THE NOVEL RNA METHYLCYTOSINE HYDROXYLASE TET IN DROSOPHILA AND ITS REQUIREMENT IN NEURONS AND MUSCLES

By

FEI WANG

A dissertation submitted to the

School of Graduate Studies

Rutgers, the State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Cell and Developmental Biology

Written under the direction of

Ruth Steward

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

OCTOBER, 2017

©2017

Fei Wang

ALL RIGHTS RESERVED

ABSTRACT OF THE DISSERTATION

THE NOVEL RNA METHYLCYTOSINE HYDROXYLASE TET IN DROSOPHILA AND ITS REQUIREMENT IN NEURONS AND MUSCLES

BY: FEI WANG

Dissertation Director:

Dr. Ruth Steward

The TET (Ten-eleven translocation) 1, 2 and 3 proteins have been shown to function as DNA hydroxymethylases in vertebrates and their requirements have been documented extensively. We show that in *Drosophila* hydroxymethylcytosine preferentially marks polyadenylated RNAs and is deposited by Tet. Due to the lack of Cytosine methylation in DNA, Drosophila serves as an ideal model to study the biological function of 5hmrC. We maped the transcriptome-wide hydroxymethylation landscape, revealing hydroxymethylcytosine in the transcripts of many genes, notably in coding sequences, and identified consensus sites for hydroxymethylation. We found that RNA hydroxymethylation may favor mRNA translation. We show that Tet is essential for viability as Tet complete loss-of-function animals die at the late pupal stage. We also characterized the temporal and spatial expression and requirement of Tet throughout Drosophila development. Tet is highly expressed in neuronal tissues and at more moderate levels in somatic muscle precursors in embryos and larvae. Depletion of Tet in muscle precursors at early embryonic stages leads to defects in larval locomotion and late pupal lethality. Although Tet knock-down in neuronal tissue does not cause lethality,

Tet is essential for neuronal function during development and affects locomotion in larvae and adults, as well as the circadian rhythm of adult flies. Further, we report the function of Tet in ovarian morphogenesis. Together, our findings provide basic insights into the biological function of 5hmrC, a modification that is likely also regulated by Tet proteins in other species, and may illuminate observed neuronal and muscle phenotypes observed in vertebrates.

ACKNOWLEGDEMENTS

First of all, I would like to express my deepest gratitude to Dr. Ruth Steward. She is not only my mentor for research, but the mentor for life. In the past five years, I have observed, listened to and learned from this amazing human being. She granted me enough freedom so I can think independently. She also provided ample guidance so I didn't get lost. Her wisdom for life enlightened and inspired me to grow into a comprehensive and significant individual. I am still far from what I want to become but I keep moving forward.

I would like to thank my wonderful committee members that have guided me throughout this process. Dr. Ken Irvine for his invaluable feedback discussing my project as well as generosity with sharing equipment and reagents. Dr. Marc Gartenberg and Dr. Mikel Zaratiegui, not only for your valuable suggestions for my thesis, but also for your encouragement when I wanted to explore the world outside academia.

Next, I would like to thank my colleagues Svetalana Minakhina, Neha Changela and William Tan. Sveta, you not only guided my science but also introduced me to the delicious world of mushrooms. Neha, you not only lent me your fly and imaging skills, but provided much needed encouragement at my darkest time in graduate school. Will, I will always remember our adventures searching for free food and your generosity feeding me while I was a hungry graduate student.

Thank you to all my friends. You guys balanced out the bitterness of PhD. I would never forget all the fun time we have had in hiking, caving, swimming, skiing and kayaking. We may, perhaps, chase our dreams to any corners of the earth, but our friendship will live forever because we were young together.

Last but not least, thank you to my dearest husband, Mr. Will Sorenson. Thank you for not showing up for the first four years of my PhD so I could focus and finish it. ☺

Table of Contents

Abstract of the Dissertation ii
Acknowlegdements iv
Table of Contents v
List of Tables vii
List of Illustrations viii
Chapter 1: Introduction 1
Epigenetics and 5mC DNA methylation 1
DNA demethylation and Tet protein 2
Tet protein structure 3
Tet in other organism 3
DNA methylation and Tet in Drosophila 4
Possible additional function of Tet 4
RNA modifications 5
Functions of RNA modifications 6
Chapter 2: Transcriptome-wide distribution and function of RNA hydroxymethylcytosine 8
I. Preface 8
II. Abstract 9
III. Main Text 9
IV. Supplementary Results 20
V Materials and Methods 35
Chapter 3: Drosophila Tet is required in neurons, muscle precursors and the developing ovary 45
I. Preface 45
II. Abstract 45
III. Introduction 46
IV. Results 49
V. Discussion 79
VI. Materials and methods 84
Chapter 4: General discussion 88
RNA modifications in neurons 88

RNA 5hmrC function and Tet in vertebrates 89
5hmrC decorates mRNA 91
Identification of writer, eraser, and reader proteins of modifications 92
Summary 94
Appendix 1: Supplementary Figures 95
References 97

LIST OF TABLES

Table 1: The Tet alleles 50
Table 2: Phenotypes of Tet KD with different drivers. 53

LIST OF ILLUSTRATIONS

Figure 1 RNA hydroxymethylation by dTet in Drosophila S2 cells. 16
Figure 2 Transcriptome-wide distribution of hmrC in Drosophila cells. 17
Figure 3 RNA hydroxymethylation can favor mRNA translation. 18
Figure 4 dTet-deficient fruitflies show impaired brain development, accompanied by decreased RNA hydroxymethylation. 19
Figure 5. (related to Fig. 1): No evidence for cytosine DNA hydroxymethylation in Drosophila. . 20
Figure 6 (related to Fig. 1): Dot blot controls. 22
Figure 7 (related to Fig. 1C, D): Dot blot controls in S2 cells. 24
Figure 8 (related to Fig. 2): hMeRIP-Seq in S2 cells. 25
Figure 9 (related to Fig. 2B): The enrichments observed in hMeRIP-Seq experiments are not due merely to nonspecific co-precipitation of abundant fragments. 26
Figure 10 (related to Fig. 2C): Examples of enrichment profiles (UCSC tracks) of hMeRIP-Seq done in control and dTet-depleted cells. 28
Figure 11 (related to Fig. 2): High agreement between hMeRIP-Seq experiments performed with two distinct hmC antibodies. 29
Figure 12 (related to Figure 2): Overrepresented motifs in hMeRIP peaks and Gene Ontology analysis of hmrC-containing transcripts. 30
Figure 13 (related to Fig. 2F,G): Relationship between hmrC-containing RNAs and gene expression. 32
Figure 14 (related to Figure 2): The RNA methyltransferases NSun2 and Dnmt2 do not seem to impact on hmC within Drosophila RNA. 34
Figure 15 The Tet locus. 60
Figure 16 Tet expression. 63
Figure 17 Tet requirement in larval locomotion. 66
Figure 18 Tet and the circadian rhythm. 68
Figure 19 Tet expression in PDF neurons. 70
Figure 20 Tet controls ovarian development. 72
Figure 21 Tet mRNA and Protein Expression. 74
Figure 22 Tet KD. 75
Figure 23 Mutant larval size and neuromuscular junctions. 76
Figure 24 Period length in pdf-gal4 KD flies. 77
Figure 25 Tet mutant ovary. 78

CHAPTER 1: INTRODUCTION

Epigenetics and 5mC DNA methylation

Certain traits are inherited from generation to generation through information that is not encoded in the DNA sequence. Because these changes are not encoded by the DNA sequence, they are known as "epigenetic", and they include such changes as covalent DNA or histone modifications, which are responsible for carrying the inherited information by altering the chromatin state. Chromatin is composed with DNA and histones. Nucleosome (composed of histones) spacing determines chromatin structure which can be broadly divided into heterochromatin (condensed chromatin structure) and euchromatin (open chromatin structure). Chromatin structure and gene accessibility to transcriptional machinery are regulated by modifications to both DNA and histone tails. For example, H3K9me3 is highly associated with heterochromatin (SHAHBAZIAN AND GRUNSTEIN 2007), and H3K4me is with euchromatin (SANTOS-ROSA *et al.* 2002). DNA and histone modifications cooperate to precisely regulate gene expression and to maintain genome integrity (PAPAMICHOS-CHRONAKIS AND PETERSON 2013). DNA 5-methylcytosine (5mC, methylation of cytosine at the 5- carbon position) is one of the well-characterized key epigenetic marks that plays an important role in the development in mammals (LAW AND JACOBSEN 2010). Distinct genomic patterns of DNA 5mC are important features of stem cell identity. The key feature of DNA 5mC is its enrichment in CpG dinucleotides (JONES 2012).

DNA 5mC is generally associated with a repressed chromatin state and inhibition of promoter activity (IRVINE *et al.* 2002). It can function either as a transcription inhibition factor or as factors that induce changes in chromatin structure so that the access of transcription factors to the particular promoter is restricted (IRVINE *et al.* 2002). For example, methylated CpGs recruit methyl-CpG-binding proteins that further recruit 'repressor complexes' (eg. polycomb2),

resulting in histone modification. The modification of histones can further recruit repressor complexes that leads to more heterochromatin as opposed to euchromatin favored by transcription (JONES *et al.* 1998; BIRD AND WOLFFE 1999).

In vertebrates DNA 5mC maintenance is carried out by the Dnmt1 DNA methyltransferase that has a strong affinity for hemimethylated DNA which is responsible for heritability of epigenetic information (GOLL AND BESTOR 2005). Dnmt1 is supplemented by the de novo methyltransferas Dnmt3 (GOLL AND BESTOR 2005). Together, both enzymes catalyze the establishment and maintenance of the cytosine DNA methylation pattern during animal development and cell fate specification.

DNA demethylation and Tet protein

Until recently, DNA 5mC was believed to be irreversible and could only be passively removed through DNA replication. Therefore, it was a landmark discovery when the Ten Eleven Translocation protein 1 (TET1) was shown to be a DNA demethylase (TAHILIANI *et al.* 2009). The TET family contains three proteins(TET1/2/3) that catalyze the successive oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (HE *et al.* 2011; ITO *et al.* 2011). These 5mC oxidation products can serve as intermediates in the conversion of 5mC to unmodified cytosines (DELATTE 2014). TET-mediated DNA hydroxymethylation regulates key biological processes such as stem cell reprogramming (GU *et al.* 2011; WU AND ZHANG 2011) and differentiation (HAHN *et al.* 2013). In addition, disruption of the DNA 5hmC pattern has been shown to be associated with diseases such as cancer (GAIDZIK *et al.* 2012; PEREZ *et al.* 2012) and neurological disorders (JIN *et al.* 2011; MELLEN *et al.* 2012; HAHN *et al.* 2013; ZHANG *et al.* 2013).

Tet protein structure

TET proteins are large (180 - 230kDa) multidomain enzymes. All TET proteins contain a conserved double-stranded β-helix (DSBH) domain and a cysteine-rich domain with binding sites for the cofactors Fe(II) and 2-oxoglutarate (2-OG) that together form the catalytic region in the C terminus (TAHILIANI *et al.* 2009). In addition to their catalytic domain, TET1 and TET3 also have an N-terminal CXXC zinc finger domain that can bind DNA (TAHILIANI *et al.* 2009; XU *et al.* 2012). The CXXC DNA binding domain has ability to bind to unmethylated cytosines of CpG dinucleotides (LONG *et al.* 2013). The highly conserved DNA binding regions (~ 45 AA in length) are characterized by 8 conserved cysteine residues (XU *et al.* 2012). The conserved core domains distinguish the TET family of dioxygenases from other members of the dioxygenase superfamily by having CXXC DNA binding domain. The combination of all of these essential domains is what characterizes TET proteins (Although Tet2 has lost the CXXC domain probably during evolution).

Tet in other organisms

Organisms can generally be split into different groups based on the number of *Tet* genes identified in their genomes. The first group, vertebrates, has 3 *Tet* genes (ITO *et al.* 2011), except *Xenopus tropicalis* (a model organism), which only has 2 members (orthologs of *Tet2* and *Tet3*) (XU *et al.* 2012).

The second group is made up of all other metazoans from the relatively simple *Trichoplax* onwards. It includes arthropods, and insects including honeybees and ants, which all contain a single *TET* gene. Their Tet proteins are very highly conserved to the mammalian Tet proteins in their CXXC and catalytic domains (DUNWELL *et al.* 2013). One notable exception is that

Caenorhabditis elegans and other nematodes that contain no identifiable *Tet* genes, but also have no *DNMT* genes and no detectable genomic DNA methylation (DUNWELL *et al.* 2013).

As discussed below it was unexpected that the genome of *Drosophila melanogaster* also contains a Tet homolog. The Tet proteins contain both the CXXC DNA binding domain and the catalytic domain.

DNA methylation and Tet in Drosophila

For over 2 decades, the existence of DNA 5mC in the Drosophila genome has been controversial. Flies lack homologs of the mammalian DNA methyltransferase genes *Dnmt1*, *Dnmt3a*, or *Dnmt3b* but contains a gene encoding a homolog of *Dnmt2* (also known as Mt2). *Dnmt2* null mutants are homozygous viable and fertile (LYKO *et al.* 2000). DNA methylation in Drosophila was reported to be restricted to early stages of embryo development (LYKO *et al.* 2000). However, in recent publications, the conclusion was reached that the genomes of Drosophila lack DNA methylation altogether (RADDATZ *et al.* 2013). This difference in results was due to a lack of specificity and a potential contamination of the earlier applied methods. If DNA methylation is dispensable, why is Tet well conserved in Drosophila? This mysterious presence of Tet is the main drive for my thesis research.

Possible additional function of Tet

In DNA, Tet can also oxidize thymine to 5-hydroxymethyluracil (Pfaffeneder, Nature Chemical Biology 2014). Furthermore, there is structural similarity between Tets and the ALKB family of dioxygenases (Hu Cell 2014). ALKB can catalyze Fe2+ and a-ketoglutarate-dependent hydroxymethylation of adenosine in both DNA and RNA (AAS *et al.* 2003). This raises an

interesting and attractive possibility: Tet might catalyze hydroxymethylcytosine formation in RNA as well. Indeed, recent mass spectrometry analyses have revealed that mammalian TETs can hydroxymethylate RNA (FU *et al.* 2014a). When I started working on my PhD, the distribution, function(s), and biological relevance of RNA 5hmC was still a mystery.

In the journey of investigating the function of Tet in Drosophila, I have proposed multiple hypothesis to explain the mysterious presence of Tet. In Chapter 2, we provide evidence and validation for the hypothesis that Tet is indeed responsible for establishing 5hmC mark in RNA and we characterized the distribution and potential functional importance of RNA 5hmC. In chapter 3, we used a genetic approach to characterize Tet's requirement by defining its mutant phenotype and by elucidating the biological function in development.

RNA modifications

RNA, transferring the genetic information from DNA into proteins, plays multiple roles in the cell, catalyzing and controlling biochemical reactions, and regulating gene expression (SHARP 2009). Similar to the epigenome, various chemical modifications are known to decorate RNA molecules, about a hundred of which are known to date (HE 2010). These marks include pseudouridine (ψ), N6-methyladenosine (m6A), and C5-methylcytidine (5mC) (HE 2010). Until very recently, these RNA modifications were thought to be restricted mainly to the very abundant transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs). Thanks to the development of deep-sequencing techniques, researchers were able to map some RNA modifications transcriptome-wide, pseudouridine (CARLILE *et al.* 2014), methyladenosine (SCHWARTZ *et al.* 2014), and methylcytosine (SQUIRES *et al.* 2012), showing that these modifications are present in specific regions of subsets of transcripts.

The distribution of m6A has been well characterized and validated by several methods from different research labs (SCHWARTZ *et al.* 2014; LINDER *et al.* 2015; GROZHIK *et al.* 2017). In eukaryotes, it is enriched near stop codons and long internal exons. It is also prevalent in ribosome-associated mRNAs(WANG *et al.* 2014; WANG *et al.* 2015). The nonrandom distribution implies important regulatory functions. Furthermore, these modifications can be dynamic in response to various stresses (FUSTIN *et al.* 2013; BATISTA *et al.* 2014; ZHOU *et al.* 2015).

Functions of RNA modifications

m6A is the most abundant RNA modification in mRNA. Genetic analysis of m6A methyltransferases (writers), binding factors (readers) and demethylases (erasers) have elucidated the functional importance of RNA m6A modification in control of transcription, RNA splicing and translation (WANG *et al.* 2014; WANG *et al.* 2015; DU *et al.* 2016; HAUSSMANN *et al.* 2016; LENCE *et al.* 2016).

Alteration of m6A RNA modification has been linked to developmental processes and various diseases (BEN-HAIM *et al.* 2015; JAFFREY AND KHARAS 2017). In mammals, m6A is required in ES cell differentiation, circadian rhythms and dosage compensation (FUSTIN *et al.* 2013; BATISTA *et al.* 2014). In addition, hypoxia causes reduction of m6A levels by up-regulating the demethylase ALKB homolog 5 (ALKBH5), associated with breast cancer progression(MULLER *et al.* 2017). Recent studies in Drosophila have shown that depletion of m6A did not affect the viability but reduced the fertility and flight ability, also impairing the locomotion and the dosage-dependent sex determination (HAUSSMANN *et al.* 2016; LENCE *et al.* 2016). Co-localization of m6A the methyltransferase complex and RNA pol II suggested that m6A methylation was added to actively transcribed mRNA. The presence of m6A affects multiple aspect of mRNA processing including splicing, 3'-end processing, nuclear export, translation and mRNA

decay(WANG *et al.* 2014; WANG *et al.* 2015; FRYE *et al.* 2016; HAUSSMANN *et al.* 2016; LENCE *et al.* 2016). Most interestingly, in the most recent study, m6A also coordinates transcription with translation. Dynamics of RNA polymerase II impact the deposition of m6A on mRNAs, substantially regulating translation efficiency. m^6A modification of mRNAs is co-transcriptional. Suboptimal transcription rates lead to elevated m^6A content, which may result in reduced translation (SLOBODIN *et al.* 2017).

These observations suggest that RNA modification represents a novel layer of gene regulation by controlling mRNA processing. RNA modifications may be required for fine-tuning the connection between transcription and levels of translation.

In my thesis, I use genetic approaches to investigate the function of Tet (converter of 5mC to 5hmC) in Drosophila development. By analyzing Tet loss-of function mutants. In Chapter 3, I describe the *Tet* phenotypes. My results show that Tet expression and requirement is tissue specific and developmentally gated. Tet function is essential in neurons and muscle precursors. While we did not find gross abnormalities in the development of these tissues phenotypes indicate that 5hmrC is most important for the normal function of these tissues.

CHAPTER 2: TRANSCRIPTOME-WIDE DISTRIBUTION AND FUNCTION OF RNA HYDROXYMETHYLCYTOSINE

I. Preface

This chapter was published, as presented here, in Science, July 2016. My contributions to the project and paper were fly genetics, preparing material for RNA 5hmrC quactitive analysis and revising the manuscript. The original manuscript does not contain the discussion. Therefore, the General discussion focuses more on this chapter.

II. Abstract

Hydroxymethylcytosine, well described in DNA, occurs also in RNA. Here, we show that hydroxymethylcytosine preferentially marks polyadenylated RNAs and is deposited by Tet in *Drosophila*. We map the transcriptome-wide hydroxymethylation landscape, revealing hydroxymethylcytosine in the transcripts of many genes, notably in coding sequences, and identify consensus sites for hydroxymethylation. We found that RNA hydroxymethylation can favor mRNA translation. Tet and hydroxymethylated RNA are found to be most abundant in the *Drosophila* brain, and Tet-deficient fruitflies suffer impaired brain development, accompanied by decreased RNA hydroxymethylation. This study highlights the distribution, localization, and function of cytosine hydroxymethylation and identifies central roles for this modification in *Drosophila*.

III. Main Text

In DNA, vertebrate Tet methyldioxygenases (Tet1, Tet2, and Tet3) catalyze hydroxylation of 5-methylcytosine to 5-hydroxymethylcytosine (KRIAUCIONIS AND HEINTZ 2009; TAHILIANI *et al.* 2009; KOHLI AND ZHANG 2013). Tet also catalyzes the formation of hydroxymethylcytosine in RNA (referred to here as hmrC) (RADDATZ *et al.* 2013; FU *et al.* 2014a). To date, however, the distribution, localization, and functional relevance of hmrC remain unknown.

In the present study, we have sought to provide a better understanding of hmrC. For ease of interpreting results, we analyzed hmrC in *Drosophila melanogaster* because (i) cytosine methylation in *Drosophila* DNA is either absent or very low, being restricted

to specific cellular contexts (RADDATZ *et al.* 2013; TAKAYAMA *et al.* 2014), and (ii) we have found no evidence of DNA hydroxymethylation in this organism (fig. S1). To detect 5hmrC RNA modification, we used an antibody raised against 5-hydroxymethylcytosine (STROUD *et al.* 2011; WU *et al.* 2011). To confirm that it will bind to hmrC, we performed dot blot experiments using in vitro transcribed templates containing either unmethylated, methylated, or hydroxymethylated cytosines (table S1). The antibody to 5hmC specifically detected 5hmrC-containing RNA. In addition, detection of 5hmrC was abolished after ribonuclease (RNase) A treatment (A).

We detected hmrC in dot blot experiments on total RNA extracted from *Drosophila* S2 cells (Figure 1A and Figure 5 B and C). Isolation of polyadenylated RNA from S2 cells followed by immunoblotting showed strong enrichment in hmrC signal as compared with that of total cellular RNA (Figure 1Band Figure 5, D and E). No signal was detected in fractions enriched in small RNAs or ribosomal RNAs (Figure 1C and Figure 7, A and B). *Drosophila* possesses only one conserved Tet ortholog, *CG43444* (dTet) (DUNWELL *et al.* 2013; ZHANG *et al.* 2015). Depletion of dTet in S2 cells, by using RNA interference for dTet (dTet KD), revealed a 54% decrease in *dTet* transcripts, as compared with control cells (Figure 1D). Dot blotting with antibody to 5hmrC showed a similar decrease in hmrC—44%—upon dTet knockdown (Figure 1D and Figure 7, C and D).

To map the hmrC modification landscape in a transcriptome-wide manner, we adapted a recently used method [methylated RNA immunoprecipitation followed by sequencing (MeRIP-seq)] (DOMINISSINI *et al.* 2012; MEYER *et al.* 2012), which we call hMeRIP-seq. This method involves immunoprecipitation of hmrC-containing RNA with the antibody to 5hmrC followed by next-generation sequencing. hMeRIP-seq in S2 cells

yielded 3058 significantly enriched regions ("hmrC peaks," $P < 10^{-10}$) within 1597 coding gene transcripts (Figure 8, A to C, and table S2). Examples of enrichment profiles are shown in Figure 2A. Several key controls were performed so as to ensure the validity and stringency of our experimental approach: (i) Further bioinformatic analyses demonstrated that our hMeRIP-seq experiments did not merely coprecipitate abundant RNA fragments nonspecifically (Figure 2B and Figure 8); (ii) up to 79.4% of hmrC sites showed significant reduction levels in hMeRIP-seq upon dTet depletion as compared with that of control S2 cells, with 85.5% of the sites showing a more than fourfold reduction in hmrC levels (Figure 2C and Figure 11); and (iii) replicate hMeRIP-seq experiments by using an additional antibody to 5hmrC showed strong agreement between experiments performed with the two 5hmrC antibodies (Figure 11).

The distribution of hmrC peaks revealed by hMeRIP-seq analyses was significantly nonrandom (Poisson, $P < 10^{-5}$), with many of these peaks found in coding sequences (48%) (Figure 2D). Further analyses identified an overrepresented motif with occurrences in a large proportion of peaks (64%) and which tends to be highly UC-rich, containing UCCUC repeats (Figure 2E and Figure 12). Gene Ontology analysis of the hmrC targets showed enrichment for genes involved in basic cellular processes and notably in the regulation of embryogenesis and development (Figure 12). In addition, exons were more enriched in hmrC than introns (Figure 13D). Thus, the above data present the hmrC-enriched transcriptome, revealing the presence of this modification in specific mRNA regions and in specific sequence contexts.

Next, we knocked down *dTet* expression in S2 cells and performed RNA-sequencing (RNA-seq) experiments in order to assess how many mRNAs might be regulated by dTet. We found the expression of 574 mRNAs to change after dTet

depletion (50.4% were more abundant, whereas 49.6% were less abundant) (Figure 2F; Figure 13, A to C). Examples of mRNAs whose levels either increased or decreased upon dTet knockdown are shown in Figure 13, A and B. A positive correlation was observed between transcript abundance and the presence of hmrC peaks (Figure 2F and Figure 9). We then compared these dTet-regulated mRNAs with the targets identified by hMeRIP-seq and found a slight but significant percentage (26%, Poisson, $P < 10^{-43}$) of the dTet-regulated mRNAs to contain at least one hmrC peak (Figure 2G, Figure 13C). It is worth mentioning that dTet contains an N-terminal CXXC Zn-finger domain, which likely explains the reported ability of mammalian Tets to regulate gene expression independently of their catalytic activity (WILLIAMS *et al.* 2011; DELATTE 2014). It thus remains possible that dTet might affect gene expression via its CXXC domain, independently of its ability to hydroxylate methylated RNA (Figure 14, A and B). This domain, however, is likely not required for the role of dTet in *Drosophila* brain development (see below) because flies where dTet CXXC is deleted are still viable and show no specific phenotype (ZHANG *et al.* 2015).

To examine how cytosine hydroxymethylation might affect mRNA function, we examined the distribution of hmrC as a function of the mRNA translational status in *Drosophila* S2 cells. For this, we performed standard sucrose-gradient fractionation followed by dot blotting. A correlation between hmrC abundance and active mRNA translation was observed; fractions with low translation activity (free mRNAs and monosomes) were found to be poor in hmrC, whereas mRNAs heavily loaded with ribosomes (polysome fractions) showed a high hmrC content (Figure 3A). We also assessed how mrC distributes across polysome fractions and found it to be high in monosomes and low in polysomes. Next, we examined whether mRNA

hydroxymethylation might affect mRNA translation. In vitro translation of unmodified, methylated, and hydroxymethylated Firefly Luciferase–encoding RNA templates in rabbit reticulocyte lysate showed a decrease in translation of methylated RNAs. In contrast, hmrC-modified templates gave rise to near-control protein levels (Figure 3B), suggesting that hydroxymethylation can restore the translation efficiency of previously methylated substrates.

We next sought to assess hmrC in vivo in fruitflies. To determine the timing of *dTet* expression and of the appearance of hmrC during early embryogenesis of *D. melanogaster*, we performed quantitative reverse transcription polymerase chain reaction (RT-PCR) to measure *dTet*transcript levels, and immunoblotting to estimate levels of hmrC. We found that *dTet* levels correlate positively with hmrC levels during fruitfly embryogenesis (Figure 4A). We also analyzed a publicly available database of RNA-seq results from different stages of *D. melanogaster* development and found that in third-instar larvae, *dTet*expression is highest in the central nervous system. These findings suggest that dTet-mediated hydroxymethylation of RNA could play a role in the fruitfly brain. To confirm and extend these observations, we generated transgenic *Drosophila* flies expressing a *GFP-dTet* fusion construct under the control of the endogenous dTet promoter (called *dTet-Mi*). The green fluorescent protein (GFP)–tagged dTet protein was detected throughout the larval brain, the highest levels being detected in the optic lobe and central brain (Figure 4B).

It was of interest to assess whether the fly brain contains high levels of hmrC. To this end, in addition to the brain, we used ovary [because this was the organ chosen to show the role of dTet in DNA m6A demethylation (ZHANG *et al.* 2015)]. We also used another organ, the salivary gland, from which we could extract enough RNA to measure

hmrC levels. Quantitative RT-PCR and dot blotting showed higher dTet expression and hmrC content in the brain than in the ovary or in the salivary gland of the fruitfly (Figure 4C). Hence, by revealing that the hmrC signal is highest in the brain, these data support the argument for the importance of hmrC in this organ.

