Characterization of mammalian NEF-sp a testis-specific 3'→5' exoribonuclease

Sara Moreira da Silva

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Biologie Moléculaire**

Arrêté ministériel : 25 mai 2016

Présentée par

Sara Moreira da Silva

Thèse dirigée par **Dr. Ramesh Pillai**

Préparée au sein du **European Molecular Biology Laboratory (EMBL), Grenoble**

Dans **l'École Doctorale de Chimie et Science du Vivant**

Caractérisation ARN Exonucléase 5/NEF-sp: une exoribonucléase 3'→ 5' spécifique du testicule

Thèse soutenue publiquement le **29 Juin 2017**
devant le jury composé de :

Prof. Winfried Weissenhorn
Président du jury
Prof. Dónal O'Carroll
Rapporteur
Dr. Alena Shkumatava
Rapporteur
Dr. André Verdel
Examinateur

Résumé

Les éléments transposables sont des parasites génomiques qui ont la capacité de se déplacer d'un endroit à l'autre dans le génome et peuvent donc menacer l'intégrité de l'hôte. Environ 50% des génomes de la souris et de l'homme sont composés de ces éléments répétitifs. La voie de l'ARNpi a un rôle essentiel dans la protection de la lignée germinale contre l'activité préjudiciable des transposons. Les ARNpi sont une classe de petits ARN non codants qui se chargent dans le sous-groupe PIWI des protéines Argonaute. Ce complexe de silencieux induit par ARNpi (piRISC) peut transposer le silence au niveau transcriptionnel ou post-transcriptionnel. La perturbation de la voie de l'ARNpi dans la souris entraîne l'activation du transposon dans les testicules, les défauts de la spermatogenèse et la stérilité spécifique chez l'homme.

La voie piRNA a été intensément étudiée au cours de la dernière décennie. Plus de vingt protéines différentes ont déjà été impliquées. Cependant, certaines caractéristiques mécanistes ne sont pas entièrement comprises. Au cours de la biogenèse primaire de ARNpi, un long précurseur primaire de ARNpi est traité par une endonucléase à des fragments plus courts appelés pré-ARNpi. Ceux-ci sont généralement plus longs que les ARNpi matures. Par conséquent, après le chargement dans les protéines PIWI, leur extrémité 3' doit être taillée à la taille correcte. Bien que la coupe ait déjà été observée *in vitro*, la protéine responsable de la maturation des ARNpi était encore inconnue.

Notre hypothèse initiale était que la nucléase de mammifère non caractéristique appelée NEF-sp était la nucléase impliquée dans la maturation des ARNpi. Dans la première partie de cette thèse, nous avons montré que Nef-sp est exclusivement exprimé dans les testicules de souris. hNEF-sp a une activité exoribonucléase 3'→ 5' et active sur les substrats ssRNA. Nous avons également démontré que le NEF-sp devrait fonctionner dans le compartiment nucléaire/nucléolaire. Notre analyse de la souris Nef-sp mutante n'a révélé aucun phénotype évident. Nous avons observé que les homozygotes des deux sexes sont viables et présentent une fertilité normale. Les éléments transposables ne sont pas actifs dans les cellules germinales masculines et la spermatogenèse n'est pas affectée. Par conséquent, nous avons conclu que NEF-sp ne joue pas un rôle dans la voie de l'ARNpi. NEF-sp peut être important pour le découpage précis du (des) substrat(s) inconnu(s). Le manque de phénotype dans les souris mutantes pourrait être dû à une complémentation potentielle par d'autres nucléases. Néanmoins, notre étude fournit une caractérisation biochimique et génétique de l'exoribonucléase NEF-sp mammalienne.

Dans la deuxième partie de cette thèse, nous avons étudié un nouveau type de méthylation de l'ADN, le marqueur N6-méthyladénine (6mA), récemment démontré être présent dans le génome de mammifère. Cette protéine devait être impliquée dans le dépôt de marqueur N6-méthyladénine (6mA) chez les mammifères. Ce type de modification de l'ADN a été récemment étudié comme un marqueur de méthylation présent dans le génome des eucaryotes supérieurs. Cependant, son rôle est encore

controversé et une analyse plus détaillée est nécessaire pour rassurer l'importance de ce marqueur. Tout au long du deuxième projet, nous avons essayé de comprendre si METTL4 joue un rôle dans le dépôt de 6mA chez les mammifères et les mouches. Nous avons essayé de caractériser biochimiquement et structurellement la souris METTL4. En outre, nous avons analysé sa signification biologique en utilisant des souris mutantes et des mouches en tant que système modèle. Nous ne pouvons pas montrer que METTL4 est une ADN-méthyltransférase active *in vitro*. Les animaux mutants sont également viables et fertiles. Par conséquent, nous n'avons pas été en mesure de comprendre le rôle de METTL4 et l'importance biologique, le cas échéant, de la N6-méthyladénine.

Summary

Transposable elements are genomic parasites that have the ability to move from one location to another in the genome. Therefore, when these elements are active, they can threaten the integrity of the host genome. Around 50% of mouse and human genomes are composed by these repetitive elements. piRNA pathway plays an essential role in protecting the germline against the detrimental activity of transposons. piRNAs are a class of small non-coding RNAs that bind the PIWI subclade of Argonaute proteins. These piRNA-induced silencing complexes (piRISCs) can silence transposons either at transcriptional or post-transcriptional level. Disruption of piRNA pathway in mice results in transposon activation in the male germline leading to defects in spermatogenesis and male sterility.

piRNA pathway has been intensively studied during the last decade. More than twenty different proteins have been shown to be important to ensure successful transposon silencing. However, some of the mechanistic features are still not fully understood. During primary piRNA biogenesis, a long primary piRNA precursor is processed by an endonuclease to shorter fragments called pre-piRNAs. These are generally longer than mature piRNAs. Therefore, after loading into PIWI proteins, their 3' end needs to be trimmed to the correct size. Although trimming has already been observed *in vitro*, the protein responsible for maturation of piRNAs was unknown.

