (Huber et al., 2015). The DESeq function of DESeq2_1.25.10 bioconductor package (Love et al., 2014) was used to obtain log2 fold changes of gene expression between wild-type and mutant samples and the adjusted p values. Adjusted p value 0.1 was used as a threshold for statistical significance. The individual mouse tissues were analyzed separately. In MA plots the genes with significantly different expression between the mutant and wild-type samples were highlighted in red. The sets of genes upregulated and dowregulated in the mutant tissues were compared using Venn diagrams (VennDiagram 1.6.20). The Volcano plots were plotted using EnhancedVolcano function from EnhancedVolcano 1.3.5 package (https://github.com/kevinblighe/EnhancedVolcano). Significantly differentially expressed genes with at least two fold change in expression between mutant and wild-type were highlighted in red. We put a specific attention on pseudogenes and predicted genes which were among the highly dysregulated genes. In pseudogene category we included the genes which contain the "pseudogene" string in Ensembl biotype annotation. The genes containing "predicted gene" string in Ensembl description were considered as predicted genes. The frequency of the pseudogenes and predicted genes was compared among the expressed genes (genes with defined p value and log2 fold change), genes upregulated and downregulated in the mutant (adjusted p value < 0.1) and top 200 and 100 upregulated (adjusted p value < 0.1 and log2 fold change >0) and downregulated (adjusted p value < 0.1 and log2 fold change < 0) genes, ordered based on log2 fold change. In individual tissues we also compared the proportion of the pseudogenes which are upregulated, downregulated, and not changed in the mutant (adjusted p value > = 0.1). To check whether the expression of parental genes of dysregulated pseudogenes is affected, we obtained the information about the parental genes for the pseudogenes using C57BL/6NJ data from http://mouse.pseudogene.org/. Only for 20%–30% of dysregulated pseudogenes in individual tissues we were able to obtain the parental gene identifier. The gene expression of the dysregulated pseudogenes and their parental genes was visualized using boxplots with the individual genes plotted as gray dots. To see whether whole chromosome expression might be affected, we produced boxplots of log2 fold expression changes for genes lying on individual chromosomes with the y-axes limited to $< -1,1 >$ to better see whether the median fold change deviates from zero. To inspect the chromosomal localization of individual dysregulated genes, we plotted the ideograms where we highlighted their localization. To assess whether the transcripts of dysregulated genes preferentially start with specific nucleotide, for every expressed gene we computed the fractions of its transcript isoforms annotated in Ensembl starting with A,U,G and C. These fractions made for individual starting nucleotides (considered as TSS nucleotides) were summed for selected groups of genes and used to calculate the fraction of genes starting with individual nucleotides. Barplots were made to compare the genes which were dysregulated in the mutant with the genes without detected significant difference. To see whether the starting nucleotide affects the overall gene expression changes, we filtered only genes which produce transcripts with specific starting nucleotide and divided them into the groups based on that nucleotide. We then computed empirical cumulative distribution function on log2 fold changes between mutant and wild-type and plotted the distribution.

The group of genes found to be significantly up- or downregulated in the mutant were searched for enriched Gene Ontology terms in the Biological Process ontology using ENRICHR (Chen et al., 2013; Kuleshov et al., 2016) and the enriched categories were shown.

To find out whether the dysregulated genes in the *Pcif1* mutant testes, are expressed in the specific spermatogenic populations, we re-analyzed the published dataset containing gene expression data for Sertolli cells, spermatogonia, spermatocytes, spermatids, and spermatozoa (Soumillon et al., 2013). Using the boxplots we plotted the expression in these populations for the genes we found to be dysregulated in the mutant testes.

Analysis of ribosome footprinting data

To analyze the translational differences between the mutant and wild-type brain samples for individual genes, we compared the ratio of footprint versus the input reads (i.e., translation efficiency) for individual genes. We tested the differences in translation efficiency for individual genes using DESeq2 design: '~assay(input/footprint) + condition(KO/WT) + assay:condition' and applying the likelihood ratio test which removes the interaction term in the reduced model. For each gene we obtained log2 fold changes of gene translation efficiency difference between wild-type and mutant samples and the adjusted p value. Downstream analysis was analogous to the analysis of gene expression data.

Analysis of m^{6}Am-Exo-Seq data

Reads were sorted into individual libraries based on the barcodes and mapped to the mouse genome (Ensembl release 95) using STAR (Dobin et al., 2013) with parameters "–outFilterType BySJout–limitOutSJcollapsed 50000000–limitIObufferSize 1500000000." The genomic coverages for individual samples were created in R using GenomicAlignments::coverage method. Cumulative coverages of input and m^{6}A IP samples around the annotated TSS positions were calculated and plotted for −500:+500 distance, as well as an m^{6}A enrichment (log2 of m^{6}A/input ratio). As expected the sequenced fragments came predominantly from 5′ ends of the transcripts as demonstrated by increased coverage close to TSS. The TSSs were then divided according to the starting nucleotide and the m^{6}A enrichment around the TSSs was plotted. The m^{6}A enrichment at the starting nucleotide was

compared between in *Pcif1* knockout and wilt-type mice (t.test, n = 3). Significantly lower m^6A enrichment in *Pcif1* knockout mice was detected only for A starting transcripts (pval = 0.009).

For individual genes we tested the differences in m^6A enrichment of their TSSs using matrix of read counts covering TSS of individual genes and DESeq2 design: '~assay(input/m^6A_IP) + condition(KO/HET) + assay:condition' and applying the likelihood ratio test which removes the interaction term in the reduced model. For each gene we obtained log2 fold m^6A enrichment difference between wild-type and mutant samples and the adjusted p value. We identified 1370 genes which had significantly lower m^6A TSS enrichment in the mutants. Importantly these genes tend to have decreased expression in the mutant as shown by different empirical cumulative distribution curves of log2 fold changes in expression between mutant and wild-type mice.