We wished to evaluate RNA hydroxymethylation levels in a complete loss-of-function mutant of dTet. In agreement with recent observations (ZHANG *et al.* 2015), dTet-deficient animals survived through the larval stages but died at the pupal stage (no adult animals survived; $n > 5000$ dTet-null animals analyzed). Morphological defects were observed at larval stages: The brains of mutant larvae were smaller than those of normal larvae and showed abnormal organization of neuroblasts in their central part (Figure 4D). To quantify the observed changes, we measured the width of the medulla region (based on 20 examined brains). We found it to be significantly reduced in $dTet^{null}$ animals ($P < 2.4 \times 10^{-9}$), likely reflecting a lower number of neuroblasts (Figure 4D). To measure the effects of dTet loss on the RNA hydroxymethylation level, we performed immunoblotting analyses with antibody to 5hmrC on brains from $dTet^{null}$and wild-type larvae. These experiments showed a decreased hmrC content in the brains of dTet-deficient larvae (Figure 4E).

Our understanding of the posttranscriptional modifications that decorate RNA, a regulatory layer positioned between DNA and proteins, is in its infancy. We have conducted a study addressing the distribution, localization, and function of cytosine hydroxymethylation in RNA, using *Drosophila melanogaster* as a model. Our work has yielded the following key findings: (i) It provides a picture of the hydroxymethylated transcriptome, (ii) reveals an unrecognized function for hmrC, and (iii) suggests a central role for this RNA modification and dTet in the *Drosophila* brain. All in all, we

expect this study to change the way we think about the roles played by cytosine hydroxymethylation and the Tet proteins. Our findings open new research prospects in an emerging realm of biological regulation: epitranscriptomics.

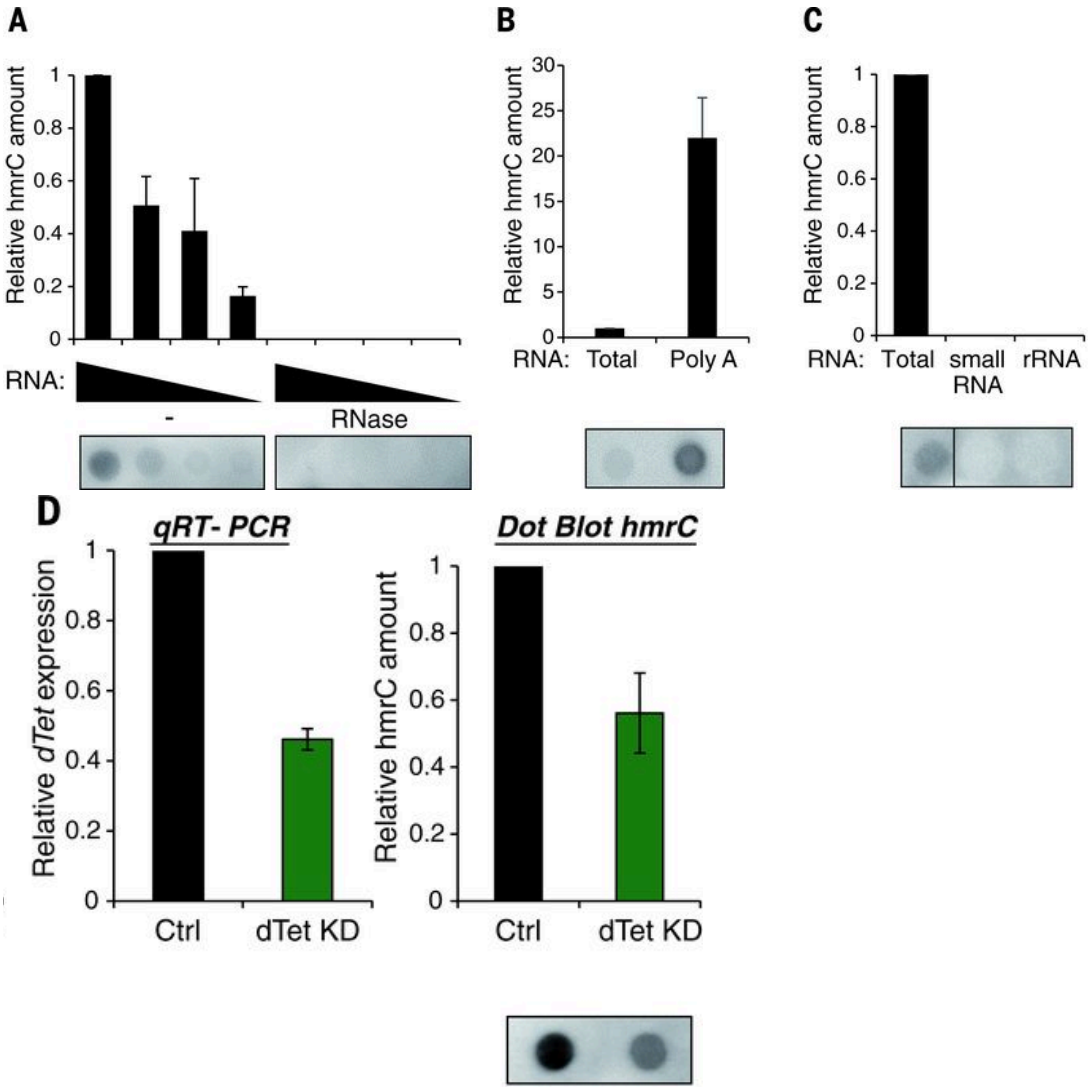


Figure 1 RNA hydroxymethylation by dTet in Drosophila S2 cells.

(**A**) Dot blotting on total RNA from *Drosophila* S2 cells with antibody to 5hmC, treated or not with RNase A (serially halved amounts of RNA, starting at 1 µg). Data are mean ± SD (n = 4 experiments run) with a representative blot shown. (**B**) Immunoblotting with anti-5hmC antibody was performed on polyadenylated and total RNA from S2 cells. Data are mean ± SD (n = 3 experiments run). (**C**) hmrC content of total RNA as well as fractions enriched in small RNA or rRNA was assessed by dot blotting. Data are mean ± SD (n = 3 experiments run). A vertical line indicates juxtaposition of lanes within the same blot, exposed for the same time. (**D**) dTet knockdown leads to reduced hmrC levels. (Left) Quantitative RT-PCR analysis. (Right) Dot blotting. Data are mean ± SD (n = 4 experiments run).

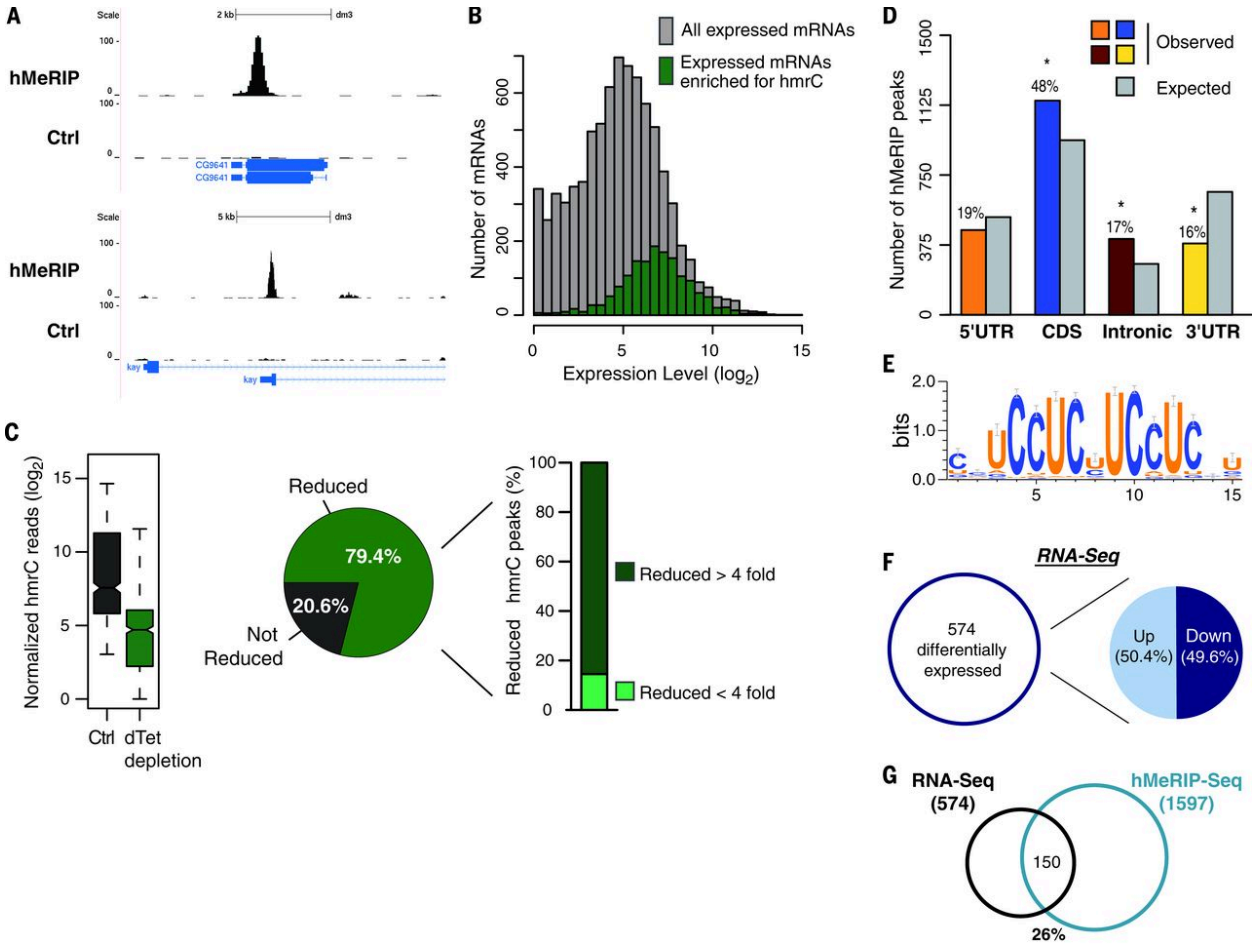


Figure 2 Transcriptome-wide distribution of hmrC in Drosophila cells.

(**A**) Representative UCSC Genome Browser plot from hMeRIP-seq data. (**B**) Distribution of all expressed (gray) or hmrC-enriched (green) transcripts, showing the number of mRNAs as a function of their expression levels. In hMeRIP-seq, enrichment in both abundant and less abundant fragments was observed. (**C**) hMeRIP-seq in cells depleted of dTet shows reduced hmrC levels at the majority of target regions. (Left) Box plot of the normalized number of hmrC reads in dTet-depleted cells versus control cells. (Center) Pie chart showing the percentage of reduced hmrC peaks, with (right) a more than fourfold reduction at most targets. (**D**) Distribution of hmrC peaks according to the type of structural element within the transcript. *$P < 10^{-}$). (**E**) Sequence motif identified within hmrC peaks. (**F**) RNA-seq performed on dTet-knockdown and control S2 cells. (**G**) Partial overlap between hMeRIP-seq and RNA-seq data sets.

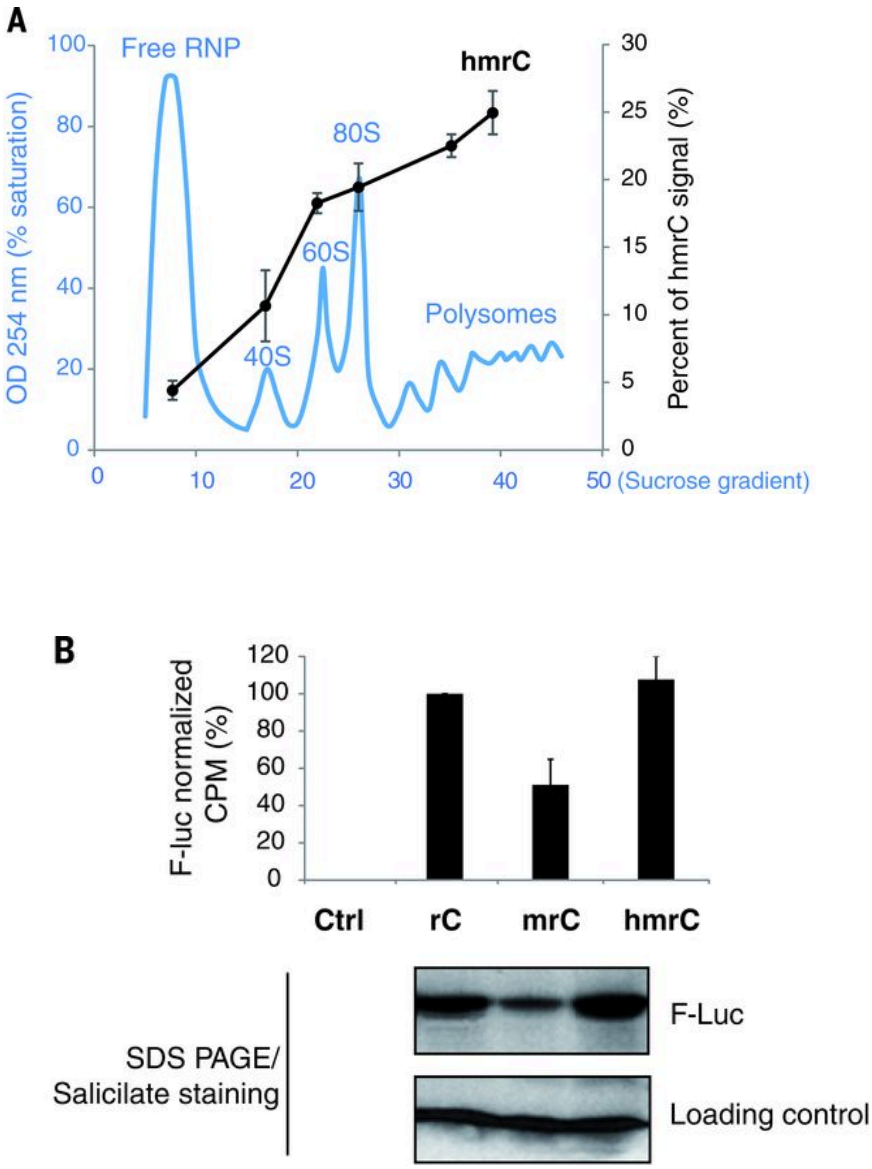


Figure 3 RNA hydroxymethylation can favor mRNA translation.

(**A**) Sucrose gradient fractionation followed by dot-blot quantification shows that active translation is associated with a high hmrC content. Data are mean ± SD (*n* = 3 experiments run). (**B**) In vitro translation of C-, mC-, and hmC-containing RNAs, as measured by incorporation of S-radiolabeled methionine, shows that methylation decreases translation and hydroxymethylation restores it. Shown are the normalized scintillation counts (top) and the results of SDS–polyacrylamide gel electrophoresis (SDS-PAGE) followed by fluorography (bottom). Data are mean ± SD (*n* = 3 experiments run).

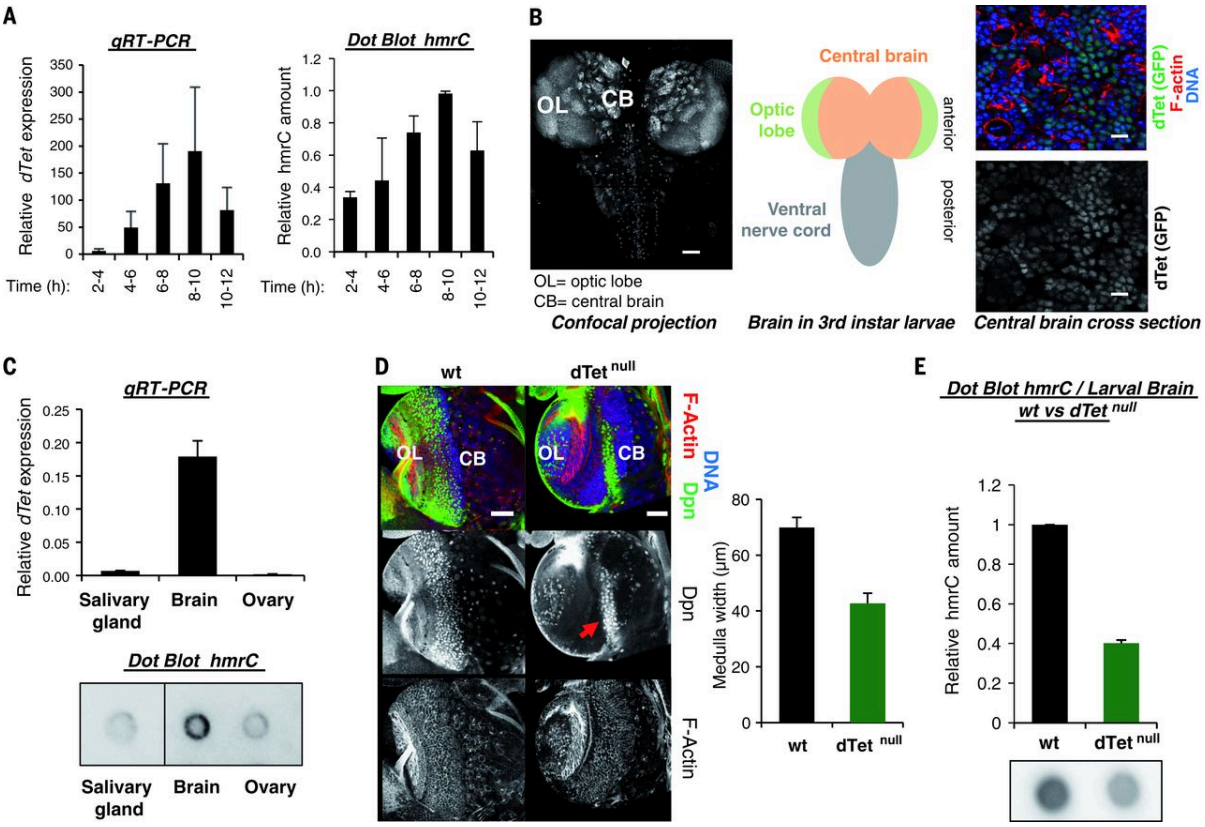


Figure 4 dTet-deficient fruitflies show impaired brain development, accompanied by decreased RNA hydroxymethylation.

(**A**) *dTet* expression and hmrC levels during Drosophila embryogenesis. Data are mean ± SD (*n* = 4 experiments run). (**B**) (Left) Pattern of endogenous GFP-tagged dTet in the larval brain. Scale bar, 50 µm. (Center) Scheme of the larval brain. (Right) Confocal brain section showing the expression of endogenous dTet (green) and F-actin (red). DNA, blue. Scale bar, 10 µm. (**C**) (Top) *dTet* expression in the salivary gland, brain, and ovary. Data are mean ± SD (*n* = 3 experiments run). (Bottom) Immunoblotting with 5hmC antibody in RNA from salivary gland, brain, and ovary. Vertical line indicates juxtaposition of lanes within the same blot, exposed for the same time. (**D**) (Left) *dTet*$^{-}$ brains are smaller than wild-type brains. Dpn (neuroblasts), green; F-actin, red; DNA, blue. Scale bar, 50 µm. (Right) Average width of the medulla region containing the neuroblasts (left, red arrow). Error bars represent 95% confidence intervals ($p < 2.4\ 10^{-}$) for 20 brain lobes. (**E**) Brains of dTet-deficient larvae show a decrease in hmrC. Data are mean ± SD (*n* = 3 experiments run).

IV. Supplementary Results

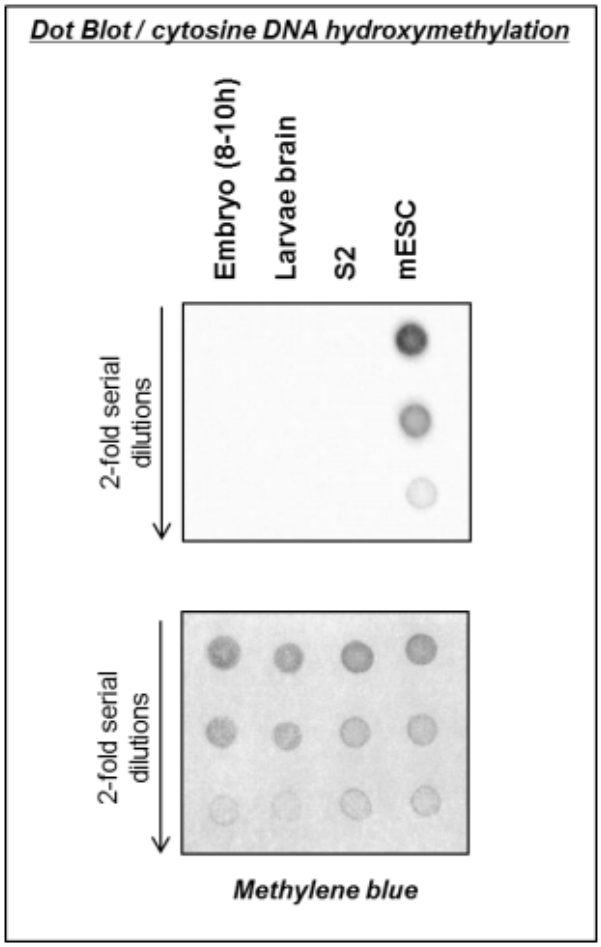


Figure 5. (related to Fig. 1): No evidence for cytosine DNA hydroxymethylation in Drosophila.
To directly assess the cytosine hydroxymethylation level in Drosophila DNA, (i) dot blotting was performed with serially halved amounts of genomic DNA (starting with 300 ng) from S2 cells and from fruitfly embryonic stages where dTet expression is highest (8- 10 h, cf. Fig. 4A). The blots were probed with the anti-hmC antibody. We detected no DNA hydroxymethylation in S2 cells or in the late embryonic stages; (ii) since the larval brain shows the highest level of dTet expression (Fig. 4C), we sought to detect hmC in DNA in this setting. By immunoblotting with anti-hmC antibody, we detected no hmC in larval brain DNA. Loading was assessed by methylene blue staining of the membrane. Mouse ES cells (mESC) were used in parallel as relevant positive controls (starting from 300 ng). Taken together, while we cannot exclude the possibility that hmC might exist in Drosophila DNA (e.g. because it is either very low or restricted to specific cell contexts), we found no evidence of its presence. Regarding the existence of cytosine DNA methylation in Drosophila, this remains controversial and a debated

issue. Drosophila lacks orthologs of mammalian Dnmt1, Dnmt3a and Dnmt3b but contains an ortholog of Dnmt2, a known RNA methyltransferase (27), which has been suggested to be able to methylate DNA. Bisulfite sequencing has indicated that Drosophila lacks cytosine methylation, but this work was performed on embryonic stages 0-2/0-3 hours, and at these time points the level of Dnmt2 expression is close to zero. Adding to this debate are recent work suggesting that DNA methylation does exist in Drosophila: (i) embryos at 2–3 hours postfertilization showed cytosine methylation but independently of Dnmt2 (7); (ii) m5C in DNA is present in the adult fly, but at levels below the detection limit of bisulfite sequencing.

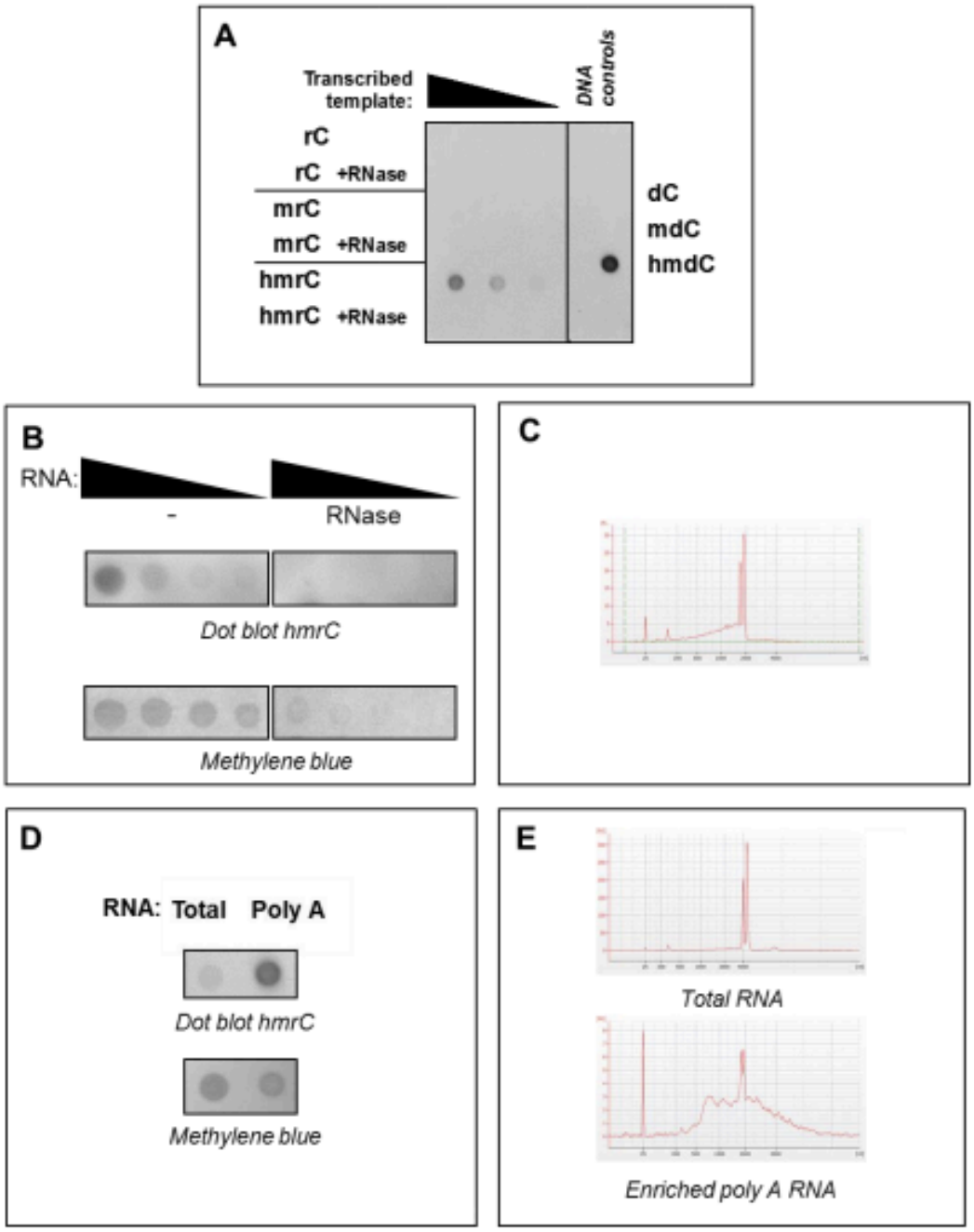


Figure 6 (related to Fig. 1): Dot blot controls.
(A) Antibody specificity and sensitivity for hmC-containing RNA. Serially halved amounts (starting at 50 ng) of transcribed template containing either rC, mrC or hmrC were dot blotted and detected with an anti-hmC antibody. A representative experiment is 14 shown, which was successfully repeated 3 times. (B) Drosophila S2 cells contain hydroxymethylcytosine in RNA. The hmrC content of serially

halved quantities (starting at 1µg) of untreated and RNase A-treated total RNA was detected by dot blotting. Loading was assessed by methylene blue staining of the membrane. (C) The quality of RNA from S2 cells was assessed on a Bionalyzer electropherogram. (D) The polyadenylated RNA pool is enriched in hmrC. The hmrC content of total RNA (200ng) and poly A-enriched RNA (100ng) from S2 cells was detected by dot blotting. Loading was assessed by methylene blue staining of the membrane. (E) The quality of total (top) and poly-A enriched (bottom) RNA from S2 cells was assessed on a Bioanalyzer electropherogram.