Our initial hypothesis was that an uncharacterized mammalian nuclease called NEF-sp was involved in piRNA maturation. In the first part of the thesis, we have shown that *Nef-sp* is exclusively expressed in mouse testes. hNEF-sp has 3′→5′ exoribonuclease activity and is active on ssRNA substrates. We also demonstrated that NEF-sp is likely to function in the nuclear/nucleolar compartment. Our analysis of the *Nef-sp* mouse knock-out mutant revealed no obvious phenotype. We observed that homozygous animals of both sexes are viability and display normal fertility. Transposable elements are not active in male germ cells and spermatogenesis is not affected. Therefore, we concluded that NEF-sp does not play a role in piRNA maturation. NEF-sp may be important for precise trimming of the unknown substrate(s). Lack of phenotype in the mutant mice might be due to potential complementation by other nucleases. Nevertheless, our study provides a biochemical and genetic characterization of the mammalian NEF-sp.

In the second part of the thesis we explored the role of N^6–methyladenine mark (6mA), a new type of DNA methylation that has been recently shown to be present in mammalian genome. The functional role of a putative DNA methyltransferase called METTL4 was investigated. This protein was predicted to be involved in the deposition of the 6mA mark in the mammalian genome. Throughout the second project, we tried to understand if METTL4 plays a role in the deposition of 6mA mark. We have tried to biochemically and structurally characterize mouse METTL4. In addition, we studied its biological significance in mutant mice and flies. METTL4 did not show DNA methyltransferase activity in the *in vitro* tested conditions. The mutant animals were also viable and fertile. METTL4 function remains elusive as well as its relevance for the deposition of the N^6–methyladenine mark.

Abbreviations

18S rRNA	18S ribossomal RNA
28S	28S ribossomal RNA
3' UTR	Three prime untranslated region
5mC	5-methylcytosine
6mA	N^6–methyladenine
6xHis	Polyhistidine-tag

A

A	Adenine
A-bias	Adenine bias
ADN	Acide désoxyribonucléique
Ago	Argonaute
ARN	Acide ribonucléique
ARNpi	ARN interagissants avec Piwi
ATP	Adenosine triphosphate
Aub	Aubergine

B

Bm	*Bombyx mori*
BmN4	Bombyx mori ovary-derived cell line
BMP	Bone morphogenetic protein
BSA	Bovine serum albumin

C

cDNA	Complementary DNA
CGIs	CpG islands
CNT	Control

D

DEDD	
Dm	*Drosophila melanogaster*
DMAD	*Drosophila* DNA 6mA demethylase
DNA	Deoxyribonucleic acid
DNMT1	DNA methyltransferase 1
DNMT2	DNA methyltransferase 2
DNMT3A	DNA methyltransferase 3A

DNMT3B	DNA methyltransferase 3B
DNMT3L	DNA methyltransferase 3 Like
dpc	Days post coitum
dsRNA	Double-stranded RNA
DTT	Dithiothreitol

E

E. coli	*Escherichia coli*
E16.5	Embryonic stage 16.5
EDTA	Ethylenediamine tetraacetic acid
ES cells	Embryonic stem cells
ESRF	European Synchrotron Radiation Facility

F

FL	Full-length

G

gRNA	guide RNA

H

HA	Influenza hemagglutinin epitope
hNEF-sp	Human NEF-sp
HTX	High-throughput crystallization facility

I

IAP	Intracisternal A-particle
IBS	Institut de Biologie Structurale
ICM	Inner cell mass
IF	Immunofluorescence
IgG	Immunoglobulin G
IPTG	isopropyl-β-D-thiogalactopyranoside

K

Kb	Kilobase
kDa	Kilodalton
KI	Knock-in
KO	Knock-out

L

L	Liter
LC-TOF	Liquid chromatography time-of-flight
LINE-1	Long Interspersed Nuclear Element 1
LTR	Long terminal repeats

M

M	Molar
MeRIP-seq	Methylated RNA immunoprecipitation followed by sequencing
Mg^{2+}	Magnesium Ion
miRNA	micro RNA
mL	Mililiter
mM	Milimolar
Mn^{2+}	Manganese Ion
mNEF-sp	Mouse NEF-sp
MS	Mass spectrometry
MTase	Methyltransferase catalytic domain

N

N	N-peptide
Ni-NTA	Nickel-charged affinity resin
NLS	Nuclear localization signal
NMAD-1	N6-methyl adenine demethylase 1
NMD	Nonsense-mediated mRNA decay
nt	Nucleotides

O

ORF	open reading frame

P

P	Pellet
P14	14 days after birth
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PGCs	Primordial germ cells
piRISC	piRNA-induced silencing complex

piRNA	Piwi-interacting RNA
PIWI	P-element induced wimpy testis
PMSF	Phenylmethylsulfonyl fluoride
PolII	Polymerase II

R

rDNA	Ribossomal DNA
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNP	Ribonucleoprotein complex
RRM	RNA Recognition Motif
rRNAs	Ribossomal RNAs
RT	Room temperature
RT-PCR	Reverse Transcriptase-PCR

S

SDN1	SMALL RNA DEGRADING NUCLEASE 1
SDS	Sodium dodecyl sulfate
Sf21	Spodoptera frugiperda 21
SINE	Short interspersed nuclear elements
SNP	Supernatant
SNP	Supernatant
ssRNA	Single-stranded RNA
Strep	Streptavidin

T

TE	Transposon elements
TEV	Tobacco Etch Virus
TGS	Transcriptional gene silencing
tRNAs	Transfer RNAs
TSS	Transcription start site

U

U-bias	Uridine bias
UPF2	Up-Frameshift Suppressor 2
UPF3b	Up-Frameshift Suppressor 3b
UTR	Untranslated region

UV	Ultraviolet

W

WB	Western blot

Z

Zn^{2+}	Zinc Ion
µL	Microliter
µM	Micromolar

Table of Contents

I. INTRODUCTION

PART I

piRNA pathway has been one of the main research interests of our group. One of the projects that I developed during my PhD "**Characterization of mammalian NEF-sp as a testis-specific 3' → 5' exoribonuclease"** was based on the hypothesis that NEF-sp could potentially act as a nuclease involved in piRNA maturation. Therefore, in the first part of the introduction I will focus on piRNA biogenesis and its role in genome defence and importance for spermatogenesis in mice.