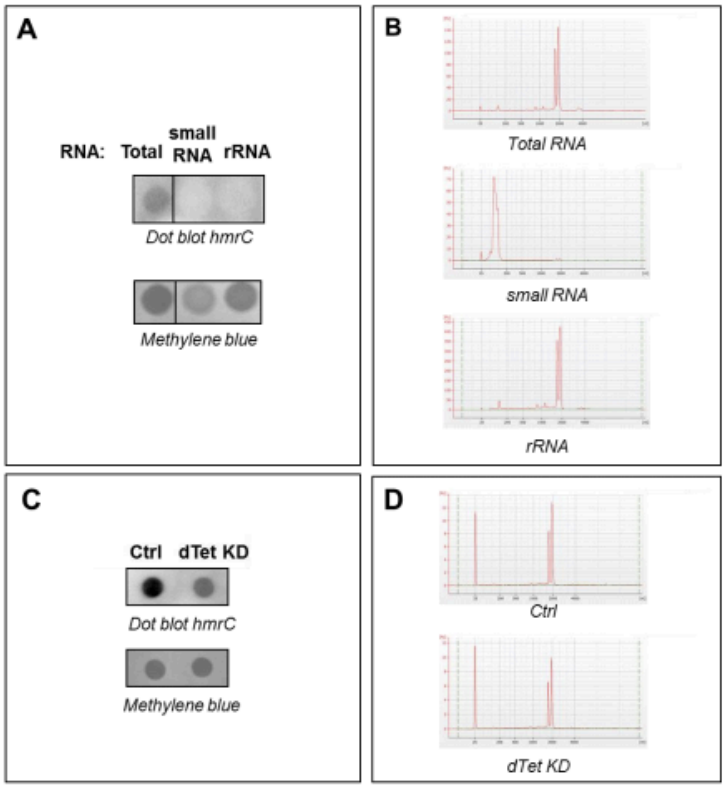


Figure 7 (related to Fig. 1C, D): Dot blot controls in S2 cells.
(A) hmrC is mostly absent from small RNA and rRNA. The hmrC content of total RNA, small RNA and rRNA (300ng of each) from S2 cells was detected by dot blotting. Loading was evaluated by methylene blue staining of the membrane. Vertical lines indicate juxtaposition of lanes within the same blot, exposed for the same time. (B) The quality of total RNA (top), small RNA (middle) and rRNA (bottom) from S2 cells was assessed by Bioanalyzer electropherogram. (C) hmrC in dTet knockdown S2 cells. hmrC content of control and dTet knockdown total RNA (400ng of each) from S2 cells was detected by dot blot. Loading control was assessed by methylene blue staining of the membrane. (D) The quality of control (Ctrl, top) and dTet knockdown total RNA (bottom) from S2 cells was assessed by Bioanalyzer electropherogram.

A

	Peaks	Transcripts
hMeRIP-Seq	3,058	1,597

B

Sample	Read Count (M)	Mapped Reads (M)	Coverage
hMeRIP-Seq S2 (#1)	13.38	11.11	18.46
hMeRIP-Seq S2 (#2)	12.66	10.86	18.04
hMeRIP-Seq Input S2	13.46	11.45	19.02
RNA-Seq S2 r1	8.99	4.17	6.93
RNA-Seq S2 r2	7.73	3.13	5.20
RNA-Seq dTet S2 r1	8.17	3.57	5.93
RNA-Seq dTet S2 r2	9.10	4.22	7.01

C

Chr	Peak Start	Peak End	Enrichment Score	Gene Symbol
chrX	2010975	2012011	323	*ph-d*
chr3L	639330	641109	322	*CR43334*
chr4	676881	679230	322	*gw*
chr3L	13032598	13034084	321	*SRm160*
chrX	10357926	10359411	321	*nocte*
chr3R	6149596	6150774	317	*Bruce*
chr3R	9209085	9210278	316	*Droj2*
chrU	7887856	7889998	314	*CR40572*
chr3R	25598755	25599936	312	*kay*
chrX	8497185	8497969	312	*Caf1-180*
chr3L	8447240	8449412	312	*Gug*
chr2L	189276	191507	310	*spen*
chr2L	235973	239354	310	*kis*
chr2L	1695288	1698106	310	*chinmo*
chr2L	2974862	2975966	310	*CG9641*

Figure 8 (related to Fig. 2): hMeRIP-Seq in S2 cells.
(A) hMeRIP-Seq in S2 cells reveals the presence of hmrC in many transcripts. (B) Deep sequencing statistics. The number of raw reads, usable reads after mapping and filtering, and coverage are shown (r1: replicate #1; r2: replicate #2). hMeRIP-seq S2 #1: Diagenode antibody; hMeRIP-seq S2 #2: Active Motif antibody. (C) Most significant hmrC peaks (top 15) with their associated gene transcript symbols and enrichment scores (-log10 Pvalue).

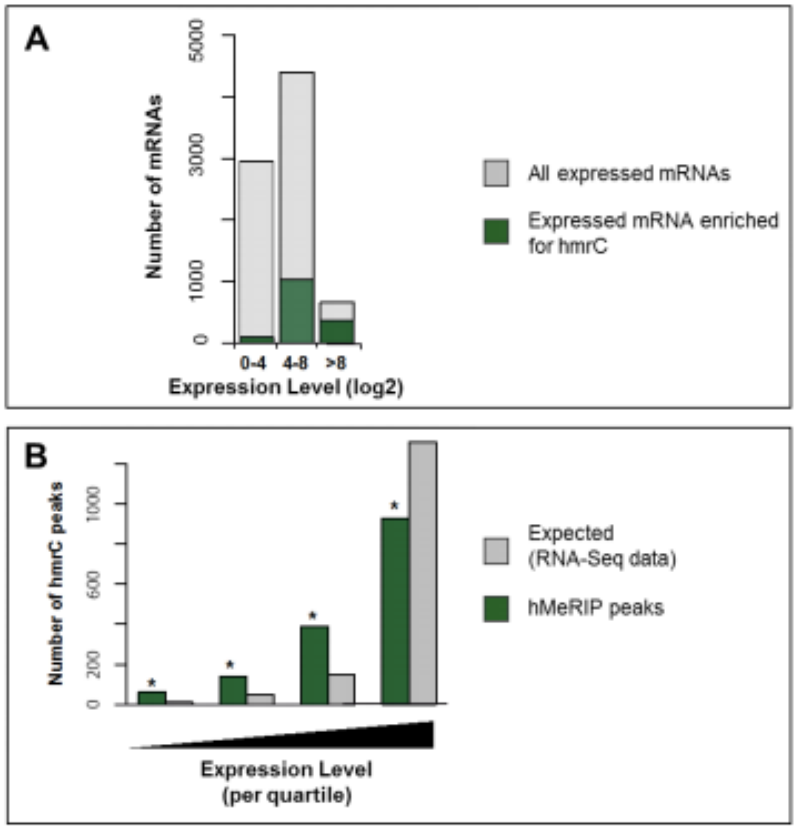


Figure 9 (related to Fig. 2B): The enrichments observed in hMeRIP-Seq experiments are not due merely to nonspecific co-precipitation of abundant fragments.

(A) To control hMeRIP-Seq for nonspecific co-precipitation of abundant mRNAs, we determined whether the gene expression level correlated with the presence of hmrC peaks. For this, we compared the overall distribution of (log2 scale) transcript abundances (gray) with the distribution observed for transcripts containing hmrC peaks (green)(see also Materials and Methods). These analyses are similar to those reported in Fig. 2B, the histogram being restricted to only 3 classes for easy comparison. Firstly, while around half of highly abundant transcripts harbored hmrC, this was not the case for all such transcripts. Secondly, a significant proportion of transcripts present at medium to low levels did contain hmrC. Thus, our hMeRIP-Seq data do not merely reflect nonspecific co-precipitation of abundant fragments. (B) hmrC peaks are found in transcripts of both highly and lowly expressed genes. As depicted, transcripts of highly expressed genes frequently contain hmrC (green), but less frequently than expected on the basis of a random distribution (gray). Importantly, transcripts of less expressed genes still harbor hmrC (green), and this occurs more frequently than expected on the basis of a random distribution (gray). For each transcript abundance quartile (from low to high) the number of hmrC peaks identified by

hMeRIP-seq is plotted in green, while the number peaks expected on the basis of a random model is reported in gray. Within a category (quartile), the difference between the expected value and the observed value was considered statistically significant at P-value

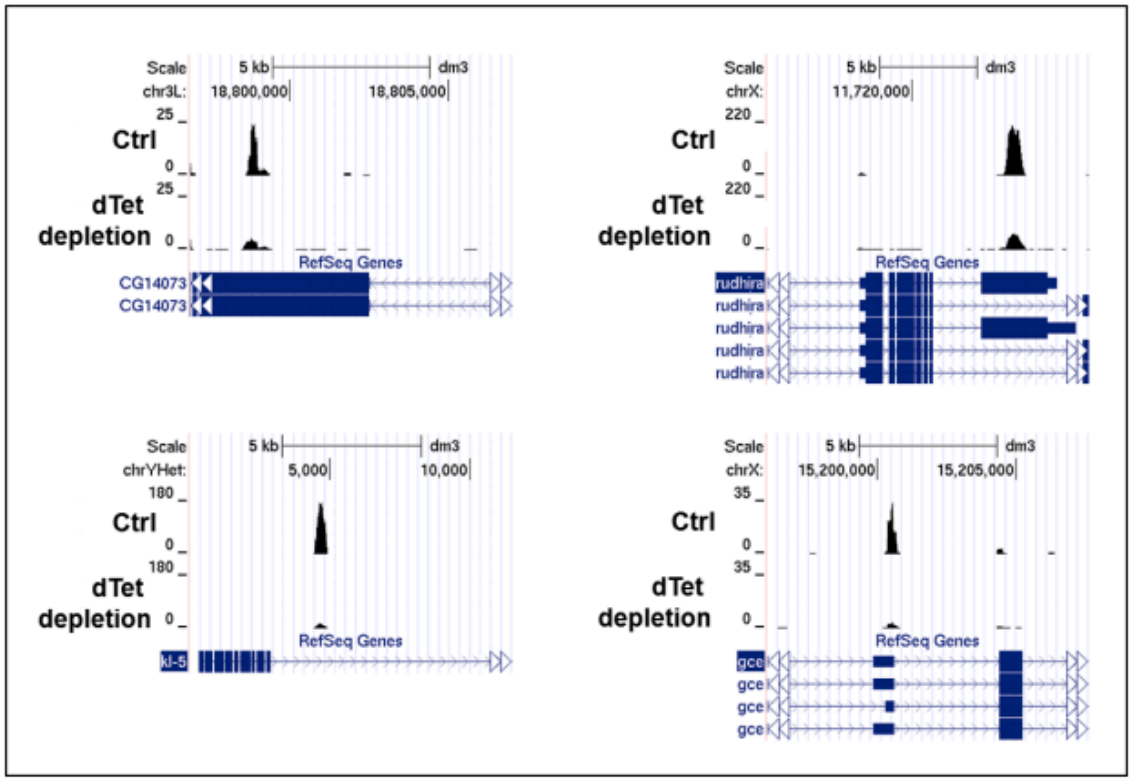


Figure 10 (related to Fig. 2C): Examples of enrichment profiles (UCSC tracks) of hMeRIP-Seq done in control and dTet-depleted cells.

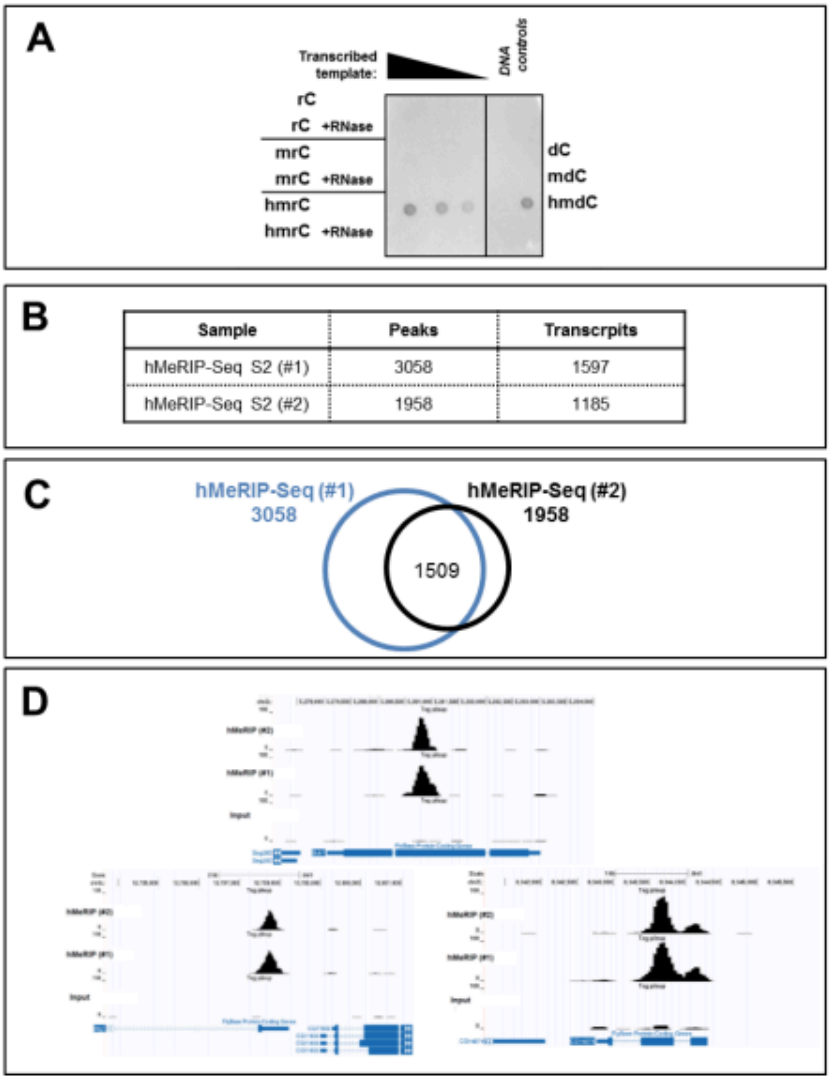
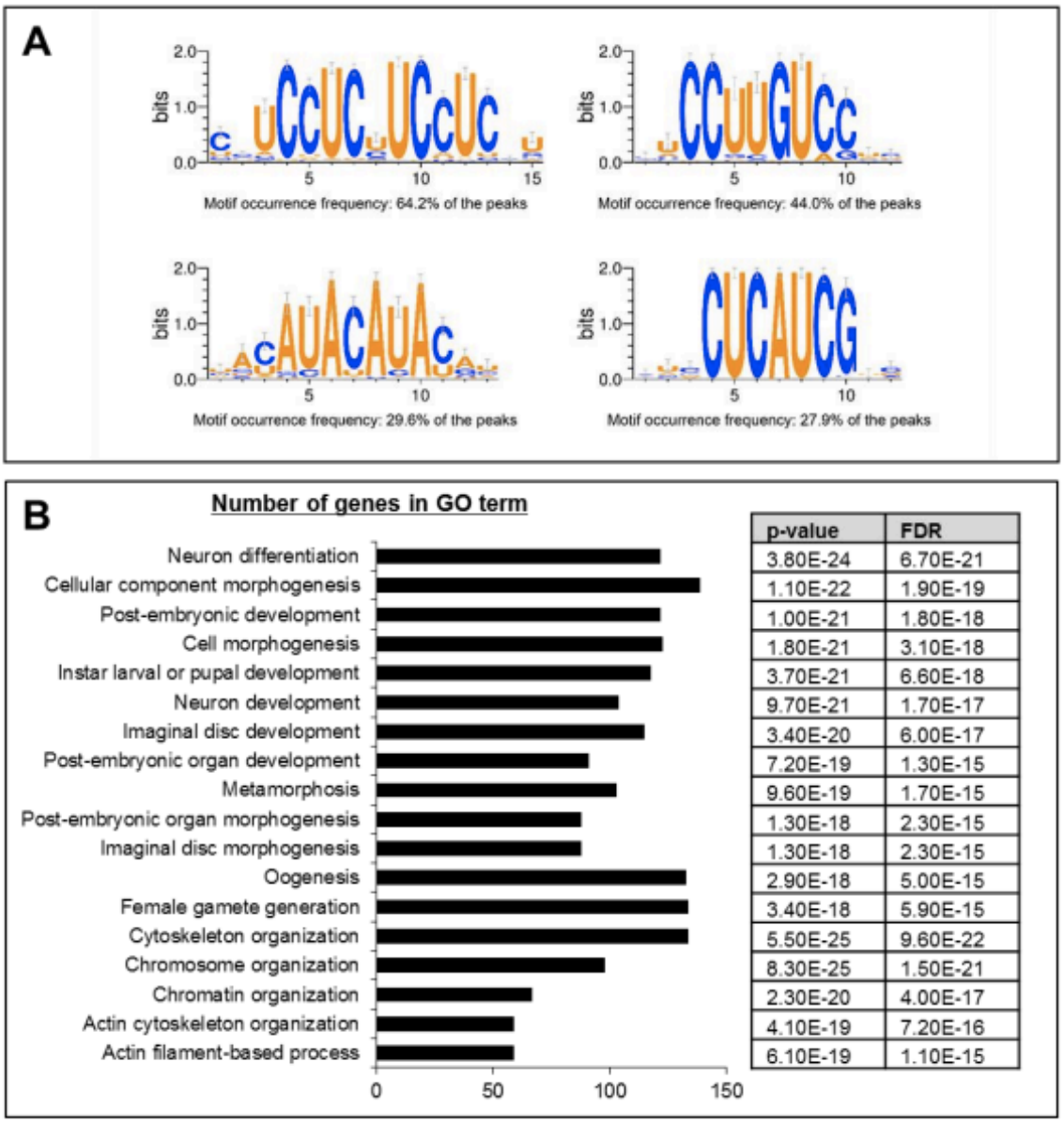


Figure 11 (related to Fig. 2): High agreement between hMeRIP-Seq experiments performed with two distinct hmC antibodies.
(A) Validation of hmC antibody #2 (Active Motif) specificity and sensitivity for hmCcontaining RNA. The hmrC content of decreasing amounts of transcribed templates containing either rC, mrC, or hmrC was detected by dot blotting with hmC-specific antibody. PCR controls are shown on the right. A blot representative of two experiments is shown. (B) hMeRIP-Seq in S2 cells reveals the presence of hmrC in many transcripts. Shown are the total numbers of statistically significant (P-value



p-value	FDR
3.80E-24	6.70E-21
1.10E-22	1.90E-19
1.00E-21	1.80E-18
1.80E-21	3.10E-18
3.70E-21	6.60E-18
9.70E-21	1.70E-17
3.40E-20	6.00E-17
7.20E-19	1.30E-15
9.60E-19	1.70E-15
1.30E-18	2.30E-15
1.30E-18	2.30E-15
2.90E-18	5.00E-15
3.40E-18	5.90E-15
5.50E-25	9.60E-22
8.30E-25	1.50E-21
2.30E-20	4.00E-17
4.10E-19	7.20E-16
6.10E-19	1.10E-15

Figure 12 (related to Figure 2): Overrepresented motifs in hMeRIP peaks and Gene Ontology analysis of hmrC-containing transcripts.
(A) Motif logos of the discovered motifs in hMeRIP peaks. For each motif, the percentage of peaks showing at least one occurrence of the motif is given. What might be the functional significance of the UC-rich motif, the most abundant identified motif (top left)? Interestingly, this motif resembles a known exonic splicing enhancer (ESE) sequence highlighted in Drosophila (30). ESE sequences are known to direct, or enhance, accurate splicing of pre-mRNA to messenger RNA. During splicing, these sequences are bound by positively acting factors, many of which are members of the serine-arginine-rich (SR) family of proteins (31, 32). In vitro assays have shown that a UC-rich motif similar to the sequence

identified here can function as a splicing enhancer (30), although no RNA binding proteins associated with this motif have yet been characterized. Thus, while our 21 work provides evidence for an effect of RNA hydroxymethylation on mRNA translation (Fig. 3), one needs not assume that dTet-deposited hmrC affects RNA biology in only one way. Transcript splicing is another level at which hmrC might play a role, as suggested by our transcriptome-wide analysis of hmrC, showing that about 17% of the hydroxymethylated regions lie within introns (Fig. 2D). In support of the notion that hmrC might affect RNA function at multiple levels, adenosine RNA methylation (m6A, an abundant modification in polyadenylated mRNAs) is also located in both exons and introns and is known to affect both RNA translation and splicing (33, 34). (B) Gene Ontology analysis was performed on hmrC-containing transcripts as identified by hMeRIP-Seq in S2 cells. Gene Ontology analysis highlights pathways involved in a variety of basic cellular processes, and more specifically in processes related to embryogenesis, morphogenesis, and other developmental categories.

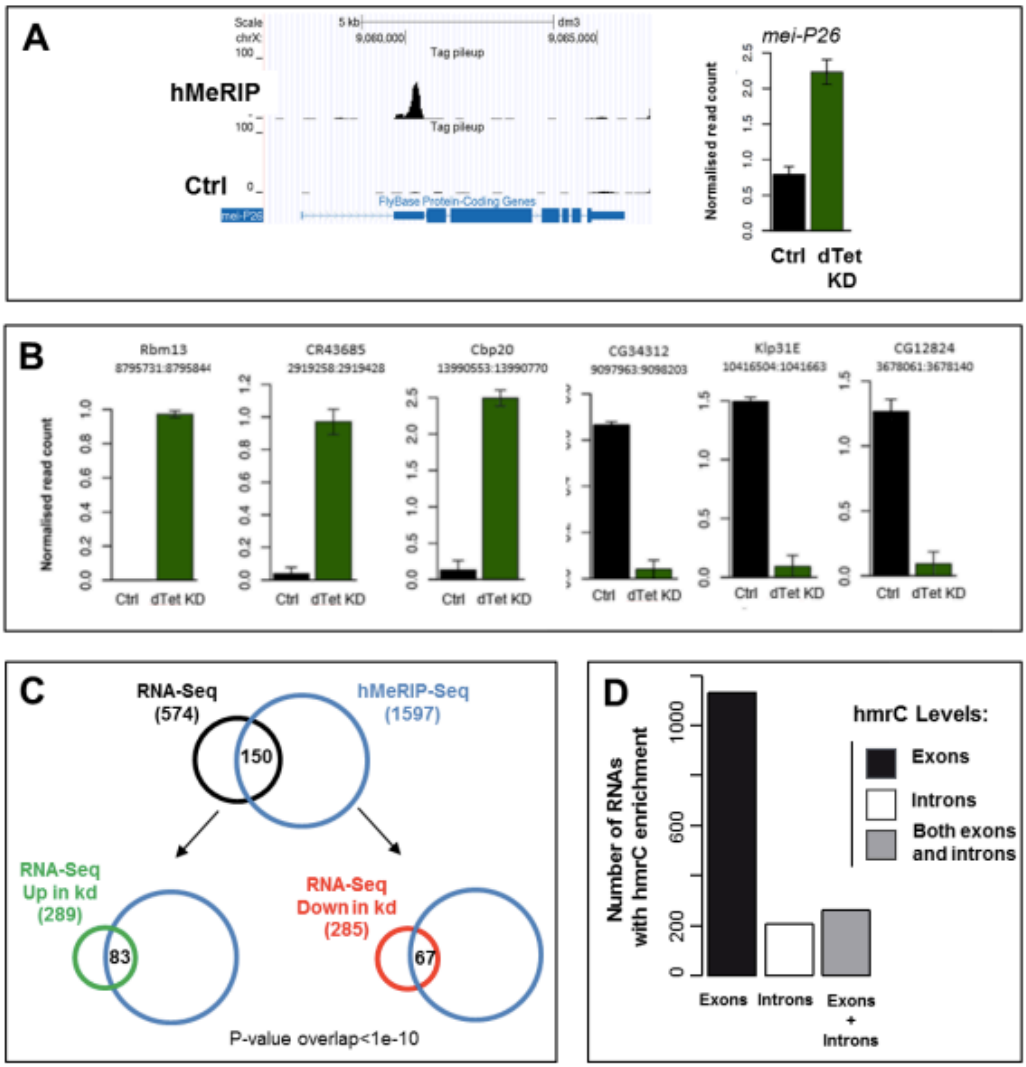


Figure 13 (related to Fig. 2F,G): Relationship between hmrC-containing RNAs and gene expression.
(A) UCSC browser representation of an example target showing enrichment for hmrC- (left) as well as differential expression between dTet KD and control S2 cells (right). (B) Examples of differentially expressed mRNAs in dTet KD identified by RNA-Seq in S2 cells. The normalized expression for the most differentially quantified exons are shown for both Ctrl (black) and dTet KD (green) samples. (C) mRNAs that are up- or down-regulated in the overlap between hmrC-containing transcripts and mRNAs whose levels are altered upon dTet knockdown. Based on data presented in Figure 2G, similar analyses were performed to determine the number of mRNA that harbor hmrC and that are up-regulated or down-regulated. As shown, 83 hmrC-overlapping mRNA (55.3%) are up-regulated while 67 (44.7%) are down-regulated respectively upon dTet knockdown. Since dTet can also modify DNA methyladenine (11), could the decreased hmrC levels observed in dTet- 23 depleted cells be indirect? By

comparing the hmrC-containing transcripts (identified by hMeRIP-Seq) with the dTet-regulated mRNAs (as determined by RNA-Seq), we found over a quarter of the dTet-regulated mRNAs to harbor RNA hydroxymethylation, being equally up- or down-regulated (Figure 2F and Figure S9C). Hence, the decreased hmrC level observed in dTet-depleted cells cannot be due to a generally decreased availability of RNA substrates for dTet (e.g. through dTet-mediated DNA m6A modification). Our data thus favor a direct impact of dTet on the hmrC levels. (D) Proportion of hmrC-enriched regions containing exonic peaks as compared to intronic peaks. We compared the observed proportion of hmrC peaks in exonic regions vs intronic ones. The bar plot displays the number of mRNAs with hmrC signal identified in exonic part exclusively (black bar), intronic part exclusively (white) or in both exonic and intronic parts (gray). For a high proportion (75 %) of mRNAs showing hmrC enrichment, signals are found in exonic part only.

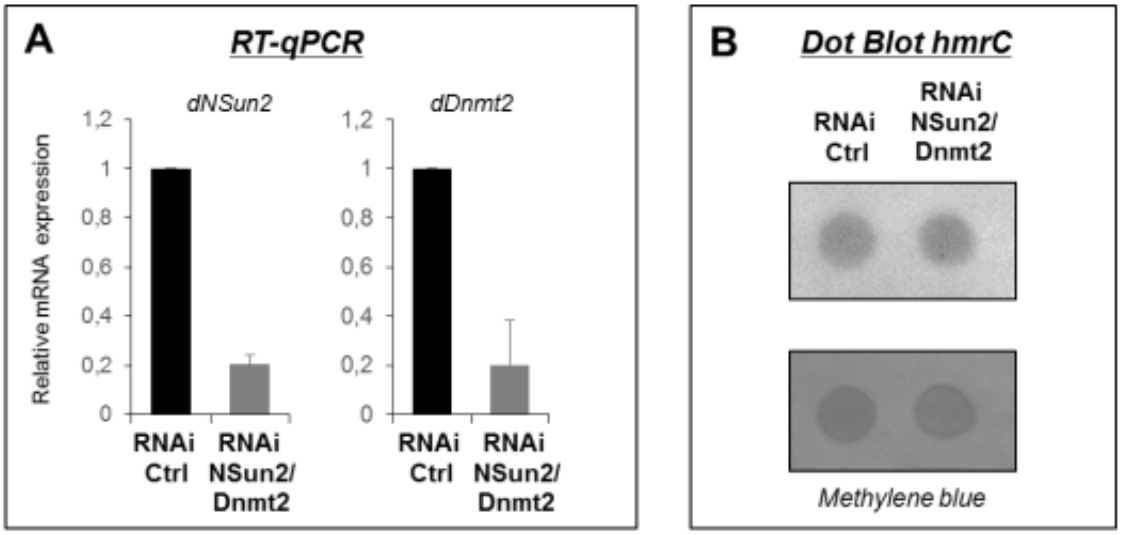


Figure 14 (related to Figure 2): The RNA methyltransferases NSun2 and Dnmt2 do not seem to impact on hmC within Drosophila RNA.
Given the presence of a CXXC Zn-Finger domain in dTet, it is still possible that dTet might affect gene expression via this domain, independently of its ability to hydroxymethylate RNA (cf. Main text). To assess the role of the CXXC motif of dTet, we sought to use a similar approach as described for Tet1, i.e. knockdown of Tet1 in Dnmt TKO cells and measure of transcription (15). Two RNA methyltransferases have been described in Drosophila: Dnmt2 (27) and NSun2 (35). We sought to knock them down in S2 cells and assess their impact on RNA hydroxymethylation. Despite efficient knockdown of Dnmt2 and NSun2 in S2 cells (cf. panel A), immunoblotting revealed no changes in hmCcontaining RNA (panel B). Hence, other RNA methyltransferases should be seeked to evaluate their impact on hmrC in Drosophila. It follows that the above approach proved inappropriate for analyzing the role of the CXXC motif in dTet-mediated regulation of transcription. (A) Efficient double knockdown of NSun2 and Dnmt2 in S2 cells. Cells were treated with the indicated siRNA (control or NSun2 and Dnmt2). Quantitative real-time PCR analysis of NSun2 and Dnmt2 mRNA is shown. Data are mean ± SD (N=4). (B) Immunoblotting after Dnmt2 and NSun2 knockdown shows no change in hmrC. The hmC content of total RNA extracted from control S2 cells and cells depleted of both NSun2 and Dnmt2 was detected by dot blotting with hmC antibody. Loading control was evaluated by methylene blue staining of the membrane.

V Materials and Methods

Cell culture

S2 cells (Invitrogen) were maintained in serum-free medium (Express Five; Life Technologies) supplemented with 18 mM L-glutamine at 24°C. Mouse embryonic stem cells (mESCs) (used in Figure S1) were maintained in DMEM (GIBCO) supplemented with 15% fetal bovine serum (GIBCO), 1% penicillin/streptomycin (GIBCO), 1% non-essential amino acids (GIBCO), 1% Glutamax (Invitrogen), 1% sodium pyruvate (Invitrogen), 0.2% β-mercaptoethanol (GIBCO) and 2000 U/ml recombinant leukaemia inhibitory factor. ESCs were cultured on irradiated feeder cells for 2 to 3 passages, then cultured without feeder cells on pre-gelatinized culture dishes for 2 additional passages before being collected.

Dot blotting for 5-hydroxymethylcytosine quantification in RNA.

RNA was extracted from S2 cells, Drosophila larval brains, ovaries or salivary glands with the RNeasy Mini or Midi Kit (Qiagen) according to the manufacturer's instructions. RNA was then spotted onto a nylon membrane (GE Healthcare Hybond-N+). The membrane was dried and then cross-linking was performed twice with 200 000μJ/cm2 UV. The membrane was stained with 0.04% methylene blue in 0.5M sodium acetate for RNA quantification and rinsed with PBS + 0.1% Tween-20 for 5 min. The membrane was then blocked in 3%(w/v) non-fat dry milk in PBS + 0.1% Tween-20 for 1h before transfer into blocking solution supplemented with rat anti-5hmrC monoclonal antibody (Diagenode MAb-633HMRC) diluted 1:500 and incubated overnight at 4°C. Thereafter, it was washed 3 times with PBS + 0.1% Tween-20 for a total of 30 min. The membrane was transferred into blocking solution supplemented with HRP-linked anti-rat IgG

(Abcam Ab6734) diluted 1:1000 before incubation for 1 h at room temperature followed by 3 washing steps with PBS + 0.1% Tween-20. Peroxidase activity was detected with ECL (Western Lightning Plus-ECL, Perkin-Elmer) or SuperSignal West Femto Chemiluminescent Substrate (Thermo Scientific) and signal detection was done with the ChemiDoc XRS system (Biorad). ImageJ software was used for signal quantification. For RNase controls, RNA samples were incubated with 0.7 U RNase A (Qiagen) at 56°C for 30 min. DNA controls were produced by PCR amplification of a 140bp fragment of the human actin gene in the presence of deoxyribo-CTP (Promega), -mCTP (Zymo Research) or -hmrCTP (Zymo Research), with the MethylTaq polymerase (Diagenode). All indicated replicates are from independent biological experiments, except for Figs. 1D, 4A and S10, for which error bars represent standard deviations from biological (at least in duplicates) as well as technical replicates. Normalization for dot blotting was done relative to: hmrC value using 1 µg of RNA set at 1 (Figure 1A); methylene blue and hmrC value for total RNA set at 1 (Figs. 1B, 1C); methylene blue (Figure 4A); and methylene blue and hmrC value for Ctrl or wt RNA set at 1 (Figs. 1D and 4E, respectively).

Dot blotting for RNA 5-methylcytosine or DNA 5-hydroxymethylcytosine.

Detection of mrC after polysomal fractionation (Figure S12) was performed similarly to dot blotting for hmrC, except for the following: blocking was performed in 5% (w/v) bovine serum albumin and 5% (v/v) horse serum in PBS + 0.1% Tween-20 for 1h before the membrane was transferred into blocking solution supplemented with mouse anti-5mC monoclonal antibody (Diagenode C15200003) diluted 1:500 and incubated overnight at 4°C. The 5 secondary antibody incubation was performed with HRP-linked anti-mouse IgG (GE Healthcare NXA931) diluted 1:3000 for 1 h at room temperature.

For dot blotting of DNA 5-hydroxymethylcytosine quantification (Figure S1), DNA was extracted with the QIAamp DNA Blood Mini Kit (Qiagen) according to the manufacturer's instructions. Detection of hmrC was then performed by dot blotting, similarly to hmrC dot blotting, except for the following: blocking was performed in 5% (w/v) non-fat dry milk in PBS + 0.1% Tween-20 for 1h and then the membrane was transferred into blocking solution supplemented with rabbit anti-5hmrC polyclonal antibody (Active Motif 39769) diluted 1:1000 and incubated overnight at 4°C. Incubation with secondary antibody was performed with HRP-linked anti-rabbit IgG (GE Healthcare NA934V) diluted 1:5000 for 1 h at room temperature.

In vitro transcription of (hydroxy)methylated RNA template.

In vitro transcription was performed with the MEGAscript T3 or the T7 Transcription Kits (Life Technologies), according to the manufacturer's instructions. For methylated and hydroxymethlated transcripts, ribo-CTP nucleotides were replaced in the reaction with ribo-mCTP or ribo-hmrCTP (TriLink Biotechnologies).

RNA interference, RT-qPCR analysis and CRISPR/Cas9 genome-editing.

For RNA interference, double-stranded RNA (dsRNA) was produced with the MEGAscript RNAi kit (Life Technologies) according to the manufacturer's instructions. For dTet siRNA, S2 cells were incubated for 6 days with 10 µg/ml dsRNA (a 1:1 mix of 2 sequences targeting dTet expression, or a sequence targeting GFP expression for the control), then replated and incubated for one day at 1.5 106 cells/ml with 10 µg/ml dsRNA before further analysis. For NSun2 and Dnmt2 siRNA, S2 cells were plated at 106 cells/ml and incubated for 72h with 5µg/ml dsRNA. Knockdown efficiency was checked

by RTqPCR analysis as described previously (16). The results were normalized with respect to the RpL32 housekeeping gene. This gene was also used for normalization of RT-qPCR performed in Drosophila embryos (Figure 4A) and dTetnull flies (Figure S13D). RpL32 expression levels were stable in the various RT-qPCR assays, e.g. across the analyzed time points of fruitfly embryogenesis (Figure S13A). The level of dTet transcripts (or NSun2 and Dnmt2) in cells treated with the GFP-targeting control RNAi was set at 1. Primer sequences are shown in Table S1. For CRISPR/Cas9, used in hMeRIP-Seq upon dTet depletion (Figure 2C), Yellow- and dTet-mutant cells were generated using the CRISPR-editing tool as described in Bassett et al. 2013 (17). Briefly, the sgRNA sequence for dTet (5'- GAGTTCAAGGACGTTATGGCCGG-3') was inserted in the BspQI site of the pAcsgRNA-Cas9 vector (Addgene, plasmid 49330). S2 cells were transfected with the resulting pAc-dTetsgRNA-Cas9 construct or the pAc-y1sgRNA-Cas9 control plasmid (Addgene, plasmid 49331) for targeting of the Yellow gene, and selected in 5 mg/ml puromycin. For genotyping, genomic DNA was extracted using the QIAamp DNA Mini Kit (Qiagen). The region surrounding the targeted dTet site was amplified by PCR using the following primers: Forward: TACCACAGAACACACCGGC; Reverse: AATCCCAATCCCGACTGCAC. PCR product was purified from agarose gel and ligated in a pCRII TOPO TA vector and used to transform Top10 bacteria (Invitrogen). DNA was extracted from bacteria colonies and sent for subsequent sequencing to determine the presence of an insertion or deletion event (Beckman Coulter Genomics). The resulting 6 sequences were aligned using the NCBI nucleotide Blast. For dTet-targeted mutation, all four clones sequenced showed frameshift-inducing insertion/deletion at the expected site.

<u>Enrichment in RNA species.</u>

For polyA RNA enrichment, GeneElute™ mRNA Miniprep (Sigma) was used according to the manufacturer's instructions. Isolation of small RNA was performed with the miRNeasy Mini kit (Qiagen) and the RNeasy MinElute Cleanup kit (Qiagen) according to the manufacturer's instructions. In the remaining fraction, polyA+ RNA was separated from ribosomal RNAs with the GeneElute™ mRNA Miniprep (Sigma) according to the manufacturer's instructions. All enrichment efficiencies were then estimated with Bioanalyzer RNA chips (Agilent).

HydroxyMethylated RNA Immunoprecipitation ("hMeRIP").

The procedure was performed in S2 cells essentially as described in Dominissi et al., 2013 (18). Briefly, 1 mg of total RNA (1 µg/µl) was chemically fragmented by incubation for 40 seconds at 94°C in fragmentation buffer (100 mM Tris-HCl pH=7, 100 mM ZnCl2). EDTA was added (50 mM) to stop the reaction. RNA was then ethanol precipitated, resuspended in RNase-free ddH2O and the fragmentation efficiency was checked on a Bioanalyzer RNA chip (Agilent). For immunoprecipitation, RNA fragments were denatured by heating at 70°C for 5 min, chilled on ice for 5 min, then incubated overnight at 4°C with or without 12.5 ug of anti-5-hmrC antibody (Diagenode rat monoclonal MAb-633HMRC), in freshly prepared 1X IP buffer (50 mM Tris-HCl pH=7.4, 750 mM NaCl and 0.5% Igepal CA-630, RNasin 400 U/ml and RVC 2 mM) supplemented with protease inhibitors (cOmplete, Mini, EDTA-free, Roche). Samples were then incubated at 4°C for 2.5h with 60 µl equilibrated Dynabeads Protein G (Life Technologies), washed three times for 5 min with 1 ml IP buffer, and eluted by addition of 1 ml TriPure Reagent (Roche) followed by RNA extraction according to the manufacturer's instructions. Samples were then analyzed by deep-sequencing. For CRISPR/Cas9, used in hMeRIP-Seq upon dTet depletion (Figure 2C), we used the

CRISPR/Cas9 system in S2 cells which allowed to produce enough RNA to do hMeRIP-Seq.

Library preparation and deep-sequencing workflow.

Library preparation was done with the TruSeq ChIP Sample Prep Kit (Illumina) after reverse transcription of pulled-down RNA and synthesis of a second strand (NEBNext mRNA second strand synthesis module (NEB)). Briefly, 5 to 10 ng dsDNA was subjected to 5′ and 3′ protruding end repair. Then, non-templated adenines were added to the 3′ ends of the blunted DNA fragments. This last step allows ligation of Illumina multiplex adapters. The DNA fragments were then size selected in order to remove all unligated adapters and to sequence 200-300-bp fragments. 18 cycles of PCR were carried out to amplify the library. DNA was quantified by fluorometry with the Qubit 2.0 and DNA integrity was assessed with a 2100 bioanalyzer (Agilent). 6 pM of DNA library spiked with 1% PhiX viral DNA was clustered on cBot (Illumina) and then sequenced on a HiScanSQ module (Illumina). For RNA-seq, total RNA from RNAi-treated S2 cells (control or dTet KD) was used to obtain a fraction enriched in polyA+ RNA. The RNA in this fraction was fragmented as described above, before library preparation and deep sequencing.

hMeRIP sequencing analysis.

The BWA software (19) was used to map S2-cell hMeRIP sequencing reads to the Drosophila melanogaster genome (UCSC dm3). Reads not uniquely mapping to the reference genome were discarded. Raw data are available in the Gene Expression Omnibus database:
www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=mhmtoqggdviptuz&acc=GSE66090.

Local read enrichment was evaluated with the MACS software (20) with a peak P-value cut-off < 10-5). Individual category enrichment was evaluated by modeling the random fluctuation from the expected value (λ) estimated by means of the random model using a Poisson distribution P(λ). Statistical significance was then assessed by computing the Poisson P-value (over- and under- representation) of the observed number of peaks belonging to one category compared to the expected value. A category was considered significantly enriched when P-value < 10-5 . 8 RNA-seq analysis. Sequencing reads obtained by RNA sequencing were mapped to the Drosophila melanogaster genome (UCSC dm3) with the BWA software. Reads not uniquely mapping to the reference genome were discarded. Read density was then computed by removing duplicate reads and normalizing the local read count with respect to the total mapped read count. For all annotated transcripts found in the FlyBase gene track (UCSC dm3), exons were quantified using the averaged normalized read count (FPKM). The fold change in FPKM and the absolute FPKM difference between knockdown and control conditions were calculated. A gene was considered differentially expressed at transcript level when at least one exon of its transcript consistently showed an absolute fold change >2 for the duplicates (unpaired t-test P-value 0.75) according to the transcript abundance observed in RNA-Seq performed on S2 control cells. For each category, the number of peak-containing regions was compared to the expected number of regions. The expected values were computed using a uniform distribution of peak location and by taking into account the transcript abundances observed in the RNA-seq data. Enrichment in individual categories was evaluated by modeling the random fluctuation estimated using the random model using a Poisson distribution P(λ), as described above (cf. "Distribution of hMeRIP peaks").

RNA-seq and hMeRIP-seq overlap.

To assess the statistical significance of the overlap between hmrC-containing and dTet-regulated transcripts, the P-value of the overlap between the two sets was estimated using the hypergeometric distribution, with the total number of annotated mRNAs (20193) as the reference set size (see also Table S5).

Motif analysis in hMeRIP peaks.

To perform the motif analysis, the genomic regions corresponding to the peaks identified by hMeRIP-seq were first extracted with the Regulatory Sequence Analysis Tools (www.rsat.eu) 'fetch sequences' software from the Drosophila melanogaster (UCSC dm3) reference genome using the peak coordinates. The corresponding sequences were then submitted to the Regulatory Sequence Analysis Tools 'peak motifs' tool using Markovian models computed on peak sequences as background models and default 9 parameters. Only the motifs with the highest occurrence frequency were kept (top 4). The discovered motifs were then converted back to RNA motifs and plotted with the WebLogo software (http://weblogo.berkeley.edu).

In vitro transcription and translation.

In vitro transcription was performed with the MEGAscript® T3 or T7 Transcription Kits (Life Technologies). The manufacturer's instructions were followed, with the exception that the GTP concentration was lowered four-fold and CAP analog was added to the reaction at a 1:4 GTP to CAP analog ratio (NEB). For methylated and hydroxymethlated transcripts, ribo-CTP nucleotides were replaced in the reaction with ribo-mCTP or ribo-hmrCTP (TriLink Biotechnologies). In vitro translation was performed in nuclease-treated rabbit reticulocyte lysate (Promega) in the presence of [35S]-methionine (PerkinElmer,

0.4 mCi/ml; 43.48 TBq/mmol) according to the manufacturer's instructions. [35S]-methionine incorporation was measured by scintillation counting after protein precipitation with TCA. Alternatively, 35S-labeled proteins were revealed by SDS-PAGE followed by sodium salicylate gel staining and fluorography. Normalization was done as follows: each result is expressed as a percentage of the signal generated with unmodified F-Luc, after subtraction of the background count as measured for R-Luc alone.

Polysome Fractionation.

The procedure was performed according to Vindry et al. 2012 (21). Briefly, S2 cells were treated with cycloheximide (100 µg/ml, Sigma Aldrich) for 5 min and harvested in buffer containing 25 mM Hepes, 100 mM KCl, 5 mM MgCl2, 0.5% Nonidet P-40, 2 µg/ml heparin, and 10 µg/ml cycloheximide. Cell extracts were cleared by centrifugation at 12,000 × g for 10 min and loaded onto 15–50% linear sucrose gradients. Samples were ultra-centrifuged at 200,000 × g for 3 h in a SW40.1 rotor and subsequently fractionated with a gradient fractionator (Brandel).

Drosophila Genetics.

The dTet^{null} allele was generated by FRT site recombination between two PBac(SANSON *et al.*) (f05022 and f03062) insertions, obtained from Exelixis at Harvard Medical School (22). The dTet-lacking mutant chromosome is kept as a stock over a balancer chromosome carrying a dominant marker. While dTetnull animals are seen in expected number at the larval and pupal stages, the only adult flies that hatch from pupae carry the dominant marker (dTetnull/balancer, n >5000). The lethality rate was estimated on the basis of the observation of >5000 heterozygous survivors. A GFP-tagged genomic

dTet gene under the control of the endogenous promoter was created using the MIMIC insertion MI05009 (BL#38002) (23), injected with plasmid DGRC#1298 by BestGene Inc. Animals were selected by genotype and no animals were excluded. The animals were picked out at random out of a bottle containing between 100 and 300 animals of different genotypes/phenotypes. Analysis of mutant phenotype (Figure 4D and figs. S13C-E) was done by at least two investigators who were blind to the genotype of the sample. For quantification of medulla size (Figure 4D, right panel), the width of the medulla region was measured at 3 different sites on confocal projections of 20 brain lobes per genotype at 3 different sites (top, middle, bottom). Error bars represent 95% confidence intervals (P- 10 value < 2.4 10-9).

Immunofluorescence Staining.

Brains were fixed in 4% formaldehyde and stained. A Dpn-specific antibody (rabbit, 1:500, kind gift from Jan Lab) was used for detection of Dpn. F-actin was stained with Alexa Fluor 546-phalloidin (1:300, Invitrogen), and DNA with Hoechst (Invitrogen). The experiments in Figure 4B were done 3 times (5 brains each).

CHAPTER 3: DROSOPHILA TET IS REQUIRED IN NEURONS, MUSCLE PRECURSORS AND THE DEVELOPING OVARY

I. Preface

This manuscript has been submitted. My contributions to this project and paper were writing the paper and all of the experiments.

II. Abstract

The TET (Ten-eleven translocation) 1, 2 and 3 proteins have been shown to function as DNA hydroxymethylases in vertebrates and their requirements have been documented extensively. Recently, the Tet proteins have been shown to also hydroxylate 5-methylcytosine in RNA. 5-hydroxymethylcytosine (5hmrC) is enriched in messenger RNA but the function of this modification has yet to be elucidated. Due to the lack of Cytosine methylation in DNA, Drosophila serves as an ideal model to study the biological function of 5hmrC. Here, we characterized the temporal and spatial expression and requirement of Tet throughout Drosophila development. We show that Tet is essential for viability as Tet complete loss-of-function animals die at the late pupal stage. Tet is highly expressed in neuronal tissues and at more moderate levels in somatic muscle precursors in embryos and larvae. Depletion of Tet in muscle precursors at early embryonic stages leads to defects in larval locomotion and late pupal lethality. Although Tet knock-down in neuronal tissue does not cause lethality, it is essential for neuronal function during development through its affects upon locomotion in larvae and the circadian rhythm of adult flies. Further, we report the function of Tet in ovarian morphogenesis. Together, our findings provide basic insights into the biological function of 5hmrC, a modification that is likely also regulated by Tet proteins in other species, and may illuminate observed neuronal and muscle phenotypes observed in vertebrates.

III. Introduction

RNA modifications represent a newly discovered layer of regulation with great importance in development (FRYE AND BLANCO 2016). RNA contains more than 100 distinct modifications, most of which are in abundant noncoding RNAs (MACHNICKA *et al.* 2013). N6-methyladenosine (m6A) is the most abundant modification in mRNAs. It plays an important role in neuronal function, and in Drosophila it is also required for sex determination (FRYE *et al.* 2016). We have recently demonstrated the presence of an additional RNA modification, 5-hydroxymethylcytosine (5hmrC) on mRNA, which is regulated by the Drosophila Tet protein (DELATTE *et al.* 2016).

In vertebrates, the TET proteins (Ten-eleven translocation) were identified as DNA modification enzymes. They function as oxoglutarate- and iron-dependent dioxygenases that oxidize 5-methylcytosine (5mC) to 5-hydroxy-methylcytosine (5hmC) ((TAHILIANI *et al.* 2009; GU *et al.* 2011; HE *et al.* 2011; ITO *et al.* 2011). 5hmrC is versatile; it can serve as a mark itself, or it can serve as an intermediate leading to the TET-dependent removal of the methyl mark through a series of intermediates which themselves serve as marks(MELLEN *et al.* 2012; KOHLI AND ZHANG 2013). It has been shown that 5hmrC is present in gene bodies, where the modification positively correlates with gene expression levels (JIN *et al.* 2011). Gain of 5hmrC is usually correlated with the loss of H3K27me3 and activation of gene expression (HAHN *et al.* 2013). In mammals, the Tet family contains three members, Tet1, Tet2 and Tet3, that share a high degree of homology within their C-terminal catalytic domain. Tet1 and Tet3 also share a CXXC patterned DNA binding domain. The regulation of gene transcription by TETs is complicated and still needs to be elucidated, especially in view of the new role in RNA hydroxymethylation. Tet1 and Tet2 are highly expressed in mouse embryonic stem cells (ESCs). ESCs depleted for both Tet1 and Tet2 retain pluripotency but show reduced levels of 5hmC in DNA, but complete loss of

either Tet1 or Tet2 does not affect viability (DAWLATY *et al.* 2011; KOH *et al.* 2011). Tet3 homozygous mutant mouse embryos develop normally, but die at birth. This is apparently due to a defect in epigenetic reprogramming during zygotic development (GU *et al.* 2011). TET3 has been shown to catalyze the conversion of 5mrC to 5hmrC in tissue culture and mouse ESCs, similar to what we have reported in Drosophila (FU *et al.* 2014a).

In Drosophila, DNA methylation is only observed in early embryonic stages at barely detectable levels and is not believed to have a function in genome-wide regulation of gene expression(LYKO *et al.* 2000; CAPUANO *et al.* 2014). The fly genome does not encode any canonical DNA methyltransferase 1 or 3 (Dnmt1/3) homologs. Dnmt2 is the only known candidate DNA methyltransferase in Drosophila and has been demonstrated to modify specific t-RNAs. Dnmt2 null mutants are homozygous viable, and whole-genome bisulfite sequencing did not uncover Dnmt-dependent cytosine methylation patterns in Drosophila (SCHAEFER *et al.* 2010; RADDATZ *et al.* 2013). The enzyme(s) that control cytosine methylation (5mrC) in Drosophila mRNA are not known.

The fly genome contains one conserved Tet gene. We have shown that the 5hmrC modification is found in Drosophila polyA^{+} RNA and that upon Tet knock down (KD) in S2 Schneider cells, the level of 5hmrC is reduced by 50%. *Tet* is essential, as null mutants are lethal at the late pupal stage. 5hmrC is enriched in neuronal tissues and is most highly expressed in 3rd instar larval brains. 5hmrC was also reduced significantly in *Tet*^{null} brains (DELATTE *et al.* 2016). The transcriptome-wide distribution of RNA 5hmrC has been mapped in Drosophila S2 cells and shows that the modification targets specific transcripts. 5hmrC modified mRNA is found in association with polysomes, suggesting that this modification facilitates mRNA translation (DELATTE *et al.* 2016). The lack of 5mC and 5hmC on DNA makes Drosophila an excellence platform to investigate the importance of RNA 5hmrC.

Here we characterize the expression pattern of Tet and its functional requirement. Tet is expressed most highly in neuronal cells during development and at lower levels in muscle precursors during embryogenesis as well as in the larval imaginal discs. Tissue-specific and temporally controlled knock-down indicates that Tet function is required during embryogenesis for larval locomotion, ovary development, and during embryonic or larval stages in PDF neurons to control aspects of the adult circadian rhythm. Our identification of Tet requirements in Drosophila establish the importance of 5hmrC for muscle and neuronal function.

IV. Results

The Tet protein and its functional domains

The Drosophila genome contains a single *Tet* gene encoding 6 transcripts that together appear to encompass transcripts encoded by the three vertebrate genes. Two Drosophila transcripts encode only the catalytic domain, similar to the transcripts of the vertebrate Tet2 gene (*Tet-S*, Figure 15A). Similar to the gene products of vertebrate Tet1 and 3, the other four transcripts are larger and encode in addition to the catalytic domain the CXXC DNA binding domain (*Tet-L*, Figure 15A). The domain organization is maintained in the vertebrate and Drosophila proteins and the amino acid similarity is highly conserved in the active domains (DUNWELL *et al.* 2013). When we initiated this work, two deletions, *Df(3L)Exel6091* and *Df(3L)Exel6092*, that originate from the same site within the Tet coding region and extend in opposite directions were available (Figure 15A). In *trans,* these deficiencies were mostly lethal at the late pupal stage with ~ 20 % escapers that lived only for a few days, indicating that the *Tet* gene has an essential function. To further determine the importance of *Tet* we created a null allele that deletes the 5' end of all the transcripts. Homozygous Tet^{null} animals develop up to late pupal-pharate adult stages, but no flies eclose (DELATTE *et al.* 2016).

To investigate the functional requirements for the two different forms of Tet, we induced transcript-specific deletions within the Tet gene by FRT-mediated site-specific recombination of piggyBac transposons (A', B', and C' in Figure 15A). Recombination between A' and B' resulted in the deletion of the 5' end of Tet-L including the DNA binding domain (Tet^{DBD}). Recombination between B' and C' caused the deletion of the 5' end of Tet-S (Tet^{2}). Recombination between A' and C' caused the deletion of the 5' end of both isoforms (Tet^{null}). To confirm the effect of these deletions on the Tet transcripts, q-PCR was performed with RNA extracted from 3rd instar larval

brains using transcript specific primers (Figure 15B). *Tet*^{null} mutants have neither Tet-L nor Tet-S transcripts. *Tet*^{DBD} mutants do not produce any long transcripts but the short transcripts are present at about 50% of the level observed in WT. With the primers common to all transcripts, we observe 25% of total (Tet-L and Tet-S) transcripts (Figure 15B). *Tet*^{2} mutants show no short transcripts, but, while more than normal levels of fragments were amplified with the N-terminal primers, the C-terminal primers suggest that only about 50% of the transcripts were full-length (Figure 15B). *Tet*^{DBD}/*Tet*^{2} transheterozygotes and *Df(3L)Exel6091/ Df(3L)Exel6092* show similar results. About 50% of Tet-L and 5-10% of Tet-S, or 5 – 25 % of full length RNA can be detected (Figure 15B).