1.1 piRNA pathway

A large proportion of the mouse and human genome is composed by repetitive elements (Waterston *et al.* 2002). Some of these elements have the ability to move from one chromosomal location to another and can therefore threaten the integrity of the host genome. Piwi-interacting RNAs (piRNAs) are a class of small non-coding RNAs that associate with PIWI proteins, a special clade of the Argonaute family that is mainly expressed in the animal gonads (Aravin *et al.* 2006, Girard *et al.* 2006, Grivna *et al.* 2006). piRNAs and PIWI proteins exert essential roles in preserving the genome integrity of developing germ cells by silencing transposable elements (TEs) (Aravin *et al.* 2007a, Siomi *et al.* 2011). Cytoplasmic PIWI loaded with piRNAs is guided by base complementarity to transposon mRNAs and the targets are destroyed by direct cleavage. Transposons are also silenced at transcriptional level by nuclear piRNA-guided PIWI proteins. These can promote DNA methylation or establishment of repressive chromatin marks at transposon promoter regions. Studies in flies, mice and other organisms were crucial to extensively characterize piRNA biogenesis, its functional role and mechanisms of action.

piRNAs were first found in *Drosophila melanogaster* (Aravin *et al.* 2001) and later shown to associate with PIWI proteins (Saito *et al.* 2006, Vagin *et al.* 2006). Flies express three PIWI proteins: Piwi, Aubergine (Aub) and Ago3. These three proteins are essential for transposon silencing and were shown to be non-redundant (Vagin *et al.* 2004, Kalmykova *et al.* 2005, Savitsky *et al.* 2006, Li *et al.* 2009). Piwi (Cox *et al.* 1998) and Aub (Schmidt *et al.* 1999) are required for fertility in males and females and Ago3 is important for female fertility (Li *et al.* 2009). Fly ovaries are composed of two types of cells: the germline that includes the oocyte and associated nurse cells, and the somatic follicular cells that surround the germ cells. The expression pattern of PIWI proteins is different in these two cell types, Piwi is expressed in both somatic and germ cells and localizes mainly in nucleus. Aub and Ago3 are exclusively present in the germline and localize in the perinuclear cytoplasmic granules called *nuage*.

In mice, three PIWI proteins: MIWI (Deng *et al.* 2002), MILI (Kuramochi-Miyagawa *et al.* 2004) and MIWI2 (Carmell *et al.* 2007) are mainly expressed in testes and each of these are essential for effective transposon silencing and for male fertility (Figure 1.1C). These three PIWI proteins are expressed at different stages during spermatogenesis. Expression of MILI starts during embryonic development and continues after birth. MIWI2 has a narrower expression pattern from E15 to 3 days post-partum (dpp). MIWI is expressed from 14 dpp, which coincides with the pachytene stage of meiosis, and continues to be expressed into the adult stage. Lack of MILI (Aravin *et al.* 2007b, De Fazio *et al.* 2011) and/or MIWI2 (Carmell *et al.* 2007, Kuramochi-Miyagawa *et al.* 2008) leads to activation of long interspersed nuclear elements (LINE-1) and intracisternal A particle (IAP) transposons and germ cell development arrests before the pachytene stage of meiosis. MIWI mutation also leads to LINE-1 activation and spermatogenesis arrests at early spermatid stage (Deng *et al.* 2002, Reuter *et al.* 2011). The dominant length of piRNAs that bind each member of mouse PIWI proteins is different. Mainly 26nt, 28nt and 30nt size piRNAs bind MILI, MIWI2 and MIWI, respectively. PIWI proteins also bind different classes of piRNAs. MILI and MIWI2 bind the so called pre-pachytene piRNAs that are expressed early in development and are enriched for transposon sequences. The pachytene piRNAs, whose expression starts later at pachytene stage, associate with MIWI and MILI. These are derived from unique intergenic loci and are depleted of repetitive elements. In mice, PIWI proteins also have distinct cellular localization. During fetal piRNA pathway, MILI localizes in cytoplasmic granules called inter-mitochondrial cement (also referred pi-body or *nuage*). MIWI2 shuttles between piP-bodies and the nucleus. The piP-bodies often overlap with the pi-bodies where MILI accumulates (Aravin *et al.* 2009). In late meiotic spermatocytes and haploid spermatids MILI and MIWI accumulate in the chromatoid body (Kotaja *et al.* 2006).

1.1.1 Mouse germ cell development and importance of piRNA pathway for spermatogenesis

At embryonic day E7.5 a small subset of cells derived from the epiblast cells are selected to become the primordial germ cells (PGCs) (McLaren 2003). The PGCs are committed to originate the future gametes and upon arrival to the genital ridge (at E11.5), they start to differentiate in a sex-specific manner. In males, PGCs become surrounded by somatic supporting cells (Sertoli cell precursors) and form the seminiferous cords which will create the future seminiferous tubules (Cool *et al.* 2012). At this stage, the PGCs go through some morphological changes developing into gonocytes. After some rounds of proliferation, gonocytes arrest at the G0/G1 phase of mitosis (at E13.5) becoming prospermatogonia cells. After birth (P3), prospermatogonia resume mitotic proliferation giving rise to spermatogonia stem cells (A). In the accepted model of spermatogonial renewal and differentiation (Figure 1.1B) (Huckins 1971, Oakberg 1971), the A_s (single) spermatogonia are considered the stem cells of spermatogenesis. When A_s spermatogonia divide, either two new stem cells are originated or an A_{pr} (pair) spermatogonia are formed. Nearly half of the divisions are self-renewing producing two new stem cells. The A_{pr} spermatogonia cells stay interconnected by cellular bridges and further divide originating chains of 4, 8 or 16 A_{al} (aligned) spermatogonia. The A_{al} spermatogonia differentiate into spermatogonia type A (A1-A4), followed by an intermediate state (A_{In}) and finally becoming B spermatogonia. All these different classes of spermatogonia are characterized by differences in heterochromatin presence and distribution. The final mitotic division of B spermatogonia generates two pre-leptotene spermatocytes that enter meiotic division (P10). Spermatocytes go through meiotic prophase which includes leptotene (P10), zygotene (P12), pachytene (P14) and diplotene (P18) phases (Bellve *et al.* 1977). At the end of meiotic divisions four haploid round spermatids are formed (P20). The spermiogenesis process starts: the nucleus of spermatids is remodelled and DNA is extensively compacted with protamines, the acrosome is formed as well as the flagellum and mature spermatozoa are obtained (P30). The spermatogenesis culminates with the release of the mature spermatozoa into the lumen of the seminiferous epithelium (Figure 1.1A).

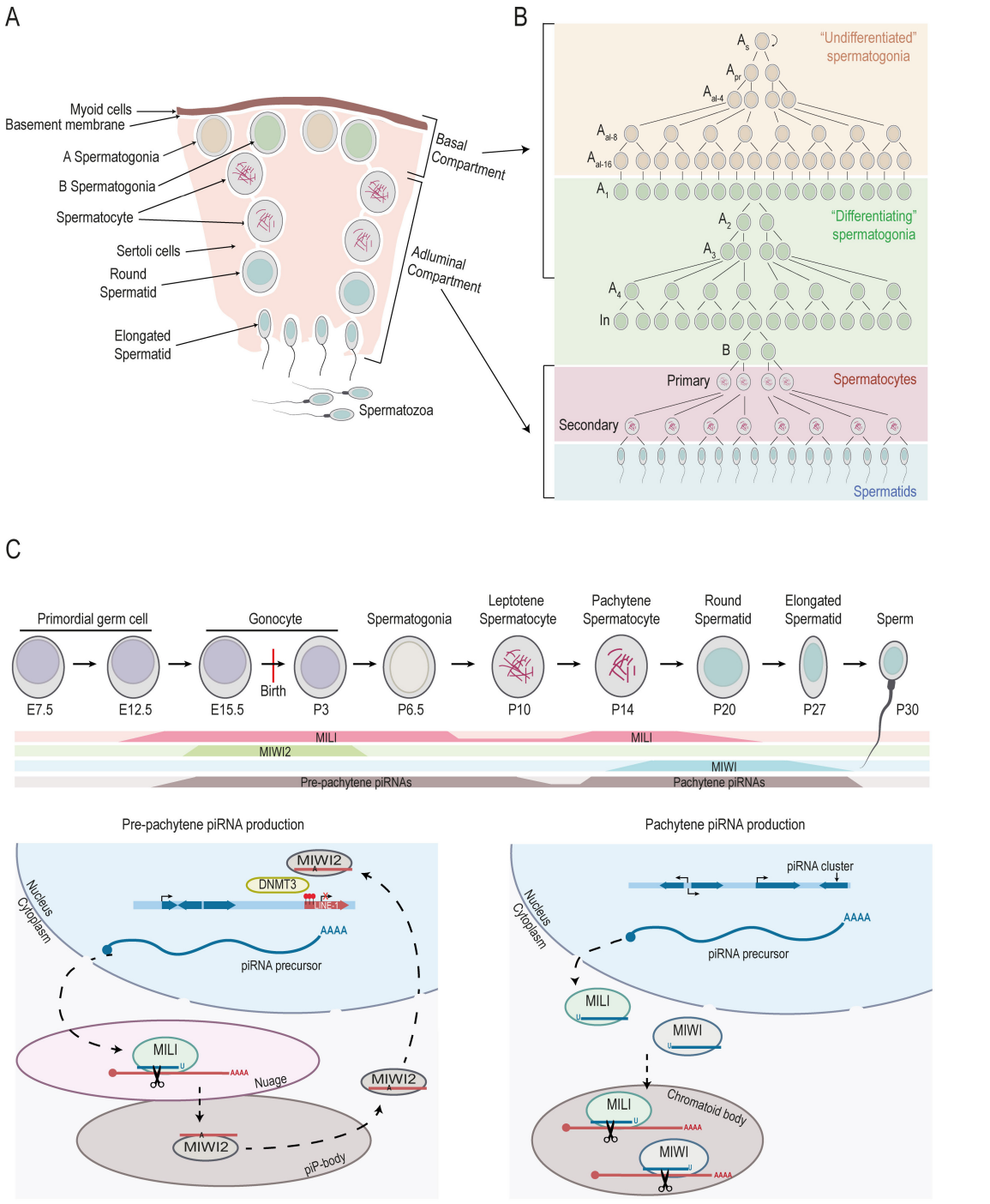


Figure 1.1 – Spermatogenesis in mice. (A) Schematic representation of a cross-section through a mouse seminiferous tubule. Spermatogonia lie on the basal membrane, followed by spermatocytes and spermatids. Differentiation progresses towards the lumen of seminiferous tubules and culminates with the release of spermatozoa. Sertoli cells compose the supporting layer were germ cells are embedded. (B) Diagram showing a model for self-renewal and differentiation of spermatogonia stem cells (see text for details) adapted from (Yoshida 2010). (C) Expression pattern of MILI, MIWI2, MIWI and piRNAs during spermatogenesis. In the pre-pachytene stage, MILI loaded with piRNAs is guided to transposon mRNAs and inactivates the target by slicing. MIWI2-piRNA complexes are guided to the nucleus where promotes DNA methylation at active transposon loci. In the pachytene stage, both MILI and MIWI silence transposon post-transcriptionally.