Despite the presence of varying levels of Tet transcripts, *Tet*^{DBD} and *Tet*^{2} over *Tet*^{null} or as homzygotes are lethal at the late pupae-pharate adult stage as *Tet*^{null}. The transheterozygous *Tet*^{DBD}/*Tet*^{2} animals show a 100% eclosion rate and the flies live for more than 2 weeks (Table 1), but they are sterile and have normal looking ovaries, full of eggs that are not laid. The *Df(3L)Exel6091/ Df(3L)Exel6092* heterozygotes show a stronger phenotype than *Tet*^{DBD}/*Tet*^{2}; about 20% of flies eclose, but die within one or two days, reflecting the lower level of Tet expression.

Table 1: The Tet alleles

Genotype	Survival stage	Adult survival rate
Df(3L)Exel6091/ Df(3L)Exel6092	Late pupa	20% (escapers)
Tet^{null}	Late Pupa	0
Tet^{DBD}	Late pupa	0
Tet^{DBD}/ Tet^{null}	Late pupa	0
Tet^{2}	Late pupa	0
Tet^{2} / Tet^{null}	Late pupa	0

Tet^{DBD}/Tet^{2}	Adults	100%
Mi{Mic}$Tet^{MI03920}/Tet^{null}$	Adults with severe moving disability	100%

An additional Tet mutant, *Mi{Mic}*$Tet^{MI03920}$, carrying the Mi{Mic} insertion in the last exon of the *Tet* gene (Figure 15A), shows 100% eclosion rate over Tet^{null}. The flies have extended wings (Figure 24A), exhibit strong defects in coordinated motor control and die within 1-2 days after eclosion (Table 1). This indicates that *Mi{Mic}*$Tet^{MI03920}$ *is* a hypomorphic allele.

Expression of Drosophila Tet

Through cassette exchange with a Mi{Mic} insertion we generated a line expressing a Tet-GFP fusion protein under the control of the endogenous Tet promoters, *Tet-GFP* (called *dTet-Mi* in (DELATTE *et al.* 2016)). *Tet-GFP* shows no phenotype over Tet^{null}, and is homozygous viable, indicating that the GFP-Tet fusion protein is functional. We find that Tet-GFP is not maternally contributed and is first detected at the germ band extension stage, mainly in the central nervous system (Figure 16A, Figure 21B), consistent with *Tet* mRNA expression (Figure 21A). Its expression overlaps largely with the postmitotic neuronal marker, ELAV (Figure 16B, C). Tet is also expressed, albeit at lower levels, in the somatic muscle precursors as they fuse into myotubes and migrate towards the dorsal aspect of the embryo. Tet is enriched in nuclei, and its expression overlaps with the nuclear mesodermal cell marker Mef2 (Figure 16C). Tet-GFP expression subsides in muscle cells at stage 16 shortly before hatching (Figure 16D), but the expression in the central nervous system is continuously present (Figure 16E). This temporal expression during embryogenesis is consistent with *Tet* RNA expression described previously (Figure 21) (DELATTE *et al.* 2016).

In addition to the CNS, Tet-GFP is present in various imaginal discs. In wing discs, Tet expression is strongest in the proneural clusters (GARCIA-BELLIDO AND DE CELIS 2009) and lower in myoblasts that develop into direct flight muscles in adults (Figure 16A). In the eye discs, Tet expresses in all neuronal cells of the developing photoreceptor cells posterior to the morphogenetic furrow (Figure 16A). Taken together our results show that Tet is most highly expressed in the neuronal lineages throughout development and is also present in the somatic muscle precursors. This expression pattern is consistent with our observation that Tet is required in neural and muscle cells.

Tissue-specific and stage-specific requirement of *Tet*

To further address the tissue-specific requirement of Tet, we depleted Tet in specific tissues and cell types using the UAS-Gal4 system. We have obtained four different *Tet*-RNAi lines (VDRC) (DIETZL *et al.* 2007). We first tested the knock-down (KD) efficiency of these different RNAi lines in RNA from 0-12h embryos. The RNAi expression was driven by the general *da-gal4* driver. RT-PCR was performed to estimate the mRNA KD. One of the lines (#v102273, we call Tet^{RNAi}) was selected for future experiments because we observed ~85% KD in mRNA levels (Figure 22A, B) and the animals die at late pupal stages similar to *Tet* null mutants. The other lines showed no lethality and more moderate KD of Tet mRNA, around 50%-70% (Figure 22A). When $UAS\text{-}Tet^{RNAi}$ was combined with other ubiquitous drivers, such as tub-gal4 or act5c-gal4, the animals also died before eclosion (Table 2). The broadly expressed mesodermal drivers mef2-gal4 or how24B-gal4 (with dcr2 to achieve a stronger KD) driving Tet^{RNAi} also cause lethality similar to that observed in null mutants. Other muscle drivers that express at later stages, such as G14-gal4 and act88F-gal4, do not cause lethality (Table 2). This suggests that the requirement of Tet

in muscle development occurs during early stages of development. Notably, Tet KD under the control of several neuron drivers (Table 2), did not result in lethality. This indicates that Tet is essential for the function of muscles, but does not rule out that Tet could also affect neuronal functions.

Table 2: Phenotypes of Tet KD with different drivers.

Driver name	Expression pattern	Lethality with Tet RNAi
da-gal4	Ubiquitous; maternal	Just before eclosion
act5C-gal4	Ubiquitous expression with an early onset	Just before eclosion
Mef2-gal4	Mesoderm, embryonic stage 12 myoblast and larval muscles	Just before eclosion
how24B-gal4	Mesoderm and larval muscles	Just before eclosion with dcr2 (viable otherwise)
G14-gal4	Muscle cells from stage 11 to larvae (2,3,4,5,6)	viable
act88F-gal4	Wing flight muscle	viable
hand-gal4	Cardiac lineages	viable
tin-gal4	heart	viable
elav-gal4	All postmitotic neurons with an early onset	viable with dcr2
OK6-gal4	all motor neurons, salivary glands, wing discs and a subset of tracheal branches commencing the first instar larval stage and persisting until pupation	viable
nrv2-gal4	Nervous system-specific expression from embryo to the adult stage	viable

***Tet* and locomotion**

We examined the *Tet*^{null} embryonic nervous systems and 3rd instar larval brains as compared to WT using cell-type specific markers and found no substantial difference in organization or morphology. This lead us to investigate potential functional defects in the nervous system. Locomotion behaviors are coordinated by neurons, muscles, and neuromuscular junctions. Given the Tet expression pattern and the effects of tissue specific knockdown it seemed a possible biological endpoint to assess. Therefore, we set up tests to study the locomotion of third instar larvae. The external morphology of WT and mutant 3rd instar larvae is indistinguishable (Figure 23A). We examined the movements of cohorts of larvae in 1 minute movies, and quantified total distance traveled and crawling speed and observed that *Tet* mutants displayed a loss of motility and contractions. *Tet*^{null} larvae showed >50% reduction in body wall contractions and crawling speed (Figure 17A, B). The partial loss of function alleles *Tet*^{DBD} (62%) and *Tet*^{2} (68%) exhibited clear, but less severe defects in both aspects of locomotion (Figure 17A, B). The differences were statistically highly significant leading to the conclusion that Tet is required for normal locomotion.

We further investigated if *Tet* function is required in neurons or muscles for normal locomotion. We performed the locomotion assays on third instar larvae in which Tet was depleted either in muscles (mef2-gal4; Tet^{RNAi}), in neurons (elav-gal4; Tet^{RNAi}), or in a motor neuron subset (OK6-gal4; Tet^{RNAi}). Mef2-gal4 driven Tet KD larvae showed the most severe defect similar to that observed in *Tet*^{null} animals (Figure 17E, F). Elav-gal4 Tet KD displayed a less severe, but significant defect (Figure 17C, D), whereas OK6-gal4 KD did not alter the locomotion behavior (Figure 17K, L). These results indicate that Tet function is required both in neurons and muscles but not in motor neurons.

Next, we examined the temporal requirement of Tet in locomotion by using the temperature-sensitive *GAL80*^{ts} system (ZEIDLER *et al.* 2004). GAL80 suppresses the activity of the

transcriptional activator GAL4, preventing expression of genes controlled by an UAS promoter, in this case, *UAS-Tet*^{RNAi}. Raising the temperature to restrictive levels inactivates GAL80^{ts} and permits expression of the *Tet*^{RNAi}. We used this system to inactivate *Tet* during different developmental stages.

We observed a significant defect in locomotion when embryos were deprived of Tet function during embryogenesis. In mef2-gal4 induced knockdown embryos reared at the restrictive temperature 0-4 h after egg lay (AEL), significant reductions of both contractility and overall motility were observed (Figure 17G, H). However, if we reduced Tet function later, when animals are between 24 h and 28 h old (first instar), the variance of the measurement became much greater, so that the defects became insignificant (Figure 17I, J). This result indicates that Tet is required during embryogenesis for larval locomotion, but is dispensable for locomotion in 1st instar larvae.

To gain functional insight into the basis of this defect, we tested for a possible role of Tet in synaptic structures by analyzing the neuromuscular junctions (NMJs) of motor neurons terminating on the somatic musculature of third instar larvae. No significant changes in the number of NMJ boutons or their morphology were observed in *Tet*^{null} mutants (Figure 23B, C).

***Tet* and circadian rhythm**

We were interested in determining if Tet was also required in additional neuronal functions. The circadian rhythm sets the day – night clock in most organisms regulating many physiological processes and is well studied in Drosophila. We decided to investigate if Tet is required for circadian behavior. The rhythm of locomotion is controlled by about 150 clock neurons in the central brain, organized into distinct subgroups. The *timeless* (*tim*) gene, expressed in all clock

neurons, is one of the major factors controlling the circadian rhythm (PANDA *et al.* 2002). Under light/dark (LD) cycles, fruit flies exhibit bimodal activity with morning and evening peaks. Small ventral lateral neurons (s-LNvs) expressing the pigment-dispersing factor (PDF) are necessary and sufficient for morning activity (GRIMA *et al.* 2004) (STOLERU *et al.* 2004).

To evaluate if Tet functions in these processes, we examined adult behaviors in flies with cell-specific KD of Tet. 3-5 d old males were trained across 12h Light-Dark (LD) cycles for 6 days, followed by constant darkness (DD). Control flies displayed two activity peaks, one before lights-on (morning anticipation), and the other before lights-off (evening anticipation), which were retained for 5 days under DD. When Tet was depleted under tim-gal4 control in all clock neurons and glia cells (tim-gal4, Tet^{RNAi}), the bimodal pattern was lost as the morning peaks were missing in Tet KD animals 2 days after switching to DD (Figure 18A). The morning peak is controlled by the PDF-positive s-LNvs neurons (PDF neurons). Therefore, we also examined animals in which Tet was depleted specifically in the PDF neurons (pdf-gal4, Tet^{RNAi}) and found that in these animals the morning anticipation was also disrupted (Figure 18B, C), while the total activities were not changed (Figure 24A).

We also investigated the temporal requirement of Tet in PDF neurons. The drug induced gene switch (GS) system provides temporal control of GAL4 (NICHOLSON *et al.* 2008). When we controlled KD of *Tet* using the pdf-GS-gal4 driver in adult males, no difference in the morning anticipation was observed between Tet KD and WT (Figure 18D).

We next investigated if we could detect any abnormalities in the pdf-positive neural network by staining adult brains with anti-PDF antibody of Tet KD flies. The PDF network looked indistinguishable in Tet KD and WT adult brains (Figure 18A). The expression of the PDF marker is normal in pdf-gal4 driven Tet KD brains, and the PDF neurons extended normal axonal and

dendritic projections. This rules out the possibility that Tet affects the circadian rhythm through impairing the gross anatomy of pdf neurons. To further address when Tet might function in PDF neurons, we examined Tet expression. We found that in embryonic stage 16, Tet-GFP is observed in the cytoplasm of PDF neurons (Figure 19D), but it is absent in 3rd instar larval or adult stages (Figure 19B, C). Our results show that Tet does not directly control the circadian rhythm, rather, that it controls the development of precursors into functional PDF neurons.

Tet requirement in ovary development

Gametogenesis in Drosophila starts in the early embryo when the primitive ovaries form on each side of the gut. During larval development the ovaries increase in volume about 50 fold. The groups of cells making up the larval ovary become subdivided into ~16 ovarioles by the migration of mesodermal cells. At the anterior end of an ovariole are the terminal filament cells, adjacent to the germ line stem cells. The ovarioles contain a series of maturing egg chambers from the youngest, to the mature eggs at the posterior end.

We examined the requirement for Tet in ovaries by characterizing the *Tet* phenotype. As *Tet*^{null} animals die as pharate adults, we examined WT and mutant ovaries dissected from larvae and pharate adults. The *Tet*^{null} larval ovaries contain all the cell types present in the WT ovary, however the organization is strongly disturbed (Figure 20B). Late larval WT ovaries show subdivision into ovarioles with the stacked terminal filament cells visible (Figure 20B). By pharate adult stage, the ovaries are similar to those of virgin females with separated ovarioles containing egg chambers that have developed to about stage 8 (Figure 20A). In *Tet*^{null} mutants, the ovarioles are not separated and no early egg chamber stages can be discerned. Moreover, the somatic portions of the ovary, such as stacked terminal filament cells cannot be clearly

identified. Nevertheless, at the posterior end of the ovary some egg chambers of about stage 8, as judged from their organization and size of their nuclei, can be identified (Figure 20A). *Tet*^{DBD} and *Tet*^{2} pupal ovaries displayed a similar phenotype with incomplete separation of ovarioles, but the phenotype is less severe. The organization of the ovary and subdivision into ovarioles while incomplete, can be detected (Figure 20A). The phenotypes observed in the larval and late pupal mutant ovaries indicate that the requirement for Tet in ovary morphogenesis occurs before the larval stages.

Next, we investigated where Tet is expressed in fly ovaries. Using the *Tet-GFP* line, we first examined the adult ovary and found no expression of Tet in either somatic or germline cells including germline stem cells. We also examined the earlier stages for Tet expression. We did not see specific expression in pole cells or the embryonic gonad in the Tet-GFP line, but we did observe weak nuclear expression in the terminal filament cells in the late 3rd instar larval gonads, which could at least in part explain the observed disruption of organization of the ovarioles in *Tet*^{null} pupal ovaries (Figure 20B).

To further investigate the requirement of *Tet* in ovary development, we induced clones by flippase recognition target (FRT)-mediated site-specific mitotic recombination (XU AND RUBIN 1993). Phenotypes were analyzed 10 or 20 days after clone induction (ACI). We first induced clones in 3rd instar larvae and found that both germ line and somatic (follicle cells) WT and *Tet*^{null} clones were similar in numbers and morphology (Figure 20D, Figure 25D). Next, we induced clones in 0-24h h embryos, which will only result in germ line clones. We detected reduced numbers of Tet germ line clones; in WT 17.8% of ovarioles contained clones compared to 8.4% of ovarioles that contained *Tet*^{null} clones (Figure 20C). However, the *Tet* clones did not show any visible phenotypes and the egg chambers looked similar to the WT egg chambers. This reduction

in clones indicated that there is a requirement for Tet function during embryonic development for ovary development.

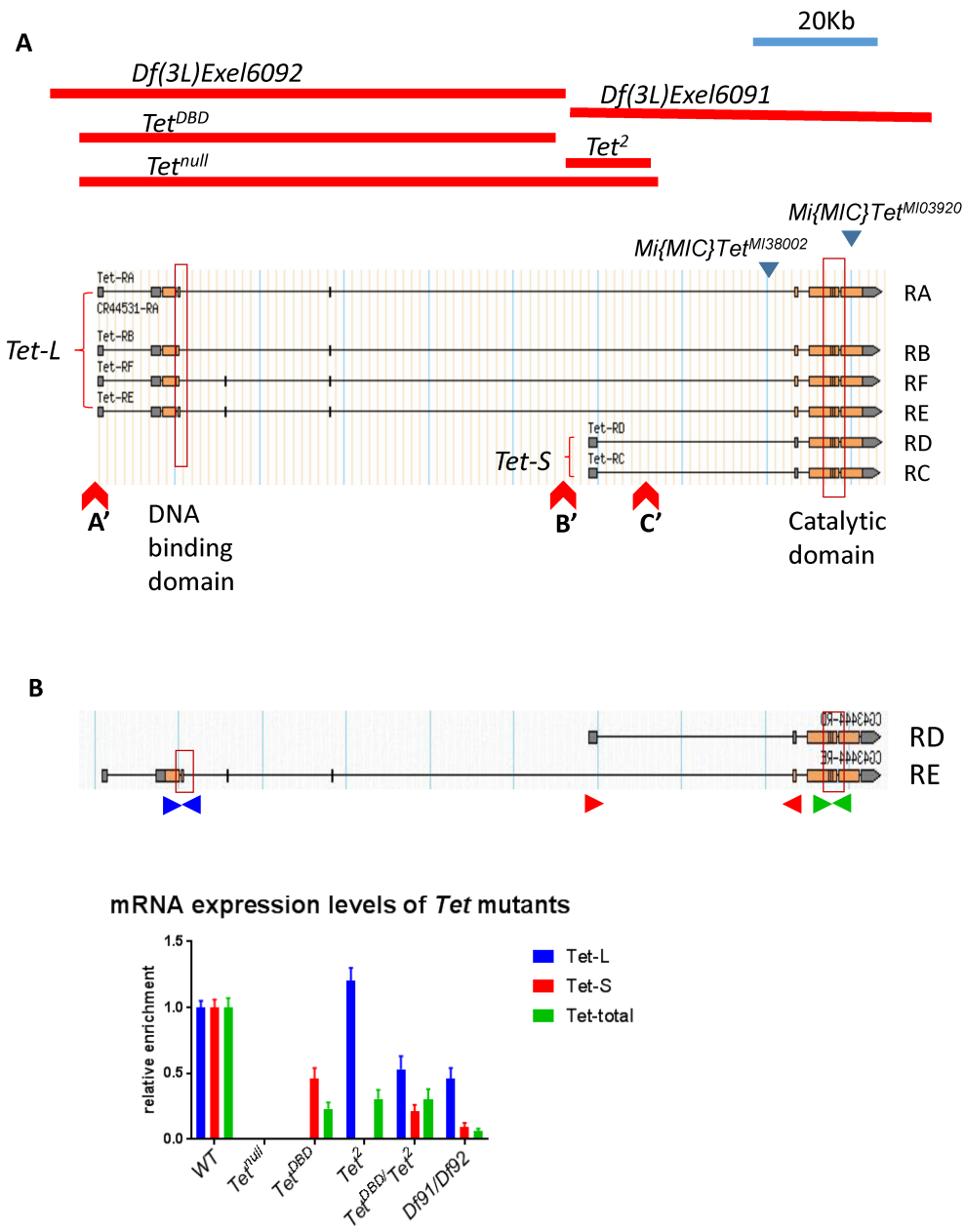
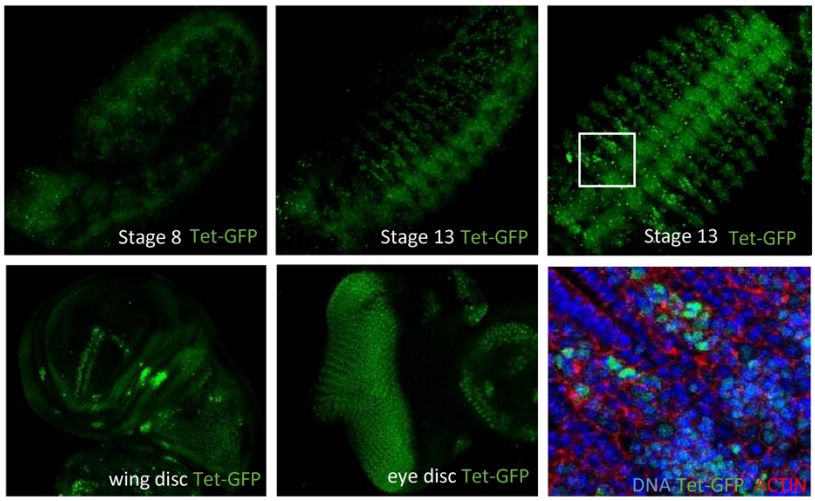


Figure 15 The Tet locus.

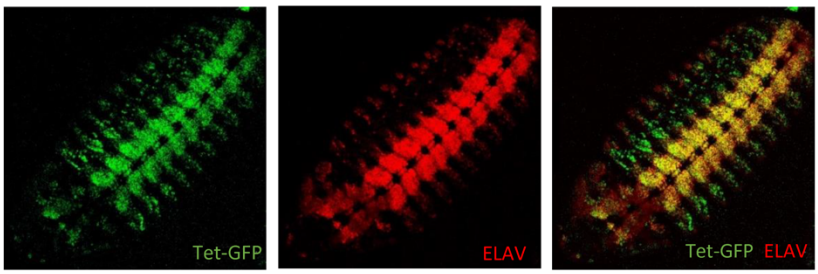
(**A**) The different Tet transcripts (Tet-L and Tet-S) and the location of the DNA-binding and catalytic domains (boxes), as well as the two Mi{MIC} insertions (blue triangles) are indicated. Red lines show the extent of the *Tet* deletions. Red Triangles mark insertion sites of the piggyback transposons used for induction of deletions. (**B**) The *Tet-L* and *Tet*-S transcripts and

their expression levels in WT and *Tet* brains. Primer pairs specific to the long transcript (TetL, blue), the short transcript (TetS, red) and all transcripts (green) were used. Df91/Df92 = *Df(3L)Exel6091/Df(3L)Exel609.* qRT-PCR results are normalized to *Rp49 (RpL32).*

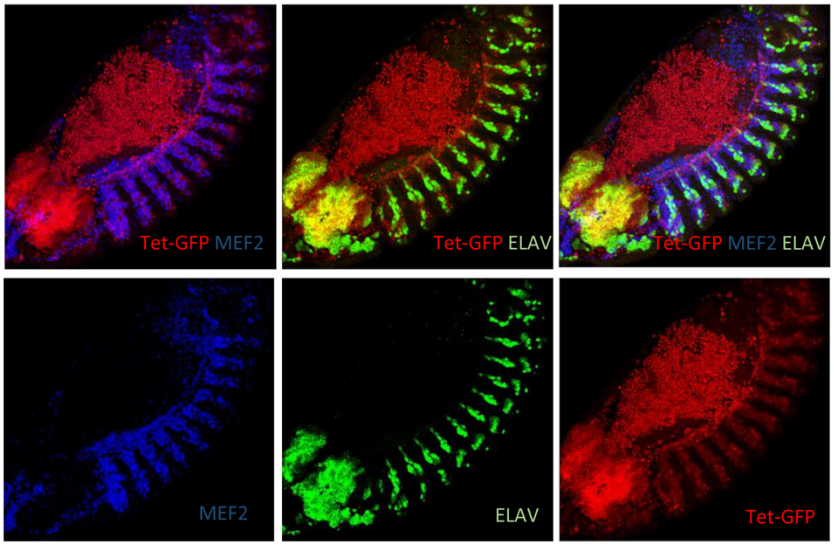
A


Stage 8 Tet-GFP
Stage 13 Tet-GFP
Stage 13 Tet-GFP
wing disc Tet-GFP
eye disc Tet-GFP
DNA Tet-GFP ACTIN

B


Tet-GFP
ELAV
Tet-GFP ELAV

C


Tet-GFP MEF2
Tet-GFP ELAV
Tet-GFP MEF2 ELAV
MEF2
ELAV
Tet-GFP

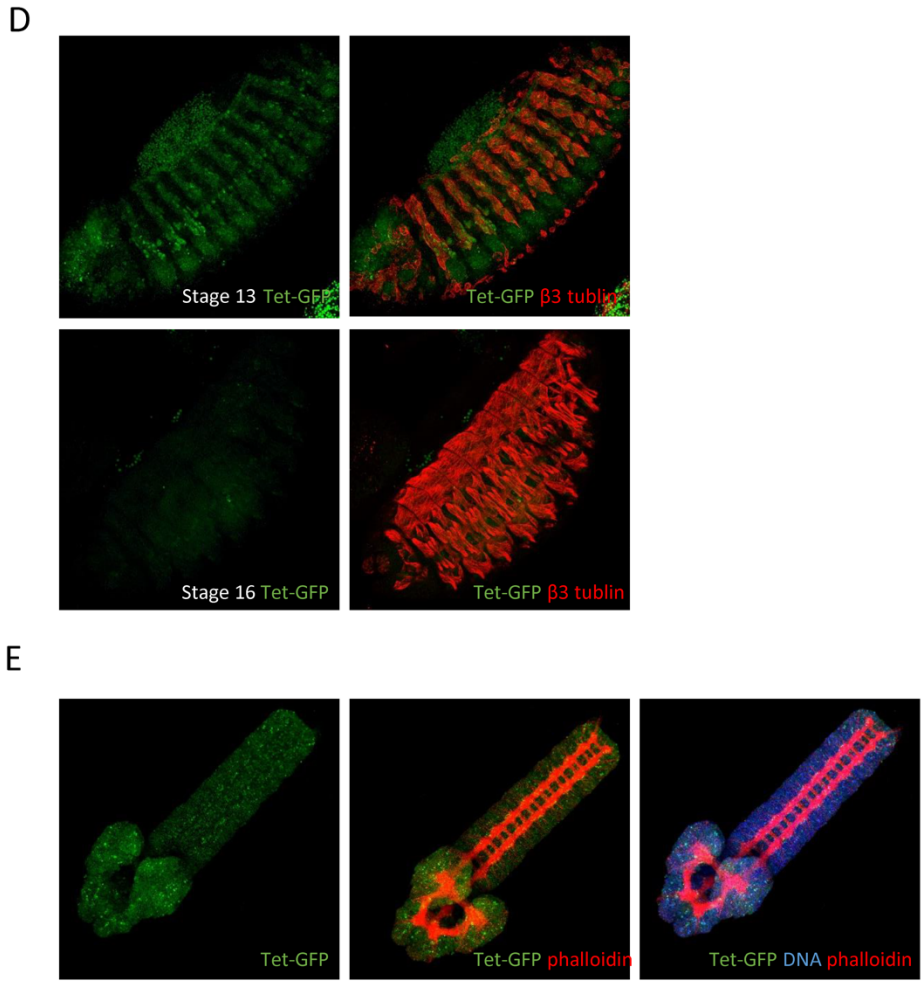
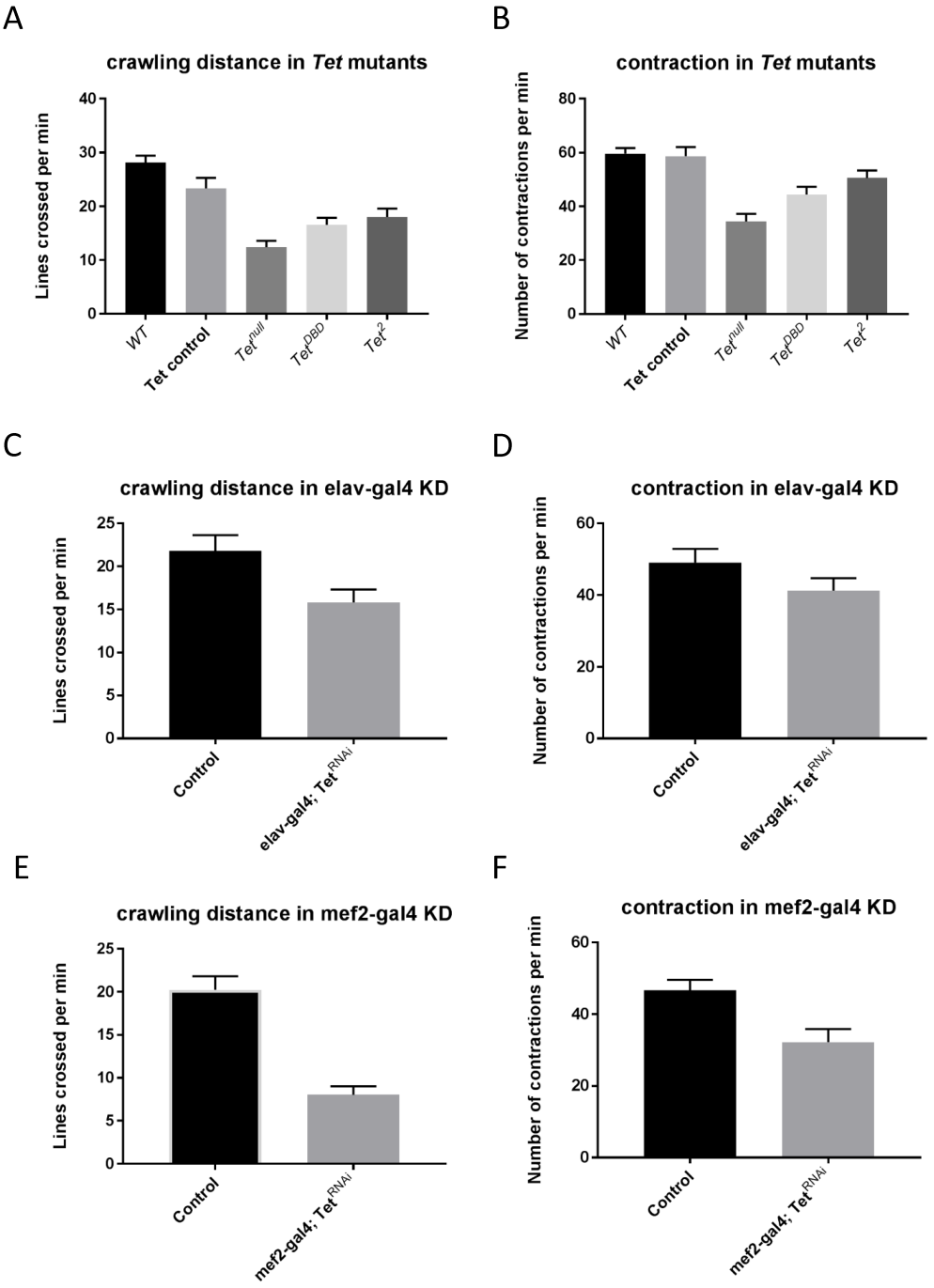


Figure 16 Tet expression.