In mice, a functional piRNA pathway is essential for the formation of spermatozoa and consequently for male fertility (Figure 1.3C). During embryonic development, PGCs undergo extensive epigenetic reprogramming that involves global DNA demethylation. This event occurs around E12.5-E13.5 and includes demethylation of retrotransposons elements such as IAP and LINE-1 (Tanaka *et al.* 2000, Lane *et al.* 2003). DNA methylation is re-stablished in the non-dividing prospermatogonia around E16.5–E18.5 by the two *de novo* DNA methyltransferases DNMT3A and DNMT3B. The catalytically inactive DNMT3L also plays an essential regulatory role (Bourc'his *et al.* 2004, Hata *et al.* 2006, Kato *et al.* 2007). MIWI2 is expressed from E15 to 3 days after birth, which coincides with the establishment of *de novo* patterns of DNA methylation in prospermatogonia. In addition, MIWI2 loaded with piRNAs accumulates in the nucleus (Aravin *et al.* 2008). In the *Miwi2* knock-out mice, germ cells enter prophase I of meiosis but cannot progress further than the zygotene stage. *Miwi2* mutants also lose germ cells over time and tubules contain only somatic Sertoli cells suggesting that *Miwi2* is also important for germline stem cell maintenance (Carmell *et al.* 2007). Analysis of the methylation status of transposable elements such as LINE-1 and IAP elements in the *Miwi2* knock-out testis revealed that the methylation is diminished at these loci. MIWI2 loaded with piRNAs is guided to active transposon loci and promotes DNA methylation at those sites, silencing transposable elements at transcriptional level (Kuramochi-Miyagawa *et al.* 2008). Therefore, MIWI2 is essential for transposon silencing in testes and for proper formation of spermatozoa.

MILI is expressed during two distinct stages of spermatogenesis. It is first detected in mitotic spermatogonia and its expression declines with the onset of meiotic division to reappear at late zygotene stage. MILI reaches the highest levels at the pachytene stage of meiosis and stops to be expressed at the early round spermatid stage (Di Giacomo *et al.* 2013). MILI and MIWI2 are co-expressed during early germ cell development at the moment that epigenetically silencing of transposable elements is critical. *Mili* knock-out mice also revealed substantial DNA demethylation at LINE-1 (Aravin *et al.* 2007b) and IAP elements (Kuramochi-Miyagawa *et al.* 2008) and shows germ cell arrest similar to *Miwi2* knock out mice. Contrary to MIWI2, MILI was shown to silence transposons post-transcriptionally. MILI loaded with piRNAs is guided to transposon mRNAs and inactivates them by cleavage.

Both *Miwi* knock-out (Deng *et al.* 2002) and *Miwi* catalytic mutant mice (Reuter *et al.* 2011) were shown to have defects during spermatogenesis and the process is arrested at the round spermatids stage. MIWI is only expressed after birth from the pachytene stage of meiosis (P14), and more abundantly in the round spermatids, until the elongating spermatids (Deng *et al.* 2002). MIWI binds pachytene piRNAs which are derived from unique intergenic loci and not so abundant in repetitive elements. However, MIWI also associates with piRNAs derived from repeat sequences and its slicer activity was shown to be essential for LINE-1 silencing (Reuter *et al.* 2011). MIWI is therefore important for transposons post-transcriptional silencing.

1.1.2 piRNA biogenesis

piRNA sequences are present in the cell in several million individual copies. They are derived from specific genomic loci, the piRNA clusters, which can range from few kilobases (kb) to 200 kb in length. piRNAs can be generated from only one strand of a piRNA cluster (uni-strand cluster) or from both strands (dual-strand cluster) (Aravin *et al.* 2006, Girard *et al.* 2006, Brennecke *et al.* 2007). In *D. melanogaster*, the germline clusters are dual-stranded and therefore generate piRNAs from both sense and antisense strands. However, the somatic cells only produce piRNAs from one strand. Such a cluster is for instance *flamenco* (*flam*), which is exclusively expressed in ovarian somatic cells (Brennecke *et al.* 2007). In mice, most of the pachytene piRNA clusters produce piRNAs from a single genomic strand (Aravin *et al.* 2008). piRNAs can also be produced from both strands (bi-directional cluster) in a divergent manner (Aravin *et al.* 2006, Girard *et al.* 2006).