(**A**) Native Tet-GFP is detected first at the germ band extension stage (side view) and becomes more pronounced in neuronal cells by stage 13 (side view). At stage 14 (ventral view) the protein is clearly nuclear (see enlargement below) The protein is also detected most strongly in proneural cells of the wing disc and in the developing photoreceptors behind the morphogenetic furrow. **(B)** Tet-GFP and the neuronal marker Elav overlap in the central nervous system. Stage 14 embryo, ventral view. **(C)** Tet expression shows significant overlap with the mesodermal marker Mef2, but it its expression in mesodermal cells is lower than in neuronal cells. Stage 13 embryo; note the autofluorescence of the yolk mass on the dorsal side of the

embryo. **(D)** Tet-GFP is expressed in differentiating mesodermal cells, labeled by beta 3 tubulin in stage 13 embryos; by stage 16 Tet-GFP is no longer detected. **(E)** Central nervous system dissected from 1st instar larva (24h). Tet-GFP is expressed in the central nervous system throughout embryogenesis. Phalloidin stain represents F-actin. NB: all figures show endogenous Tet-GFP expression, except for B which is stained with anti-GFP antibody.


A
crawling distance in Tet mutants
Lines crossed per min
40
30
20
10
0
WT
Tet control
Tet^null
Tet^DBD
Tet^2
B
contraction in Tet mutants
Number of contractions per min
80
60
40
20
0
WT
Tet control
Tet^null
Tet^DBD
Tet^2
C
crawling distance in elav-gal4 KD
Lines crossed per min
25
20
15
10
5
0
Control
elav-gal4; Tet^RNAi
D
contraction in elav-gal4 KD
Number of contractions per min
60
40
20
0
Control
elav-gal4; Tet^RNAi
E
crawling distance in mef2-gal4 KD
Lines crossed per min
25
20
15
10
5
0
Control
mef2-gal4; Tet^RNAi
F
contraction in mef2-gal4 KD
Number of contractions per min
60
40
20
0
Control
mef2-gal4; Tet^RNAi

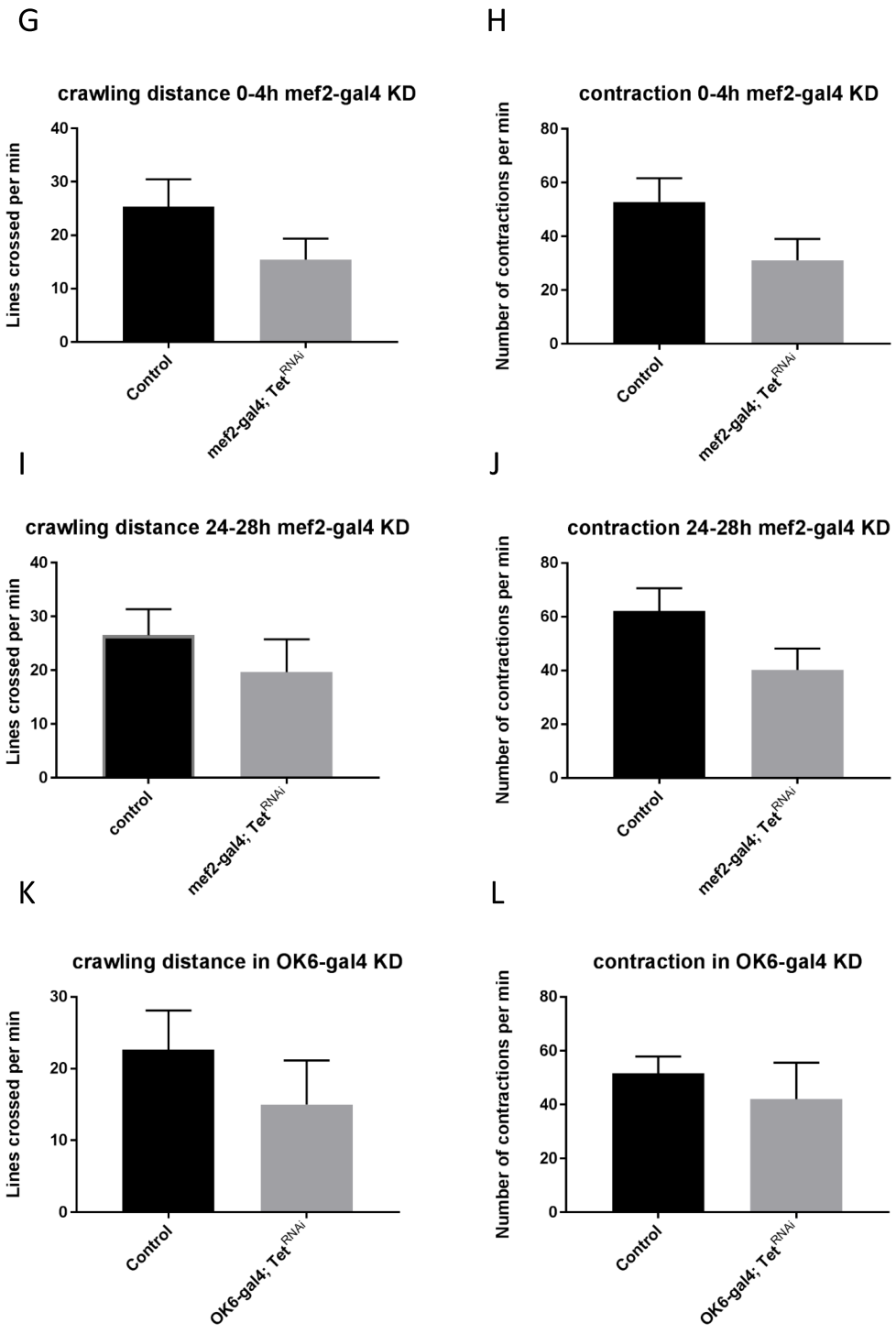


Figure 17 Tet requirement in larval locomotion.

(A, B) Locomotion and contractions are significantly affected in *Tet* mutants. Note that *Tet*^{null} shows a stronger effect than the alleles that retain some Tet function. WT, w^{118} larvae, Tet control, parental stock containing piggyback transposon PBac(SANSON *et al.*)f05022 **(C, D)** Tet KD by the *mef2-gal4* driver results in strong reduction of locomotion and contractions. **(E, F)** Tet KD by the *elav-gal4* driver results in moderate but significant reduction of locomotion and contractions. **(G, H)** Conditional Tet KD by the mef2-gal4 driver during embryogenesis affects locomotion and contractions significantly, but the difference in locomotion and contractions is alleviated when the KD is induced in first instar larvae **(I, J)**. **(K, L)** Tet KD by the *OK6-gal4* motorneuron driver did not affect the larval locomotion and contractions.

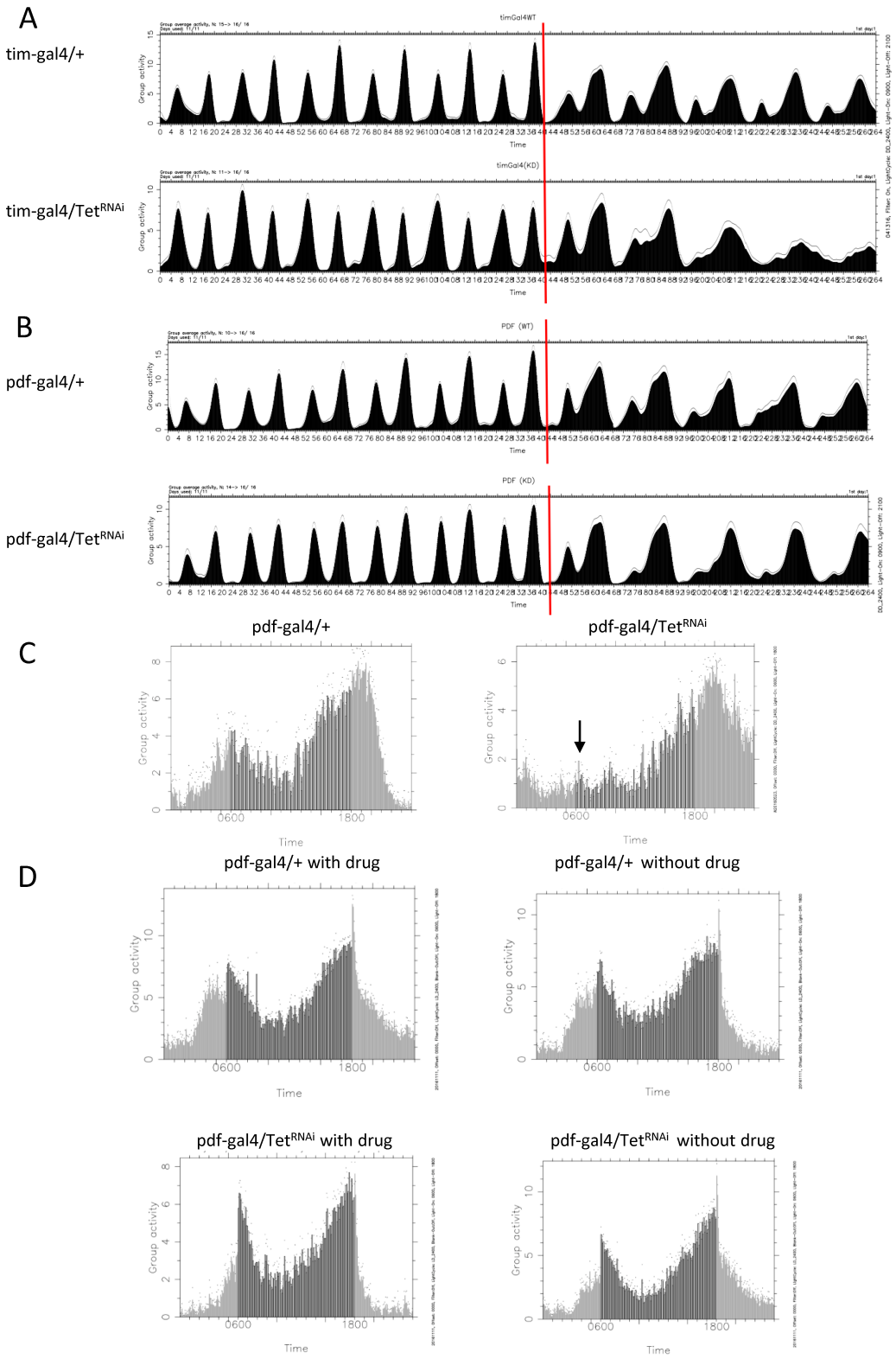


Figure 18 Tet and the circadian rhythm.

(A, B) Actograms: Depletion of Tet with the *tim-gal4* driver alters the bimodal pattern of the activity peaks after two days of DD. The morning activity peak is lost on day 2 of DD when Tet is depleted by *pdf-gal4*. **(C)** Eduction graph of average activities in DD shows the reduction of the morning peak (black arrow) if Tet is knocked-down throughout the expression period of *pdf-gal4*. **(D)** No effect on the morning peak is observed when Tet is depleted only in the adult stage (pdf-GS-gal4/Tet^{RNAi} with drug) **(A, B)** Light-Dark training, left of the red line, Dark-Dark behavior, on the right.

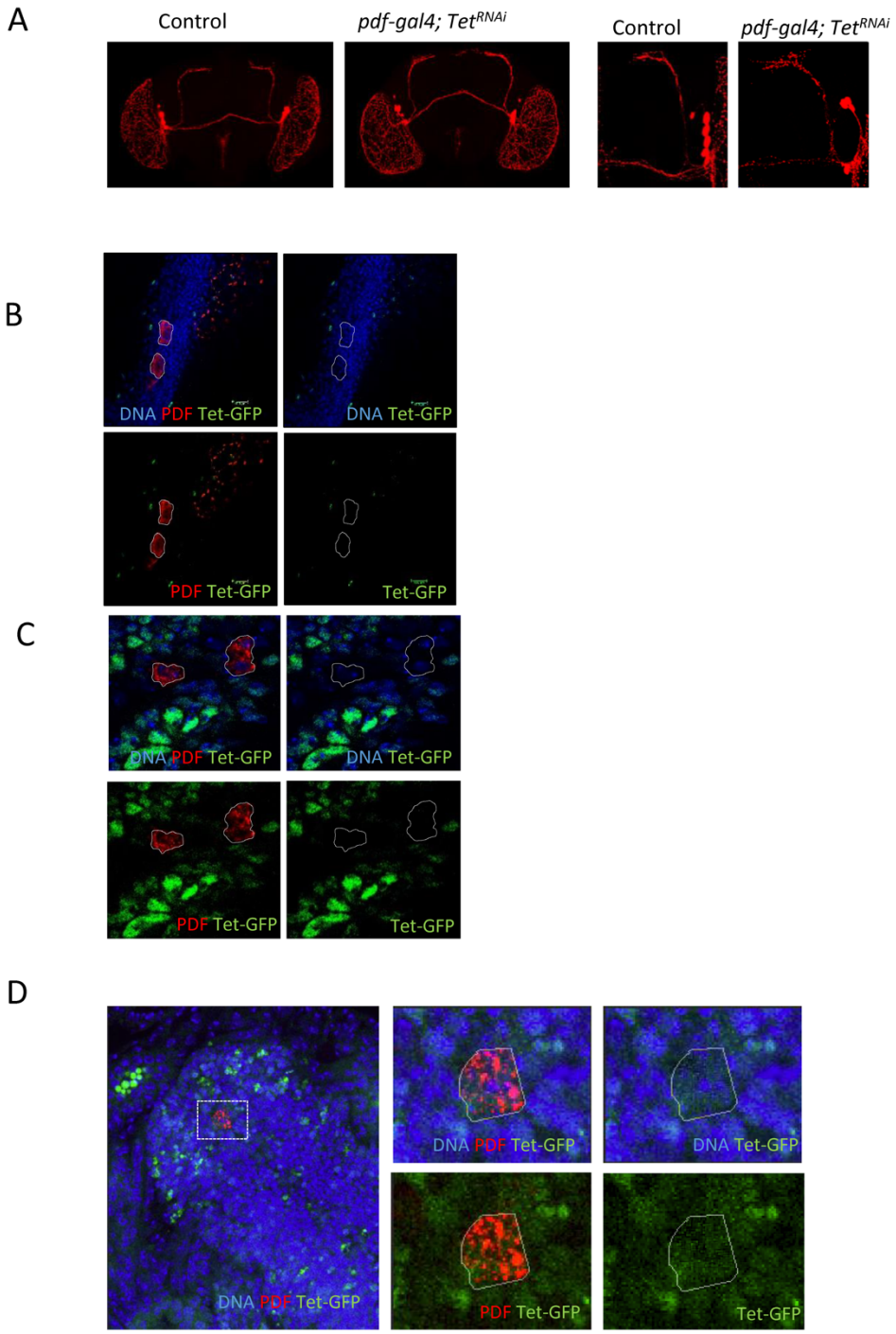


Figure 19 Tet expression in PDF neurons.

(A) *Tet KD* by *pdf-gal4* does not affect organization of PDF neurons in adult brain. Brains are stained by anti-PDF antibody. **(B)** Tet-GFP is not expressed in PDF neurons in adult males or**, (C)**

in third instar larval brains. **(D)** Tet-GFP protein is observed in the cytoplasm of PDF positive neurons in the embryonic central nervous system (stage 16).

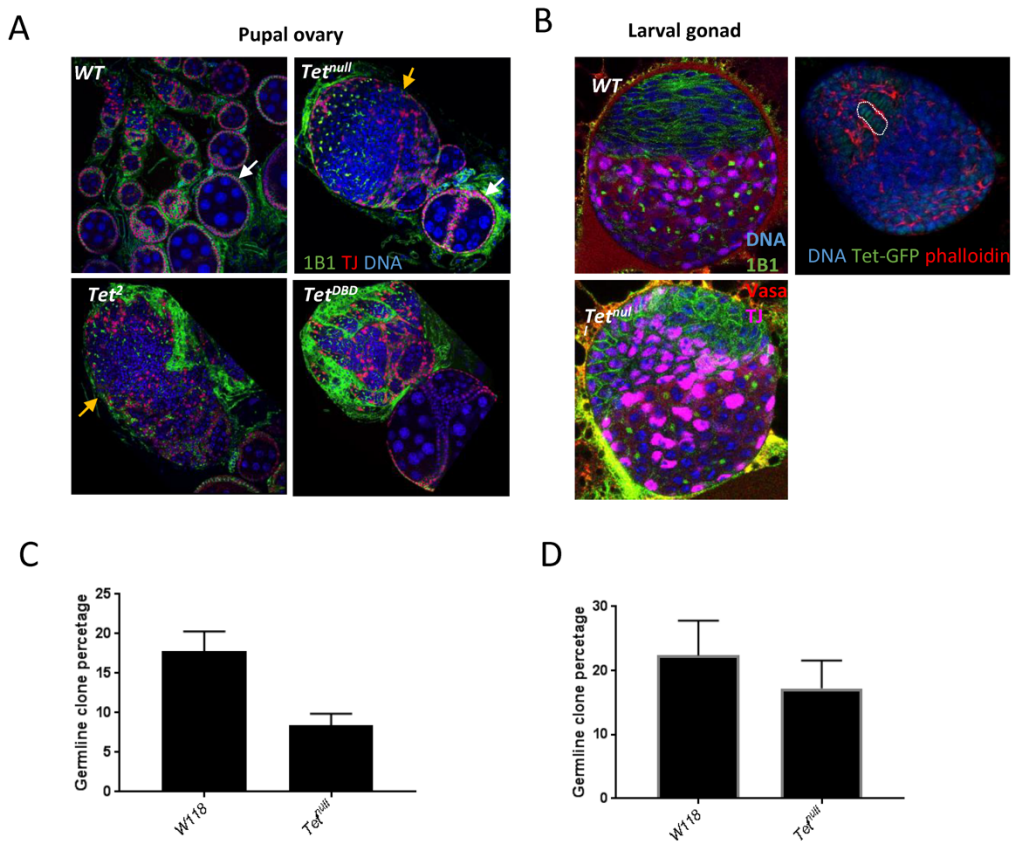


Figure 20 Tet controls ovarian development.

(A) The ovarioles are clearly separated and contain up to stage 8 egg chambers (white arrows). *Tet* mutant ovaries fail to separate their ovarioles, but do contain some disorganized egg chambers (orange arrows). The *Tet*^{null} ovaries does not show the clear separation of the somatic (TJ, red) and germline tissues. 1B1 in green outlines the somatic cells and fusome. **(B)** Left side, 3rd instar larval gonad: *Tet*^{null} gonad display the defect in cell organization. TJ (purple) labels the somatic cells and Vasa (red) labels the germline cells. Right side, Tet-GFP is observed in terminal filament cells in WT late 3rd instar larval ovaries. Red, filamentous actin, most highly expressed around the terminal filament cells. The germline clones induced in 0-24h embryogenesis showed

a strong reduction **(C),** but not the ones induced at 3rd instar larval stage **(D).** PGC, primordial stem cells; GSC, germ line stem cells.

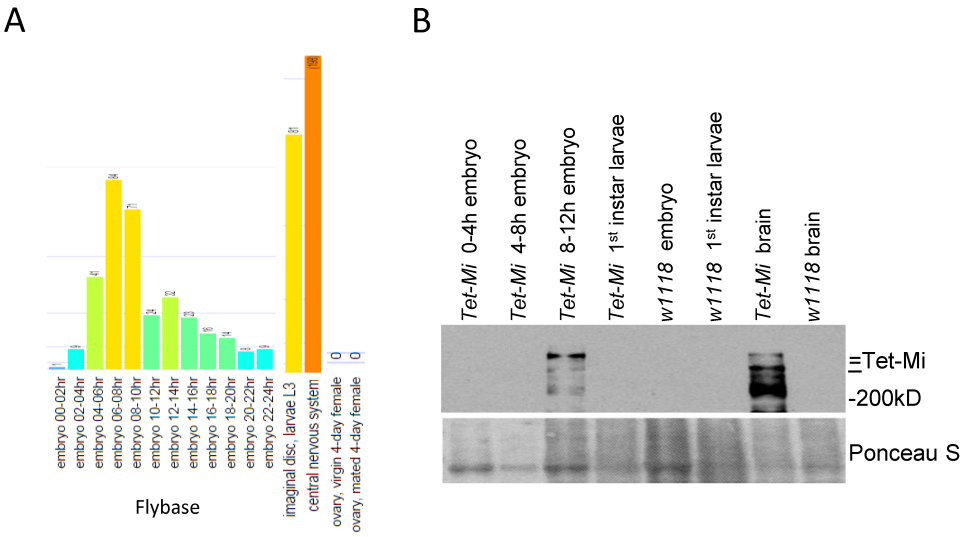


Figure 21 Tet mRNA and Protein Expression.

(A) High throughput RNA seq; Tet is most highly expressed in 2-12 h embryos and in 3rd instar discs and brains (from flybase). **(B)** Tet-GFP, similar to the Tet RNA, is detected in 8-12 h embryos and larval brains (Western blot reacted with anti-GFP antibody).

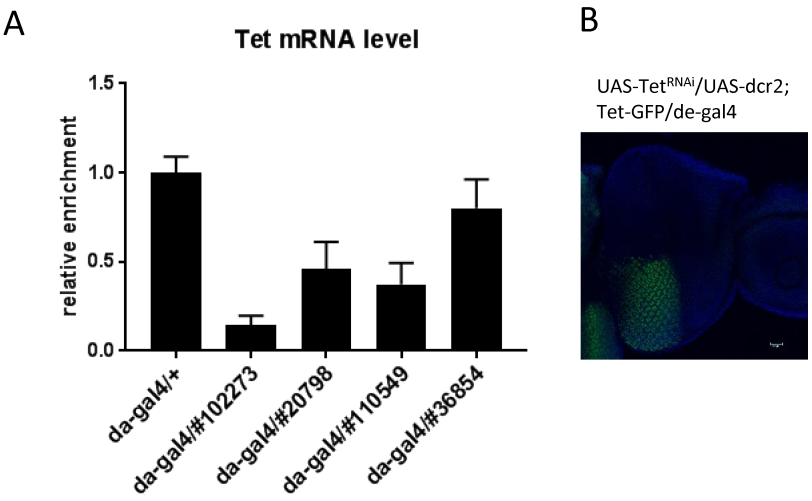


Figure 22 Tet KD.

(A) Tet RNA depletion in different Tet-RNAi lines using the *da-gal4* driver. 0-12 hour embryos were used for the experiments. **(B)** Tet-GFP is absent in half of the eye disc when Tet RNAi expression is controlled by the DE-gal4 driver, active on the dorsal side of the eye disc.

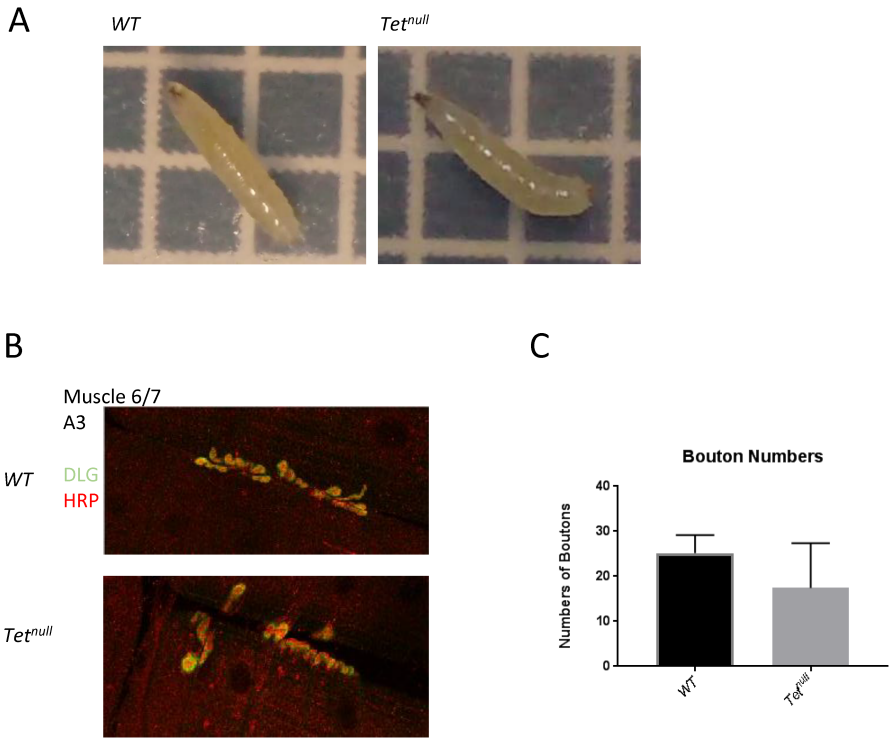


Figure 23 Mutant larval size and neuromuscular junctions.

(A) The WT and mutant 3rd instar larvae look identical in size**. (B, C)** The number and morphology of boutons is not changed in *Tet*^{null} larvae.