Analysis of piRNA biogenesis in different organisms led to the proposal of three biogenesis pathways: the primary processing pathway, the secondary processing pathway (also called ping-pong amplification) and the inchworming pathway. The primary biogenesis process starts with the transcription of a long primary piRNA precursor by RNA PolII that is later capped at the 5' end and polyadenylated at 3' end. In *D. melanogaster*, primary piRNA precursors are generally antisense to transposons and derived from piRNA clusters (Brennecke *et al.* 2007). In mice, primary pre-pachytene piRNAs have a sense orientation and are derived from transposon sequences. However, pachytene piRNAs are transcribed from defined unique genomic regions called pachytene piRNA clusters and vast majority only match to the same sites from which they are produced (Aravin *et al.* 2008). The primary transcripts are exported to the cytoplasm and processed to smaller pre-piRNAs by an endonuclease. Studies performed in *D. melanogaster* and mice identified a nuclease called Zucchini (Pane *et al.* 2007, Nishimasu *et al.* 2012) or its mammalian homolog MitoPLD (Watanabe *et al.* 2011, Ipsaro *et al.* 2012) to be essential for primary piRNA biogenesis. One of the features that primary piRNAs exhibit is the prevalence of a uridine nucleotide (U) at the 5' end. This 1U preference was thought to be a feature of the endonuclease involved in the processing of the long primary transcript. However, biochemical and structural studies of Zucchini could not explain this 1U – bias (Ipsaro *et al.* 2012, Nishimasu *et al.* 2012). Genetic studies suggested that an additional factor might direct Zucchini to cleave before U (Han *et al.* 2015, Mohn *et al.* 2015). Additionally, pre-piRNAs with a 1U nucleotide are preferentially loaded into PIWI proteins (Kawaoka *et al.* 2011, Cora *et al.* 2014). After loading into PIWI proteins, pre-piRNAs are further trimmed at the 3' end (Izumi *et al.* 2016) to the mature piRNA size. The 3' end is 2'-O-methylated by HEN1 creating the mature and functional PIWI-piRNA complexes (Horwich *et al.* 2007, Kirino *et al.* 2007, Saito *et al.* 2007). In *D. melanogaster* only Piwi and Aub get loaded with piRNAs produced by primary biogenesis pathway (Figure 1.2). In mice, MILI is the main recipient of these piRNAs in the first phase of piRNA production (pre-pachytene). However, in the pachytene phase MILI and MIWI get both loaded with primary piRNAs (Figure 1.3).