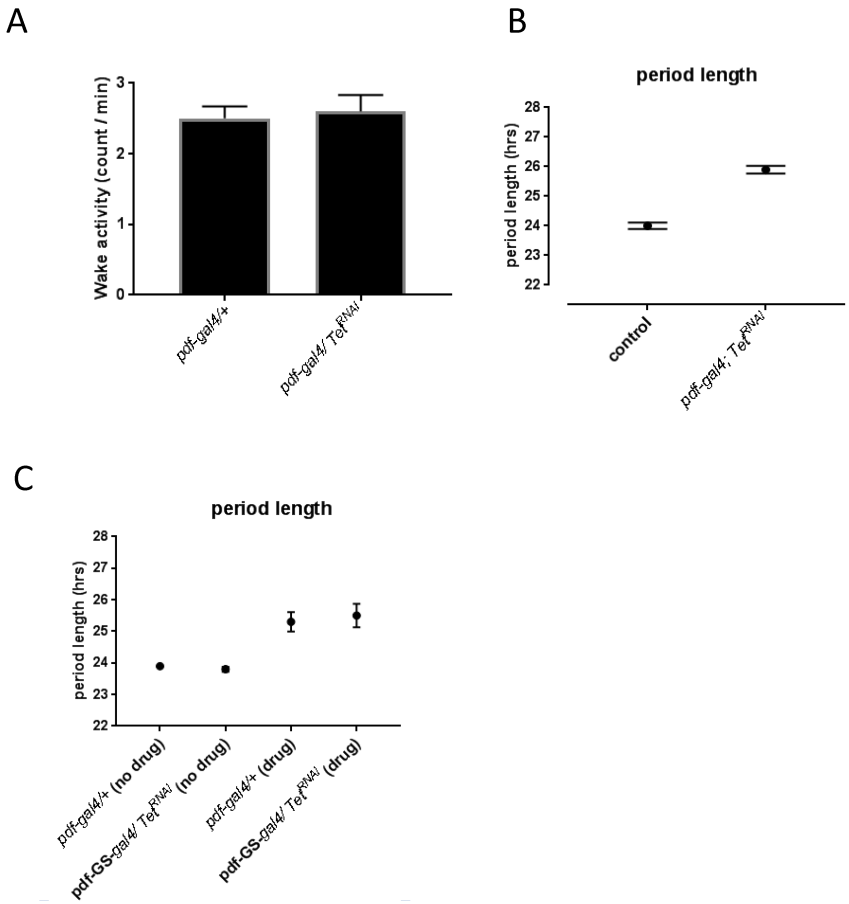


Figure 24 Period length in pdf-gal4 KD flies.

(A) Average activities are not changed in control and *pdf-gal4* Tet KD adult males. **(B)** but period length was prolonged in *pdf-gal4* Tet KD males. **(C)** Period length was not affected when Tet is only depleted in adult stage.

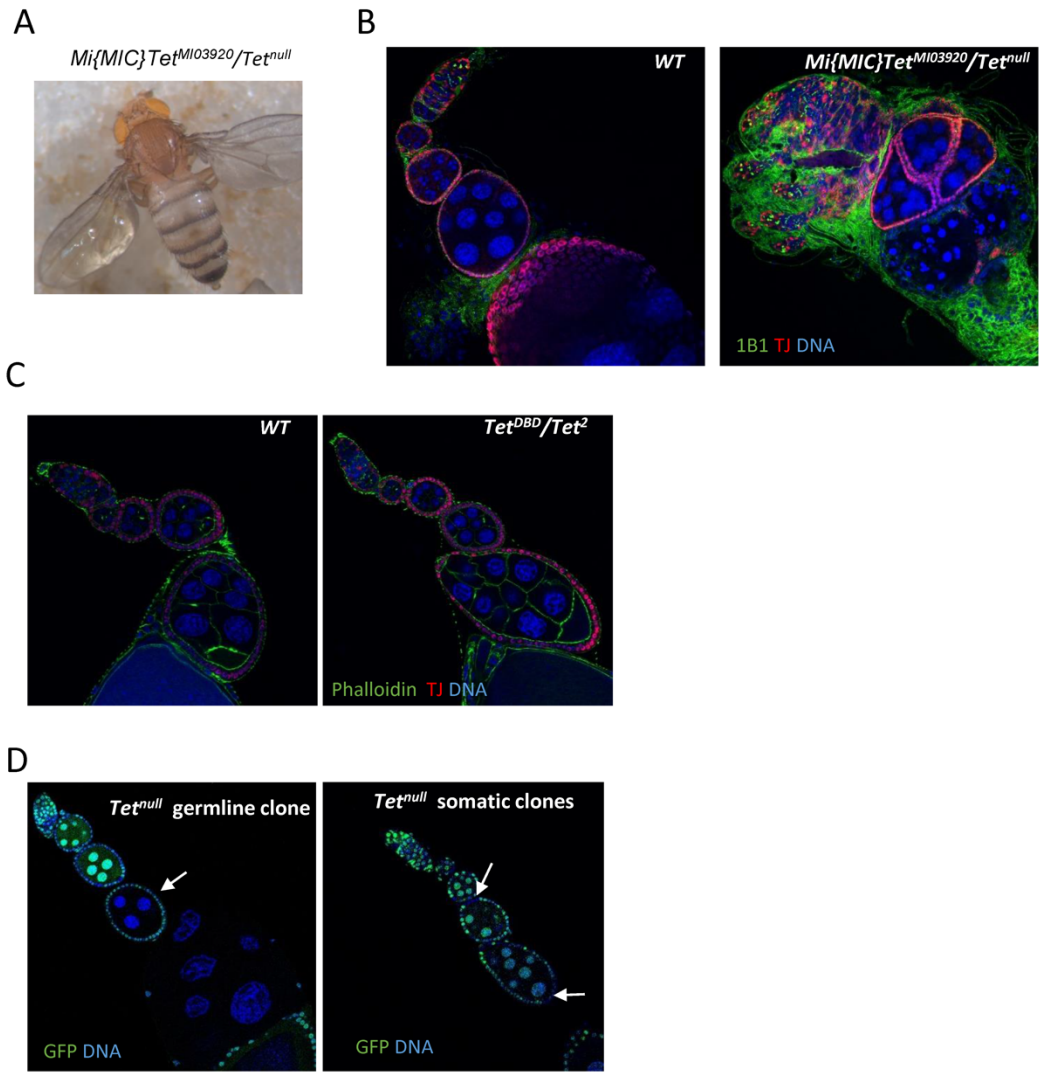


Figure 25 Tet mutant ovary.

(A) *Mi{MIC}Tet*^{MI03920}*/Tet*^{null} adults show a held-out phenotype and are uncoordinated***,*** before dying 2-3 days after eclosion. **(B)** The adult ovary phenotype of this mutant shows some separation of ovarioles, but overall the ovary does not look much different than other *Tet* alleles that do not survive well into adulthood (Fig. 6). **(C)** *Tet*^{DBD}*/Tet*^{2} ovary shows no significant difference from control. **(D)** *Tet*^{null} germline white arrow) and somatic clone (red arrow) show no phenotype.

V. Discussion

Tet functions in neurons and muscle precursors

In our previous work, we mapped the transcriptome-wide 5hmrC landscape in S2 cells and found that transcripts encoded by genes involved in neuronal differentiation contained the most 5hmrC peaks. These genes also represent a significant percentage (26%, $P<10^{43}$) of the Tet-regulated transcripts as determined by sequencing of RNAs isolated from normal and Tet depleted S2 cells. We further reported that Tet was an essential gene and that *Tet*^{null} larval brains showed a small, but significant defect in development (DELATTE *et al.* 2016). Our further experiments into the brain phenotype (induction of *Tet*^{null} clones and Tet KD with brain-specific drivers) showed that the observed reduction of the medulla of larval brains was most likely due to a developmental delay.

The results obtained in S2 cells are corroborated and extended by the *in vivo* data we report here. We found that Drosophila Tet protein is expressed most highly in the central nervous system, especially the brain, starting at germ band extension and continuing throughout development. This agrees well with observations that the 5hmrC mark is highest in larval brains (DELATTE *et al.* 2016). Although we did not observe lethality after depletion of Tet from neuronal cells, we found that Tet was required early in development in pdf neurons for circadian rhythm.

Tet protein is present in somatic muscle precursors in the embryo, where it appears to have an essential function. 3rd instar *Tet*^{null} larvae exhibit normal morphology but severe motility defects and newly eclosed hypomorphic adult flies have uncoordinated leg and wing movements. These defects could either derive from Tet loss of function in the postmitotic neurons, or from Tet loss in muscle precursor cells. Our results show that Tet expression in

neurons and muscle precursor cells is highest during embryonic development and is also required in both tissues during embryogenesis for normal larval locomotion. Further, this time window of Tet activity overlaps with the period when Tet is required in PDF expressing neurons to assure normal bimodal circadian rhythms. These results are consistent with Tet having an essential function in neuronal and muscle development during embryogenesis. Consistent with the notion that the Tet-dependent deposition of 5hmrC serves as an epigenetic mark, the absence of Tet function during embryogenesis results in phenotypes which become manifest at later developmental stages (e.g., larval or adult) whereas the absence of Tet function during late larval stages has little or no affect upon the processes investigated in this study.

In vertebrates Tet levels are especially high in ESCs and in neurons (KRIAUCIONIS AND HEINTZ 2009) (ITO *et al.* 2010) (SZWAGIERCZAK *et al.* 2010) and requirement of Tet proteins in neurogenesis has been documented in several species. In Xenopus Tet3 is strongly expressed in embryonic developing neuronal tissues and depletion of Tet3 using Morpholino antisense Oligos, resulted in abnormalities in the development of these tissues and in death (XU *et al.* 2012; SANTIAGO *et al.* 2014). Further, mouse Tet3 KO ES cells, while normal in self renewal, fail to differentiate into viable neural precursor cells (LI *et al.* 2015). Also, Tet1 KO mice develop normally with no visible morphological or groWTh defects. However, when 4 months old WT and Tet1 deficient mice were tested in the Morris water maze, a deficit in learning and short term memory was detected in the mutant animals (ZHANG *et al.* 2013). In the literature these phenotypes have been interpreted as resulting from a lack of 5hmrC in DNA, but it seems possible that 5hmrC may also be involved in these neuronal processes. Tet proteins have also been linked to muscle development and function in vertebrates, but their direct involvement has not been clearly documented so far (KINNEY *et al.* 2011; ZHONG *et al.* 2017).

Tet^{null} animals die in late pupal stages and the lethality can be separated from the larval locomotion phenotype in that Tet appears to be required in mesodermal cells throughout embryonic and larval development for viability, correlating with the expression of the *Tet* gene. It is therefore possible that Tet function may be essential in additional cells and tissues in which it is expressed, such as adult tissues derived from the Tet positive larval disc cells.

Requirement of Tet in the ovary:

We demonstrated that Tet is required for normal oogenesis in the embryonic stages. In our clonal analysis we observed a reduction in *Tet*^{null} egg chambers when clones were induced during embryogenesis, but induction in 3rd instar larvae did not reduce the number of clones in the germline or the soma. These results are consistent with *Tet* functioning in germline cells during embryogenesis and early larval stages. In addition, we found that complete loss of *Tet* in Drosophila results in a strong phenotype in larval and pupal ovaries: the ovaries are disorganized and mesodermal cells fail to migrate to form ovarioles. Together with the observed expression of Tet in somatically derived terminal filament cells, the phenotypes suggest that Tet is also required the somatic cells of the developing ovary.

Tet was shown previously to be both expressed and required in the germline in early oogenesis (ZHANG *et al.* 2015). In contrast we did not observe expression of Tet during these stages. Therefore we investigated if there was any maternal requirement for Tet function during oogenesis. To this end we induced *Tet*^{null} clones by the OvoD technique, which allows creation of homozygous mutant embryos from mutant eggs (Ref). In our experiments the homozygous *Tet*^{null} embryos derived from *Tet*^{null} stem cells did not show any enhanced phenotype compared to the embryos derived from *Tetnull/+* stem cells.

Our results on the requirement for Tet during oogenesis differ significantly from those previously published by Zhang et al. (ZHANG *et al.* 2015) We could not confirm their reported presence of the Tet protein during oogenesis and in 0-2h embryos. Further, when we induced clones in third instar larvae we did not observe any effect on early germ cell differentiation. The differences in expression pattern may be due to the fact that Zhang et al. used an anti-Tet antibody, that may not have recognized the Tet protein specifically, while we demonstrated the absence of Tet expression in adult ovaries and 0-2 hour embryos using the Tet-GFP reporter line both by fluorescent imaging and western blot (Figure 21B).

The Tet proteins

Drosophila has only one *Tet* gene, encoding several alternatively spliced isoforms from two distinct promoters. Drosophila Tet-L most closely resembles mammalian TET1 and 3, while Tet-S is similar to TET2. Both DNA binding and catalytic domains of Drosophila Tet are about 50% homologous to that of TET1/3, and the specific amino acids within the catalytic domain that are responsible for methylcytosine binding are identical in Drosophila and vertebrates (DUNWELL *et al.* 2013). Homozygous or hemizygous Tet^{DBD} and Tet^{2} animals die, like Tet^{null}, at the late pupal stage while Tet^{DBD}/Tet^{2} show strong complementation (Table 1). But the trans-heterozygotes adults live up to two weeks and have normal looking ovaries (Figure 25C), and females are sterile. This result indicates that both forms of Tet have essential functions.

Mammals have three Tet genes: Tet 1 and Tet3 encode proteins with both a DNA-binding and a catalytic domain, and Tet2 that lacks the DNA-binding domain. Xenopus however, has only two Tet genes, encoding the two basic forms of Tet, one form containing both domains and one from containing only the catalytic domain. Drosophila has only one Tet gene, encoding

the two basic forms of Tet. Vertebrate genes, while showing specific expression patterns are partially redundant. For instance, Tet1 and Tet2 KO mice survive, but the majority of double KO mice die within two days postnatally (DAWLATY *et al.* 2013), and Tet3 KO mice die one day postnatally (GU *et al.* 2011). This suggested that Tet3 has a distinct essential function which could be accounted for through its known role in RNA hydroxymethylation.

The catalytic domains of all three Tet enzymes as well as full-length Tet3 have been shown to induce the formation of 5-hmrC in RNA in mouse and human cells (FU *et al.* 2014a). Of the three Tet proteins, human TET3 is structurally most similar to Drosophila (DUNWELL *et al.* 2013) and TET overexpression experiments in a human cell line suggests that hydroxmethylcytosine formation in RNA may be controlled solely by TET3 (FU *et al.* 2014a). Further, *Tet3* expression is reported to be particularly high in the developing mouse brain (HAHN *et al.* 2013). Because both the Tet-L (similar to Tet3) and Tet-S (similar to Tet-2) are essential in flies where 5hmrC in DNA is lacking, Tet2 may be functioning with Tet3 in controlling 5hmrC also in vertebrates. But it is also possible that Tet proteins have additional, non-catalytic, essential functions in flies that are conserved in vertebrates.

In studying the expression and distribution of Tet-GFP in embryos and larvae we find that the protein is found primarily in the nucleus, but can also be detected exclusively in the cytoplasm, for instance in PDF expressing neurons in the embryo. As Tet-L contains a DNA-binding domain, we hypothesize that its major function is in the nucleus, where it may control the modification of nascent transcripts. It is possible that Tet-S represents the cytoplasmic protein. Independent of the localization of the Tet proteins, they control the modification of specific mRNAs and likely their transport and translation (DELATTE *et al.* 2016).

Together with the documented function of vertebrate proteins in RNA modification, our results invite further investigation of 5hmrC in vertebrate RNA, particularly in neuronal and muscle development and function. It is well possible that phenotypes described in mouse mutants, or symptoms associated with Tet in humans, could be due to abnormal processing of specific mRNAs, rather than resulting in response to disruption in hydroxymethylation of Cytosine in DNA.

VI. Materials and methods

Fly lines

The deletion alleles *Tet*^{DBD} and *Tet*^{2} were generated by FRT site-specific recombination between two PBac insertions, obtained from Exelixis at Harvard Medical School (Fig 1). *Tet*^{DBD} was recombined using *PBac(SANSON et al.)f0306* and *P{XP}d00815*. *Tet*^{2} was recombined using *PBac(SANSON et al.)f05022* and *P{XP}d00815*. The mutant chromosomes are kept over balancer chromosomes *TM3* or *TM6*. The Tet RNAi lines (#v110459, #v102273, #v20798, #v36178) were obtained from the Vienna *Drosophila* RNAi Center (VDRC). We found that #v102273 showed the most efficient KD and therefore this line, called TetRNAi in this work, was used for all experiments. The *mef2-gal4*, *da-gal4, tub-gal4, act5-gal4, elva-gal4, pdf-gal4, tim-gal4*, *UAS-dcr2*, and deficiency stocks were obtained from the Bloomington Stock Center. For RNAi experiments, flies were kept at 25°C. For temperature shift experiments, we used 25°C as permissive temperature and 29°C as restrictive temperature.

Clonal analysis

The FLP/FRP site-specific recombination system was used to generate mutant clones under the control of a heat-shock promoter (XU AND RUBIN 1993). The fly stocks

used to induce ovoD clones and have been previously described (PERRIMON 1998; ST JOHNSTON 2002). To induce clones, 0-24h embryos and third instar larvae of the appropriate genotype were exposed to 38°C for 90 minutes. Ovaries were dissected, fixed stained at 5-10 days after eclosion. To count and characterize clones, more than 150 ovarioles of each genotype were analyzed on a Zeiss Axioplan-2 microscope.

Antibodies and microscopy

Embryos were dechorionated in 50% bleach, fixed in formaldehyde-heptane and devitellinized with methanol. Embryos were permeabilized and blocked in PBST and 5% goat serum for 1 hours. Embryos were incubated in primary antibodies overnight at 4°C overnight, and in secondary antibodies for 2hrs at room temperature. The following primary antibodies were used: rat-anti-Elav (1:250, DSHB), mouse-anti-Prospero (1:20, DSHB), rabbit anti-Mef2 obtained from H. Nguyen, University of Erlangen-Nuremberg, Nurenberg, Germany (1:1000) (BOUR *et al.* 1995), rabbit anti-ß3 tubulin from R. Renkawitz-Pohl, Philipps-Universität Marburg, Marburg, Germany (1:1000) (LEISS *et al.* 1988), and rabbit anti-GFP (1:1000, Invitrogen). Alexa Fluor-488, 568, 647 secondary antibodies were from Molecular Probes and used at 1:500.

For staining of larval NMJ, 3rd instar wandering larvae were dissected as described in Brent et al 2009, (BRENT *et al.* 2009a; BRENT *et al.* 2009b). 3rd instar larvae were dissected in calcium free HL-3 saline and fixed in 4% paraformaldehyde in PBT (PBS + 0.05% Triton X-100). Larvae were then washed briefly in 0.05% PBT for 30 min and incubated overnight at 4 with the following primary antibodies: mouse anti-DLG, (1:100, DSHB); TRITC-conjugated anti-HRP (1:200, Jackson ImmunoResearch). Synaptic boutons and NMJ expansion were quantified with Leica software.

Confocal images were captured using a Leica TSC SP5/8 laser scanning confocal microscope (objectives 40× and 63× oil), analyzed with Leica Microsystems software and further processed using Adobe Photoshop.

Quantitative RT-PCR

We used RNA from third instar larval brain for quantifying Tet mRNA levels. Total RNA was isolated from 20 brains using the RNeasy Plus Mini kit (Qiagen). Quantitative RT-PCR was performed as described in the manufacturer's instructions using the SYBR Green Selected kit (Qiagen) and the relative standard curve method. Primers used for RT-PCT are listed in S6 table. Transcript levels were normalized to those of *Rp49. RNA purified from* WT larval brains was used as control, and all data were normalized to the transcript levels in WT (baseline=1). At least three biological and three technical replicates were performed for each genotype. Statistical significance (*P*-value) was determined using two-tailed Student's *t*-test.

Fly assays

Lethal phase analysis: Flies were cultured at 25°C and allowed to lay eggs for 12 hours. For each genotype, 100 embryos were collected and transferred to a sugar agarose plate and each plate was scored for the number of hatched larvae and pupae. Experiments were repeated for 3 times.

Larval locomotion assay: Larval locomotion was assayed as described in Louis et al. 2008 (LOUIS *et al.* 2008) but without odor source. Briefly, single mid-3rd instar larvae were placed on 3% agarose plates over a grid and animal locomotion was recorded by a camera for 1 min from

the larva's first movement. 30 to 100 animals were tested for each genotype. Student's t-test was applied to evaluate the statistical significance.

Circadian Rhythm assay: Adult male flies (2–5 day old) were used to test locomotor activity rhythms (CHIU *et al.* 2010). Flies were entrained under LD (the standard condition of 12 h of light and 12 h dark) for 6 days and released into constant darkness (DD) for at least 5 days at 25°C. FaasX software was used to analyze behavioral data to produce the actogram and eduction graph. Activity counts were collected in 5-min bins, and the data for individual flies were pooled to generate group averages of D2 in DD.

CHAPTER 4: GENERAL DISCUSSION

RNA modifications in neurons

Plasticity is a remarkable feature of the brain and can occur in response to outside stimuli, or to internal injury. It is widely accepted that this plasticity is controlled by gene expression and protein synthesis, both required for dynamic regulation of neuron functions (ALBERINI AND KANDEL 2014). Postmitotic neurons transduce signals in less than miliseconds using both chemical and electrical processes. These reactions are regulated by both small molecules and macromolecules such as proteins (SWEATT 2016). The activity-induced neuronal transcriptome is dynamic, regulated on many layers such as transcription rate, RNA processing, RNA degradation, RNA localization, as well as protein translation rate (RABANI *et al.* 2011). In fact, often there is limited correspondence between levels of mRNA and its corresponding protein (VOGEL AND MARCOTTE 2012). For example, it has been shown that only 40% of the variance in protein levels can be directly attributed to mRNA levels (SCHWANHAUSSER *et al.* 2011). The structures of neurons are special in that they have a cell body that branches out into dendrites and one long axon that communicates with other cells. After transcription in the nucleus, the processing of mRNAs is strictly controlled and plays important roles in levels and local range of proteins. For instance, some mRNAs are translated at the synaptic terminals, where the axon connects to other cells, and this local translation is essential.

RNA modification can be employed bidirectionally to regulate the life of RNA, serving as an additional layer of post transcriptional regulation. For example, RNA methylation in the form of N^6-methyladenosine (m6A) and N^1-methyladenosine (m1A) is critical for controlling RNA levels and even for the rate and fidelity of protein translation (FU *et al.* 2014b). Another highly abundant RNA modification, pseudouridine (Ψ), has recently been shown to be dynamically

responsive to stimuli (CARLILE *et al.* 2014). In addition, depending on the locus, chemical modifications of RNA can dictate patterns of alternative splicing and degradation and influence secondary structure (LIU *et al.* 2015a; SPITALE *et al.* 2015). Changes in RNA editing and RNA structure can control protein binding affinity and microRNA targeting (LIU *et al.* 2015b; BEHM AND OHMAN 2016). These mechanisms serve as an additional layer of regulation for fine-tuning the state of mRNAs, imparting functional diversity without changing the levels of transcription.

m6A in Drosophila has been shown to be of great importance for adult locomotion and behavioral coordination. The function of m6A was shown to be required specifically in the nervous system (LENCE *et al.* 2016). In my study of 5hmrC, loss of function of Tet in Drosophila, the enzyme converting 5mC to 5hmrC, resulted in severe locomotion defects in 3rd instar larvae and the enzyme proved to be essential in both neurons and muscles. The specific mRNA targets in these two organs remains to be determined. Advances in our understanding of the epitranscriptome and emerging technologies that can be used to unravel the complex nature of post-transcriptional gene regulation in the brain and muscle is still in its infancy.

RNA 5hmrC function and Tet in vertebrates

Most of what we know about TET proteins comes from studies on DNA hydroxymethylation in mammalian systems (TAHILIANI *et al.* 2009; DAWLATY *et al.* 2011). However, when it comes to the functional study of RNA hydroxymethylation, the work done in mammals provided sometimes confusing and ambiguous data. Hence, I chose *Drosophila melanogaster* to perform my studies, because it lacks DNA methylation (RADDATZ *et al.* 2013; TAKAYAMA *et al.* 2014) as well as DNA hydroxymethylation, thus making it a straightforward model system to analyze the function of hydroxymethylated RNA. Drosophila does not contain homologs of the mammalian DNA methyltransferase genes Dnmt1, Dnmt3a, or Dnmt3b but

contains a gene encoding a homolog of Dnmt2, which is a known RNA methyltransferase that methylates specifically tRNAs (LYKO *et al.* 2000). Drosophila also harbors a homolog of mammalian TET. My work provides an answer for why Tet is present and essential in Drosophilla. I show the first functional and biological role for Tet. In experiments described in Chapter 2 we showed that Tet is responsible for the formation of hydroxymethylcytosine in RNA and promotes mRNA translation. In Chapter 3 we show in addition that it is essential in neurons and muscles.

A central and common theme emerging from the results is the important role of 5hmrC and Tet in the neuron. Transcriptome-wide analyses of 5hmrC in RNA from S2 cells identified hmrC peaks mapping a total 1597 transcripts, 120 of which are important for neuronal differentiation. The 5hmrC mark of about 75% of these transcripts appears to be regulated by Tet, as assessed by anti-5hmrC immunoprecipitation followed by RNA-Seq in Tet knockdown S2 cells. Further, we found that dTet expression correlates with hmrC levels during embryogenesis and that Tet is highest in the larval central nervous system. We also observed that KD of Tet impairs larval locomotion and that the locomotion of partial loss-of function *Tet* flies is affected.

An interesting question for the future will be to assess whether this neuron and muscle-specific function of Tet in Drosophila can be extended to other organisms. Some insights so far support this proposition; morpholino-based Tet KD in Xenopus showed that Tet3 plays an essential role in neuronal development (XU *et al.* 2012). In mammals, Tet3 expression has been reported to be particularly high in the developing mouse brain (HAHN *et al.* 2013). Also, Tet2 is involved in Myogenic differentiation of skeletal myoblast (ZHONG *et al.* 2017).

The similarity of requirement of Tet in a variety of organisms is also supported by the conservation of the proteins. Drosophila Tet is the closest to human TET3 (DUNWELL *et al.* 2013)

and it is the overexpression of TET3 in human cell lines that has the strongest effect on the formation of hydroxmethylcytosine (FU *et al.* 2014a). Thus, it would be interesting to further evaluate more closely the involvement of TET3 in brain development and function of vertebrates, as e.g. in brain-specific conditional Tet3 KO mice. Such an approach would allow evaluating the contribution of TET3 to the hydroxymethylation of cytosines in mRNA as well as studying the effects of loss of TET3 on brain function and development. While there has not been any publication on the functional importance of 5hmrC in vertebrates our work on flies, the conservation of the enzymes and the high expression of Tet proteins in neurons make it clear that understanding of the pathophysiology of TET must now encompass the novel role we have identified of TET and 5hmrC in RNA.

5hmrC decorates mRNA

In Drosophila, co-staining of polytene chromosomes revealed broad co-localization of Ime4 (m^6A methyltransferase) and RNA Pol II indicating that Ime4 may be localized at active genes. In addition, the absence of m^6A affects alternative splicing. Furthermore, localization of Ime4 to sites of transcription is RNA-dependent (HAUSSMANN *et al.* 2016). These observations suggested that the m^6A marks are deposited co-transcriptionally on specific nascent RNAs.

Our results (Chapter 2) indicate that 5hmrC is found strongly enriched in polyadenylated RNA, while it seems absent in other RNA species, tRNA (small RNAs) and rRNA. The substrates for RNA hydroxymethylcytosine within tRNA and rRNA exist, as they both harbour cytosine methylation (GOLL AND BESTOR 2005; HUBER *et al.* 2015). For exemple, tRNA methylation has been demonstrated at C38 of tRNAAsp, tRNAVal and tRNAGly (GOLL AND BESTOR 2005). Hence, it remains a formal possibility that besides mRNA, tRNA and rRNA might also be

hydroxymethylated, at least in some contexts. However, these hypothetical modifications are unlikely due to Tet activity.