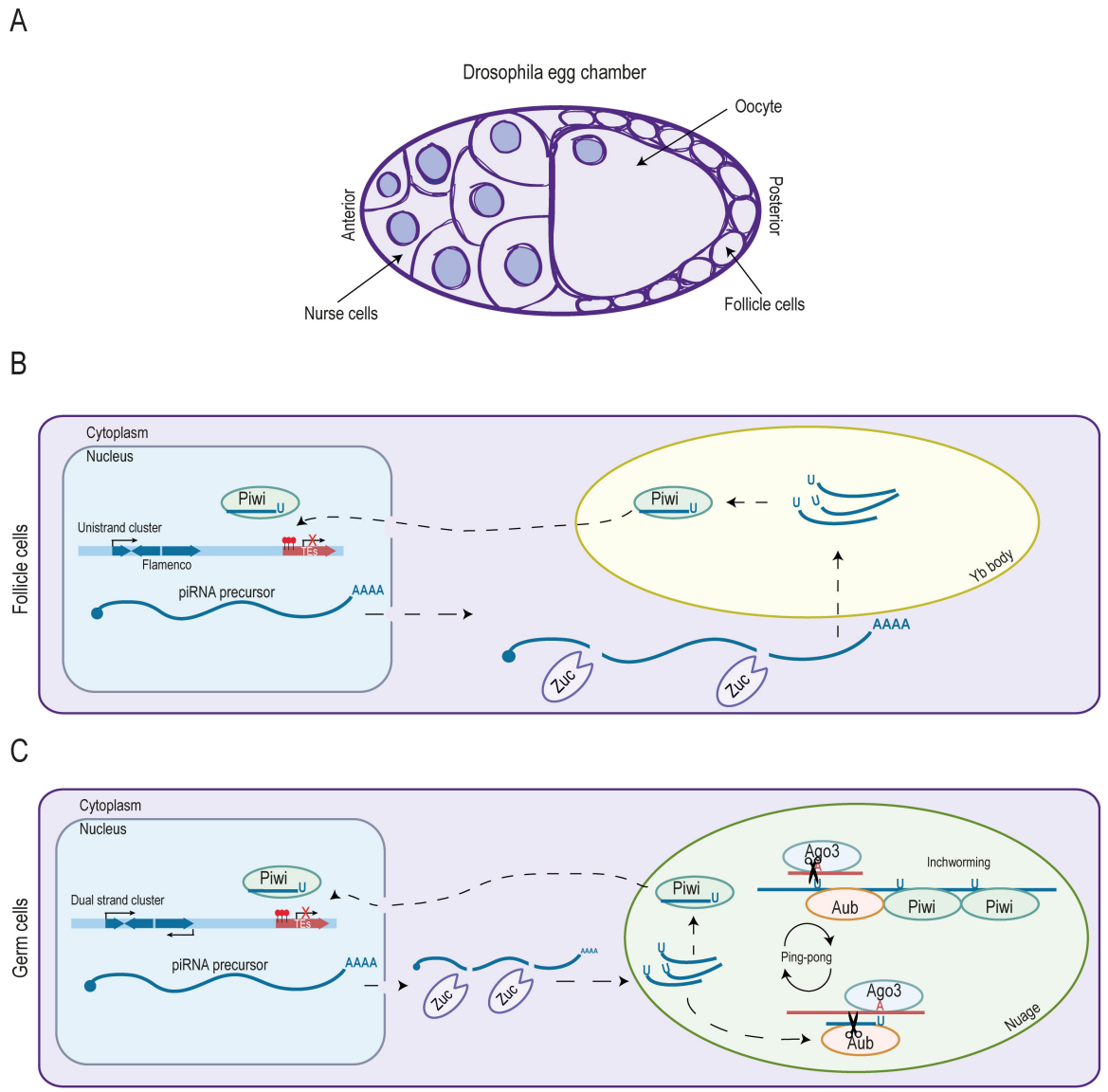


Figure 1.2 – Biogenesis pathway of *Drosophila melanogaster* piRNAs. (A) Representation of a *D. melanogaster* egg chamber at stage 10. Two types of cells are represented: the germline that includes the oocyte and associated nurse cells and the somatic follicular cells that surround the germ cells. (B) In the follicle cells, piRNAs are produced by the primary processing pathway. Only Piwi is expressed in these cells. Transcripts of piRNA clusters are processed into piRNA intermediates by the endonuclease Zucchini (Zuc). These piRNA intermediates are loaded onto Piwi. In the nucleus, the Piwi–piRNA complexes silence transposable elements by promoting deposition of repressive histone marks. (C) Besides Piwi, also Aub and Ago3 are expressed in the germline. These two proteins localize in the *nuage*. piRNAs are produced by the primary biogenesis pathway as well as the secondary processing pathway and the inchworming pathway. Aubergine (Aub) loaded with piRNAs produced by the primary biogenesis serves as a trigger to start the ping-pong amplification pathway. Secondary piRNAs get loaded into Ago3. Ago3–piRNA complexes are guided to an antisense transcript. After Ago3 slicing, the 5' end of a new piRNA is generated that gets loaded into Aub. The 5' end of the cleaved fragment is used to produce the inchworming-piRNAs that get loaded into Piwi.

The secondary piRNA processing, were the piRNA production depends on existing primary piRNAs, was first described in *Drosophila melanogaster* (Brennecke *et al.* 2007, Gunawardane *et al.* 2007). Analysis of the features of piRNAs bound to Ago3 and Aub allowed to establish a model called ping-pong cycle. While piRNAs bound to Aub are usually antisense and have a preference for a uridine at position 1 (1U-bias), the Ago3 bound piRNAs are sense and have a strong A10-bias. Moreover, the first 10 nt of piRNAs associated with Ago3 are complementary to the first 10 nt of Aub bound piRNAs. Altogether, these observations allowed establishing the following model: Aub is loaded with antisense piRNAs generated through the primary biogenesis pathway. Based on the sequence complementarity this Piwi-piRNA complex is guided to a complementary target, a sense transposon mRNA, which is cut by the slicer activity of Aub at 10 nt distance from the 5' end of the guide piRNA. The 5' cleavage fragment of the transposon transcript is then loaded onto Ago3 and matures as a secondary piRNA. This secondary piRNA guides Ago3 to cleave an antisense transcript which results in production of piRNA of the same sequence as the initial piRNA that started the cycle. Therefore, this so-called ping-pong cycle is considered an amplification step in which the initial piRNA pool, generated by the primary pathway, is amplified upon the encounter of transposon mRNA (Figure 1.2C).

In mice, the ping-pong cycle only contributes to the production of piRNAs during the pre-pachytene phase. At this stage, only MILI and MIWI2 are expressed. MILI is generally loaded with sense piRNAs derived mainly from transposable elements. These piRNAs have a strong 1U – bias. The functional MILI-piRNA complexes are guided by base complementarity to an antisense transcript. MILI has slicer activity and therefore cleaves the antisense transcripts generating the 5' end of a second piRNA. This secondary piRNA exhibit the ping-pong signatures similarly to *D. melanogaster*: bias for A at position 10 and an overlap of 10 nucleotides between the primary and the secondary piRNA (Aravin *et al.* 2007b). The antisense secondary piRNAs get loaded into MIWI2 (Aravin *et al.* 2008) and into MILI (intra-MILI ping-pong cycle) (De Fazio *et al.* 2011) (Figure1.3).