By transcriptome-wide mapping, we found that 5hmrC is found in more than 1500 mRNAs and the distribution is non-uniform. Almost 80% of 5hmrC marks are found within exons. We also found about 10% of 5hmrC was localized in or around introns. Tet protein is enriched in nuclei and is found at low levels in the cytoplasm, suggesting that this modification can be added, at least in some cases on pre-mRNA, either before or at the same time as RNA splicing in nuclei.

The long form of the Drosophila Tet protein contains a CXXC DNA binding domain and is essential. Based on the distribution of the protein and the presence of the CXXC domain, we propose that Tet can hydroxylate nascent transcripts and that it is part of ae complex with possibly a RNA methyltransferase and maybe additional protein. The deposition of m6A modification on RNA is controlled in a similar way.

In addition, we found 5hmrC to be enriched within a UCCUC motif, which resembles a known exonic splice enhancer (ESE) sequence (SCHAAL AND MANIATIS 1999). Hence, a role for 5hmrC in alternative splicing regulation warrants further investigation. It will be also important to determine how the 5hmrC mark specificity is achieved and what are mechanisms provide the sequence specificity to Tet.

Identification of writer, eraser, and reader proteins of RNA modifications

To elucidate the functional importance of RNA modifications, it is important to identify the enzymes responsible for the deposition (writer), removal (eraser), and recognition of the

modification (reader). Genetic analysis allows to study the functional importance of the modification. For example, several studies of m^6A modification discovered its importance to regulation on mRNA stability, translation efficiency, and exon inclusion. These studies were done by manipulating the methyltransferases METTL3/METTL14/WTAP (MEYER AND JAFFREY 2014; YUE *et al.* 2015; FRYE *et al.* 2016; GILBERT *et al.* 2016) . The roles of m^6A were also validated by manipulating its readers, mainly the YT521-B homology domain-containing proteins WTAP (DOMINISSINI *et al.* 2012; MEYER AND JAFFREY 2014; YUE *et al.* 2015; FRYE *et al.* 2016; GILBERT *et al.* 2016). Thus, for other newly-mapped mRNA modifications like 5hmrC, identification of the writer, eraser, and reader proteins is of fundamental importance and helpful in exploring their functions, as well as revealing the underlying mechanisms.

5hmrC also represents an additional layer of gene expression control. Its study is complicated by the-fact that it can modify both DNA and RNA. Further, 5hmrC can function either as the intermediate of RNA demethylation or possibly as its own epitranscritomic mark, attracting reader proteins or affecting the secondary structure of RNA. Without identifying the target mRNAs and the writers and readers that function with Tet, those questions remain unanswered.

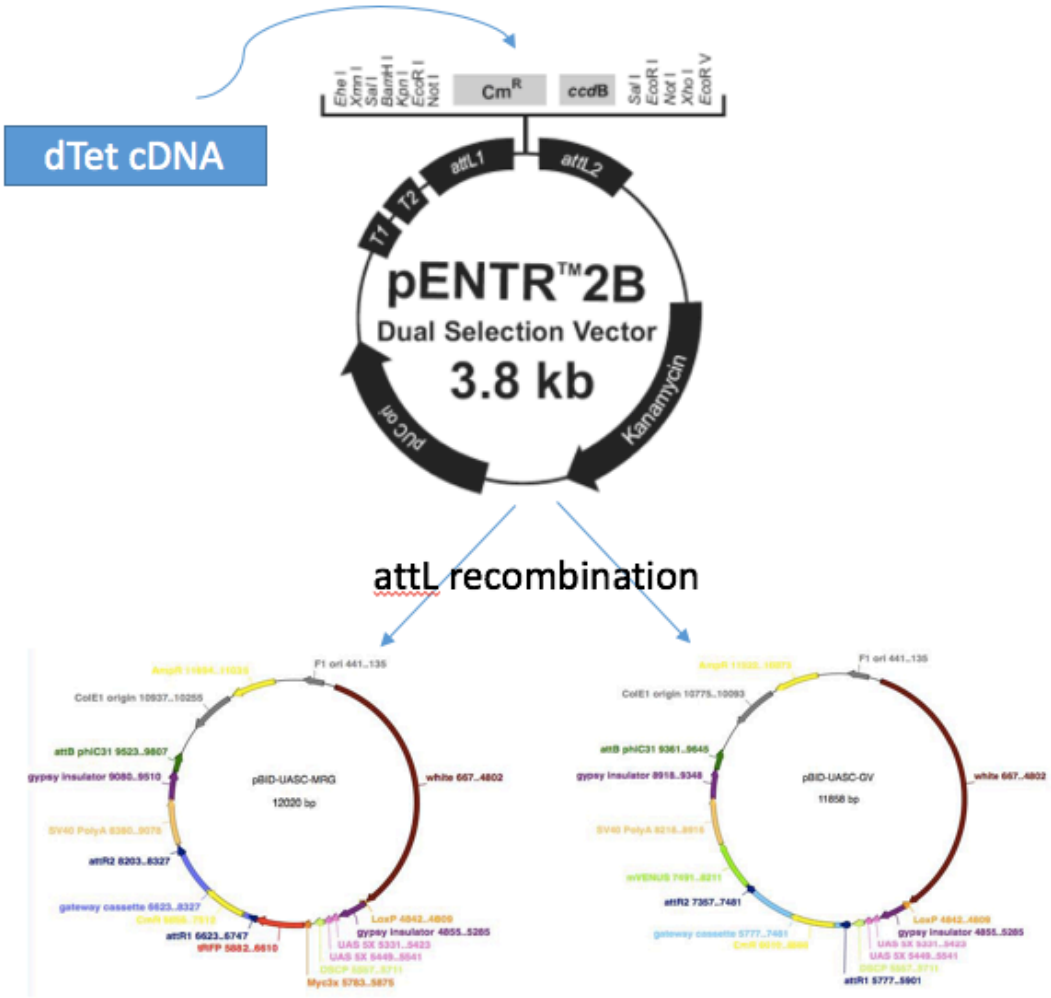
The initial discovery of m^6A readers relied on the RNA pull-down assay using synthesized oligonucleotides containing m^6A (DOMINISSINI *et al.* 2012). This approach could be repeated to identify readers for other modifications like 5mC or 5hmrC. Knock-out of Tet in Drosophila in Tet^{null} animals results in ~50% of reduction in the 5hmrC signal. This indicates that an additional hydroxylase of 5mC is likely to exist. To isolate additional writer proteins, we could purify fractions from cell extracts and assay their enzymatic activity through in vitro incubation of a synthetic target oligonuctleotide. METTL3 (also known as MT-A70), one component of m^6A writer was identified in this way (BOKAR *et al.* 1997).

In addition, modification enzymes may also be identified through bioinformatic predictions based on the conservation of functional domains of proteins, followed by the genetic validation of gain/loss of function analysis and study of the phenotypes (LIU *et al.* 2014). Tet was identified initially using this method (TAHILIANI *et al.* 2009). Because the hypothetical writer of 5mC could target mRNA, tRNA and even rRNA, the specific importance of modification of each RNA species may be difficult to determine. Under these circumstances, assays like photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) or individual-nucleotide resolution CLIP (iCLIP) could be very useful in identifying the targets of a writer transcriptome-wide.

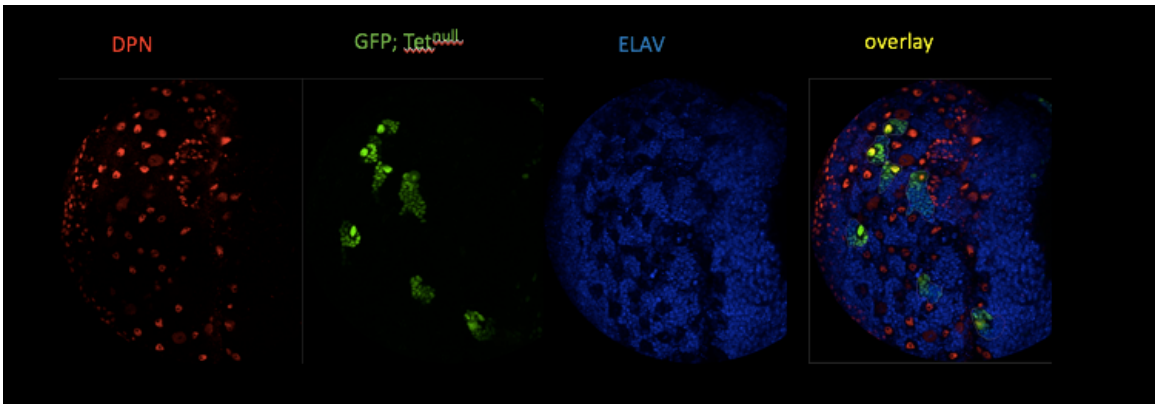
Summary

My thesis work provides the first mapping of the hydroxymethylated transcriptome, reveals the previously unrecognized function for RNA hydroxymethylation and dTet in mRNA metabolism, and highlights their central role in Drosophila neurons and muscles. I have discussed a several new methodologies that could be exploited to further understand the qualitative nature of RNA modifications and their metabolism and their effect on RNA processing and ultimately protein levels and localization. Armed with new sequencing technologies and cell-type-specific profiling, we will discover new events beyond the linear relationships between gene transcription and protein synthesis.

APPENDIX 1: SUPPLEMENTARY FIGURES



Tet cDNA was cloned into pENTR2B vector named as pENTR2B-dTet-full-length. pENTR2B-dTet-full-length transferred Tet full length cDNA to pBID-UASC-MRG and pBID-UASC-GV using attL-R recombination. UASC-MRG-Tet has N terminal tags of Myc and RFP. UASC-GV-Tet is Tet full length without any tags.



we assessed the proliferation of larval neuroblast clones using the MARCM technique. Using this strategy, the neural progeny of single neuroblasts of Tet homozygous can be labeled with GFP. We observed that *Tetnull* mutant neuroblast clones appropriately maintained a single neuroblast and could undergo multiple divisions to generate many neurons. This level of analysis does not address potential quantitative defects in neuroblast clones, nor does it rule out that a subpopulation of cells may have developed abnormally. However, Tet does not appear to be required for bulk aspects of neuron differentiation

REFERENCES

Aas, P. A., M. Otterlei, P. O. Falnes, C. B. Vagbo, F. Skorpen *et al.*, 2003 Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. Nature 421**:** 859-863.

Alberini, C. M., and E. R. Kandel, 2014 The regulation of transcription in memory consolidation. Cold Spring Harb Perspect Biol 7**:** a021741.

Batista, P. J., B. Molinie, J. Wang, K. Qu, J. Zhang *et al.*, 2014 m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell 15**:** 707-719.

Behm, M., and M. Ohman, 2016 RNA Editing: A Contributor to Neuronal Dynamics in the Mammalian Brain. Trends Genet 32**:** 165-175.

Ben-Haim, M. S., S. Moshitch-Moshkovitz and G. Rechavi, 2015 FTO: linking m6A demethylation to adipogenesis. Cell Res 25**:** 3-4.

Bird, A. P., and A. P. Wolffe, 1999 Methylation-induced repression--belts, braces, and chromatin. Cell 99**:** 451-454.

Bokar, J. A., M. E. Shambaugh, D. Polayes, A. G. Matera and F. M. Rottman, 1997 Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. RNA 3**:** 1233-1247.

Bour, B. A., M. A. O'Brien, W. L. Lockwood, E. S. Goldstein, R. Bodmer *et al.*, 1995 Drosophila MEF2, a transcription factor that is essential for myogenesis. Genes Dev 9**:** 730-741.

Brent, J., K. Werner and B. D. McCabe, 2009a Drosophila larval NMJ immunohistochemistry. J Vis Exp.

Brent, J. R., K. M. Werner and B. D. McCabe, 2009b Drosophila larval NMJ dissection. J Vis Exp.

Capuano, F., M. Mulleder, R. Kok, H. J. Blom and M. Ralser, 2014 Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. Anal Chem 86**:** 3697-3702.

Carlile, T. M., M. F. Rojas-Duran, B. Zinshteyn, H. Shin, K. M. Bartoli *et al.*, 2014 Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature 515**:** 143-146.

Chiu, J. C., K. H. Low, D. H. Pike, E. Yildirim and I. Edery, 2010 Assaying locomotor activity to study circadian rhythms and sleep parameters in Drosophila. J Vis Exp.

Dawlaty, M. M., A. Breiling, T. Le, G. Raddatz, M. I. Barrasa *et al.*, 2013 Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Dev Cell 24**:** 310-323.

Dawlaty, M. M., K. Ganz, B. E. Powell, Y. C. Hu, S. Markoulaki *et al.*, 2011 Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell 9**:** 166-175.

Delatte, B., 2014 [TET-OGT interaction potentiates transcription by regulating histone H3 methylation]. Med Sci (Paris) 30**:** 619-621.

Delatte, B., F. Wang, L. V. Ngoc, E. Collignon, E. Bonvin *et al.*, 2016 RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science 351**:** 282-285.

Dietzl, G., D. Chen, F. Schnorrer, K. C. Su, Y. Barinova *et al.*, 2007 A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448**:** 151-156.

Dominissini, D., S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar *et al.*, 2012 Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature 485**:** 201-206.

Du, H., Y. Zhao, J. He, Y. Zhang, H. Xi *et al.*, 2016 YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. Nat Commun 7**:** 12626.

Dunwell, T. L., L. J. McGuffin, J. M. Dunwell and G. P. Pfeifer, 2013 The mysterious presence of a 5-methylcytosine oxidase in the Drosophila genome: possible explanations. Cell Cycle 12**:** 3357-3365.

Frye, M., and S. Blanco, 2016 Post-transcriptional modifications in development and stem cells. Development 143**:** 3871-3881.

Frye, M., S. R. Jaffrey, T. Pan, G. Rechavi and T. Suzuki, 2016 RNA modifications: what have we learned and where are we headed? Nat Rev Genet 17**:** 365-372.

Fu, L., C. R. Guerrero, N. Zhong, N. J. Amato, Y. Liu *et al.*, 2014a Tet-mediated formation of 5-hydroxymethylcytosine in RNA. J Am Chem Soc 136**:** 11582-11585.

Fu, Y., D. Dominissini, G. Rechavi and C. He, 2014b Gene expression regulation mediated through reversible m(6)A RNA methylation. Nat Rev Genet 15**:** 293-306.

Fustin, J. M., M. Doi, Y. Yamaguchi, H. Hida, S. Nishimura *et al.*, 2013 RNA-methylation-dependent RNA processing controls the speed of the circadian clock. Cell 155**:** 793-806.

Gaidzik, V. I., P. Paschka, D. Spath, M. Habdank, C. H. Kohne *et al.*, 2012 TET2 mutations in acute myeloid leukemia (AML): results from a comprehensive genetic and clinical analysis of the AML study group. J Clin Oncol 30**:** 1350-1357.

Garcia-Bellido, A., and J. F. de Celis, 2009 The complex tale of the achaete-scute complex: a paradigmatic case in the analysis of gene organization and function during development. Genetics 182**:** 631-639.

Gilbert, W. V., T. A. Bell and C. Schaening, 2016 Messenger RNA modifications: Form, distribution, and function. Science 352**:** 1408-1412.

Goll, M. G., and T. H. Bestor, 2005 Eukaryotic cytosine methyltransferases. Annu Rev Biochem 74**:** 481-514.

Grima, B., E. Chelot, R. Xia and F. Rouyer, 2004 Morning and evening peaks of activity rely on different clock neurons of the Drosophila brain. Nature 431**:** 869-873.

Grozhik, A. V., B. Linder, A. O. Olarerin-George and S. R. Jaffrey, 2017 Mapping m6A at Individual-Nucleotide Resolution Using Crosslinking and Immunoprecipitation (miCLIP). Methods Mol Biol 1562**:** 55-78.

Gu, T. P., F. Guo, H. Yang, H. P. Wu, G. F. Xu *et al.*, 2011 The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature 477**:** 606-610.

Hahn, M. A., R. Qiu, X. Wu, A. X. Li, H. Zhang *et al.*, 2013 Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. Cell Rep 3**:** 291-300.

Haussmann, I. U., Z. Bodi, E. Sanchez-Moran, N. P. Mongan, N. Archer *et al.*, 2016 m6A potentiates Sxl alternative pre-mRNA splicing for robust Drosophila sex determination. Nature 540**:** 301-304.

He, C., 2010 Grand challenge commentary: RNA epigenetics? Nat Chem Biol 6**:** 863-865.

He, Y. F., B. Z. Li, Z. Li, P. Liu, Y. Wang *et al.*, 2011 Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 333**:** 1303-1307.

Huber, S. M., P. van Delft, L. Mendil, M. Bachman, K. Smollett *et al.*, 2015 Formation and abundance of 5-hydroxymethylcytosine in RNA. Chembiochem 16**:** 752-755.

Irvine, R. A., I. G. Lin and C. L. Hsieh, 2002 DNA methylation has a local effect on transcription and histone acetylation. Mol Cell Biol 22**:** 6689-6696.

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

Ito, S., L. Shen, Q. Dai, S. C. Wu, L. B. Collins *et al.*, 2011 Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 333**:** 1300-1303.

Jaffrey, S. R., and M. G. Kharas, 2017 Emerging links between m6A and misregulated mRNA methylation in cancer. Genome Med 9**:** 2.

Jin, S. G., X. Wu, A. X. Li and G. P. Pfeifer, 2011 Genomic mapping of 5-hydroxymethylcytosine in the human brain. Nucleic Acids Res 39**:** 5015-5024.

Jones, P. A., 2012 Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13**:** 484-492.

Jones, P. L., G. J. Veenstra, P. A. Wade, D. Vermaak, S. U. Kass *et al.*, 1998 Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet 19**:** 187-191.

Kinney, S. M., H. G. Chin, R. Vaisvila, J. Bitinaite, Y. Zheng *et al.*, 2011 Tissue-specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genomes. J Biol Chem 286**:** 24685-24693.

Klein, F., P. Mahr, M. Galova, S. B. Buonomo, C. Michaelis *et al.*, 1999 A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. Cell 98**:** 91-103.

Koh, K. P., A. Yabuuchi, S. Rao, Y. Huang, K. Cunniff *et al.*, 2011 Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell 8**:** 200-213.

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

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

Law, J. A., and S. E. Jacobsen, 2010 Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11**:** 204-220.

Leiss, D., U. Hinz, A. Gasch, R. Mertz and R. Renkawitz-Pohl, 1988 Beta 3 tubulin expression characterizes the differentiating mesodermal germ layer during Drosophila embryogenesis. Development 104**:** 525-531.

Lence, T., J. Akhtar, M. Bayer, K. Schmid, L. Spindler *et al.*, 2016 m6A modulates neuronal functions and sex determination in Drosophila. Nature 540**:** 242-247.

Li, T., D. Yang, J. Li, Y. Tang, J. Yang *et al.*, 2015 Critical role of Tet3 in neural progenitor cell maintenance and terminal differentiation. Mol Neurobiol 51**:** 142-154.

Linder, B., A. V. Grozhik, A. O. Olarerin-George, C. Meydan, C. E. Mason *et al.*, 2015 Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nat Methods 12**:** 767-772.

Liu, J., Y. Yue, D. Han, X. Wang, Y. Fu *et al.*, 2014 A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol 10**:** 93-95.

Liu, K., Y. Ding, W. Ye, Y. Liu, J. Yang *et al.*, 2015a Structural and functional characterization of the proteins responsible for N6-methyladenosine modification and recognition. Curr Protein Pept Sci.

Liu, N., Q. Dai, G. Zheng, C. He, M. Parisien *et al.*, 2015b N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. Nature 518**:** 560-564.

Long, H. K., N. P. Blackledge and R. J. Klose, 2013 ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochem Soc Trans 41**:** 727-740.

Louis, M., S. Piccinotti and L. B. Vosshall, 2008 High-resolution measurement of odor-driven behavior in Drosophila larvae. J Vis Exp.

Lyko, F., B. H. Ramsahoye and R. Jaenisch, 2000 DNA methylation in Drosophila melanogaster. Nature 408**:** 538-540.

Machnicka, M. A., K. Milanowska, O. Osman Oglou, E. Purta, M. Kurkowska *et al.*, 2013 MODOMICS: a database of RNA modification pathways--2013 update. Nucleic Acids Res 41**:** D262-267.

Mellen, M., P. Ayata, S. Dewell, S. Kriaucionis and N. Heintz, 2012 MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell 151**:** 1417-1430.

Meyer, K. D., and S. R. Jaffrey, 2014 The dynamic epitranscriptome: N6-methyladenosine and gene expression control. Nat Rev Mol Cell Biol 15**:** 313-326.

Meyer, K. D., Y. Saletore, P. Zumbo, O. Elemento, C. E. Mason *et al.*, 2012 Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell 149**:** 1635-1646.

Muller, T. A., M. A. Tobar, M. N. Perian and R. P. Hausinger, 2017 Biochemical Characterization of AP Lyase and m6A Demethylase Activities of Human AlkB Homologue 1 (ALKBH1). Biochemistry 56**:** 1899-1910.

Nicholson, L., G. K. Singh, T. Osterwalder, G. W. Roman, R. L. Davis *et al.*, 2008 Spatial and temporal control of gene expression in Drosophila using the inducible GeneSwitch GAL4 system. I. Screen for larval nervous system drivers. Genetics 178**:** 215-234.

Panda, S., J. B. Hogenesch and S. A. Kay, 2002 Circadian rhythms from flies to human. Nature 417**:** 329-335.

Papamichos-Chronakis, M., and C. L. Peterson, 2013 Chromatin and the genome integrity network. Nat Rev Genet 14**:** 62-75.

Perez, C., N. Martinez-Calle, J. I. Martin-Subero, V. Segura, E. Delabesse *et al.*, 2012 TET2 mutations are associated with specific 5-methylcytosine and 5-hydroxymethylcytosine profiles in patients with chronic myelomonocytic leukemia. PLoS One 7**:** e31605.

Perrimon, N., 1998 Creating mosaics in Drosophila. Int J Dev Biol 42**:** 243-247.

Rabani, M., J. Z. Levin, L. Fan, X. Adiconis, R. Raychowdhury *et al.*, 2011 Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nat Biotechnol 29**:** 436-442.

Raddatz, G., P. M. Guzzardo, N. Olova, M. R. Fantappie, M. Rampp *et al.*, 2013 Dnmt2-dependent methylomes lack defined DNA methylation patterns. Proc Natl Acad Sci U S A 110**:** 8627-8631.

Sanson, B., P. White and J. P. Vincent, 1996 Uncoupling cadherin-based adhesion from wingless signalling in Drosophila. Nature 383**:** 627-630.

Santiago, M., C. Antunes, M. Guedes, N. Sousa and C. J. Marques, 2014 TET enzymes and DNA hydroxymethylation in neural development and function - how critical are they? Genomics 104**:** 334-340.

Santos-Rosa, H., R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein *et al.*, 2002 Active genes are tri-methylated at K4 of histone H3. Nature 419**:** 407-411.

Schaal, T. D., and T. Maniatis, 1999 Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. Mol Cell Biol 19**:** 1705-1719.

Schaefer, M., T. Pollex, K. Hanna, F. Tuorto, M. Meusburger *et al.*, 2010 RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. Genes Dev 24**:** 1590-1595.

Schwanhausser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt *et al.*, 2011 Global quantification of mammalian gene expression control. Nature 473**:** 337-342.

Schwartz, S., M. R. Mumbach, M. Jovanovic, T. Wang, K. Maciag *et al.*, 2014 Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. Cell Rep 8**:** 284-296.

Shahbazian, M. D., and M. Grunstein, 2007 Functions of site-specific histone acetylation and deacetylation. Annu Rev Biochem 76**:** 75-100.

Sharp, P. A., 2009 The centrality of RNA. Cell 136**:** 577-580.

Slobodin, B., R. Han, V. Calderone, J. A. Vrielink, F. Loayza-Puch *et al.*, 2017 Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. Cell 169**:** 326-337 e312.

Spitale, R. C., R. A. Flynn, Q. C. Zhang, P. Crisalli, B. Lee *et al.*, 2015 Structural imprints in vivo decode RNA regulatory mechanisms. Nature 519**:** 486-490.

Squires, J. E., H. R. Patel, M. Nousch, T. Sibbritt, D. T. Humphreys *et al.*, 2012 Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res 40**:** 5023-5033.

St Johnston, D., 2002 The art and design of genetic screens: Drosophila melanogaster. Nat Rev Genet 3**:** 176-188.

Stoleru, D., Y. Peng, J. Agosto and M. Rosbash, 2004 Coupled oscillators control morning and evening locomotor behaviour of Drosophila. Nature 431**:** 862-868.

Stroud, H., S. Feng, S. Morey Kinney, S. Pradhan and S. E. Jacobsen, 2011 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol 12**:** R54.

Sweatt, J. D., 2016 Neural plasticity and behavior - sixty years of conceptual advances. J Neurochem 139 Suppl 2**:** 179-199.

Szwagierczak, A., S. Bultmann, C. S. Schmidt, F. Spada and H. Leonhardt, 2010 Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res 38**:** e181.

Tahiliani, M., K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala *et al.*, 2009 Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324**:** 930-935.

Takayama, S., J. Dhahbi, A. Roberts, G. Mao, S. J. Heo *et al.*, 2014 Genome methylation in D. melanogaster is found at specific short motifs and is independent of DNMT2 activity. Genome Res 24**:** 821-830.

Vogel, C., and E. M. Marcotte, 2012 Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 13**:** 227-232.

Wang, X., Z. Lu, A. Gomez, G. C. Hon, Y. Yue *et al.*, 2014 N6-methyladenosine-dependent regulation of messenger RNA stability. Nature 505**:** 117-120.

Wang, X., B. S. Zhao, I. A. Roundtree, Z. Lu, D. Han *et al.*, 2015 N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. Cell 161**:** 1388-1399.

Williams, K., J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. Cloos *et al.*, 2011 TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature 473**:** 343-348.

Wu, H., A. C. D'Alessio, S. Ito, K. Xia, Z. Wang *et al.*, 2011 Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature 473**:** 389-393.

Wu, H., and Y. Zhang, 2011 Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. Genes Dev 25**:** 2436-2452.

Xu, T., and G. M. Rubin, 1993 Analysis of genetic mosaics in developing and adult Drosophila tissues. Development 117**:** 1223-1237.

Xu, Y., C. Xu, A. Kato, W. Tempel, J. G. Abreu *et al.*, 2012 Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. Cell 151**:** 1200-1213.

Yue, Y., J. Liu and C. He, 2015 RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. Genes Dev 29**:** 1343-1355.

Zeidler, M. P., C. Tan, Y. Bellaiche, S. Cherry, S. Hader *et al.*, 2004 Temperature-sensitive control of protein activity by conditionally splicing inteins. Nat Biotechnol 22**:** 871-876.

Zhang, G., H. Huang, D. Liu, Y. Cheng, X. Liu *et al.*, 2015 N6-methyladenine DNA modification in Drosophila. Cell 161**:** 893-906.

Zhang, R. R., Q. Y. Cui, K. Murai, Y. C. Lim, Z. D. Smith *et al.*, 2013 Tet1 regulates adult hippocampal neurogenesis and cognition. Cell Stem Cell 13**:** 237-245.

Zhong, X., Q. Q. Wang, J. W. Li, Y. M. Zhang, X. R. An *et al.*, 2017 Ten-Eleven Translocation-2 (Tet2) Is Involved in Myogenic Differentiation of Skeletal Myoblast Cells in Vitro. Sci Rep 7**:** 43539.

Zhou, J., J. Wan, X. Gao, X. Zhang, S. R. Jaffrey *et al.*, 2015 Dynamic m(6)A mRNA methylation directs translational control of heat shock response. Nature 526**:** 591-594.