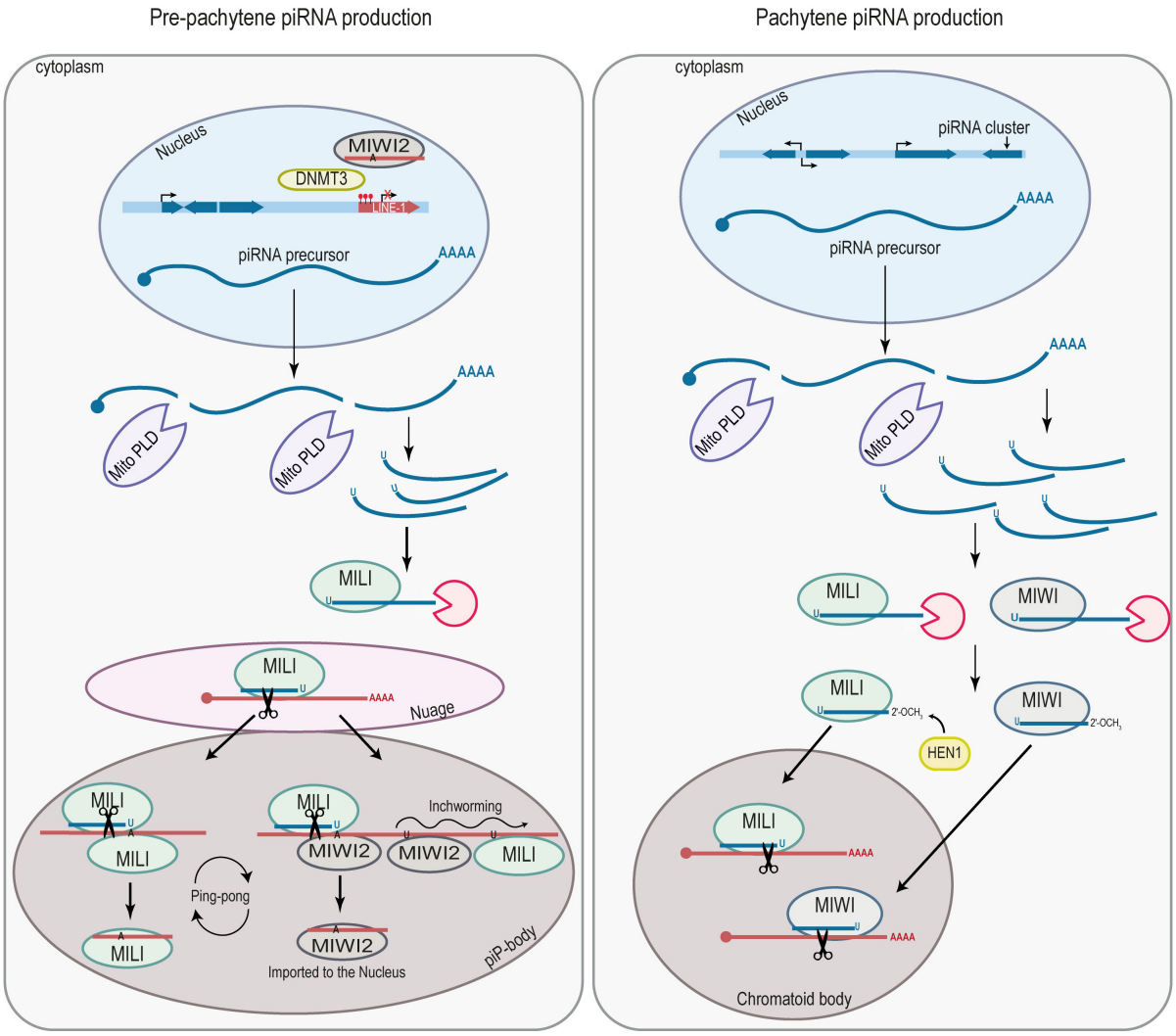


Figure 1.3 – piRNA biogenesis pathways in mice. Primary biogenesis starts with the transcription of long primary piRNA precursors. These are processed by MitoPLD in the cytoplasm to pre-piRNAs. At the pre-pachytene stage, pre-piRNAs get loaded into MILI. During the pachytene stage, both MILI and MIWI get loaded with pre-piRNAs. The 3' end of pre-piRNAs is trimmed to the mature length and 2'-O-methylated by HEN1. These functional PIWI-piRNA complexes can silence transposable elements post-transcriptionally. Secondary processing pathway and the inchworming pathway only produces piRNAs during the pre-pachytene phase. MILI loaded with primary piRNAs is guided to an antisense transcript. After cleavage, the 5' end of a second piRNA is generated and gets loaded into MIWI2 or MILI. The remaining 5' end of the MILI-cleaved transcript is used to produce the inchworming piRNAs that get loaded into MILI and MIWI2 without any apparent bias.

A third biogenesis pathway, called phased piRNA biogenesis (Han *et al.* 2015, Mohn *et al.* 2015) or inchworm processing (Homolka *et al.* 2015) was recently found in both flies and mice (Yang *et al.* 2016). This third piRNA biogenesis pathway is connected to the ping-pong cycle as it is initiated by piRNA guided slicing. As described above, in *D. melanogaster* the secondary piRNAs are generated during the ping-pong cycle and get loaded into Ago3 which is guided to the antisense transcript. The Ago3 cleaves the transcript and creates the 5' end of a new piRNA that gets loaded into Aub. However the remaining portion of the 5' cleaved fragment is not degraded but is used to produce the phased-piRNAs in a Zucchini dependent manner (Han *et al.* 2015, Mohn *et al.* 2015). These phased-piRNAs get loaded mainly in Piwi which can enter the nucleus and promote silencing of transposable elements. This mechanism creates most of the germ cell Piwi piRNAs (Figure 1.2C). A similar mechanism exists in mice, where MILI slicing generates the 5' cleavage fragment which is used for production of the secondary piRNA followed by the set of phased-piRNAs (Yang *et al.* 2016). The phased-piRNAs are loaded both onto MILI and MIWI2. MIWI2 loaded with phased piRNAs enter the nucleus and silences transposons by promoting DNA methylation and deposition of chromatin repressive marks. As only few MIWI2 piRNAs exhibit the ping-pong signatures (10A-bias and 10 nt overlap) (Aravin *et al.* 2008), the phased-piRNA production is likely to be the main mechanism responsible for biogenesis of MIWI2 piRNAs (Figure 1.3). Therefore, cytoplasmic slicing of the transposons not only degrades transposon transcripts but is also essential for transcriptional silencing of transposons by enhancing the repertoire of piRNAs bound to nuclear Piwi proteins.

1.1.3 Role of nucleases in piRNA pathway

Biochemical and genetic studies allowed identifying new proteins that play an important role in piRNA pathway. Several classes of proteins such as RNA helicases (Wenda *et al.* 2017) and Tudor domain-containing proteins (Reuter *et al.* 2009, Pandey *et al.* 2013) were shown to be essential for an effective piRNA-mediated transposon silencing. Some nucleases are also known to be important for piRNA pathway. During piRNA biogenesis, long primary transcripts are exported to the cytoplasm and processed to smaller fragments by an endonuclease. Zucchini or the mouse homolog MitoPLD is the nuclease that was shown to be involved in this process. Several studies in flies (Pane *et al.* 2007) and mice confirmed its importance for piRNA pathway (Watanabe *et al.* 2011). In mice, MitoPLD was shown to be important for a proper primary piRNA biogenesis and its absence results in meiotic arrest during spermatogenesis and activation of transposons (Watanabe *et al.* 2011). In flies, Zucchini was shown to be involved in the formation of both 5' and 3' ends of piRNAs during the biogenesis of the phased-piRNAs (Han *et al.* 2015, Mohn *et al.* 2015).

Contrary to the situation in flies, in mice pre-piRNAs are 30-40 nt in length and therefore their 3' end needs to be trimmed. Studies performed using BmN4, an ovarian cell line derived from the silkworm *Bombyx mori* and commonly used in piRNA studies (Kawaoka *et al.* 2009), identified the Trimmer responsible for processing the piRNAs 3' ends (Izumi *et al.* 2016). Depletion of Trimmer leads in BmN4 cells to an accumulation 35-40 nt pre-piRNAs that cannot efficiently silence their targets. In mice, PNLDC1, which is highly expressed in mouse testis, is the best candidate for piRNA 3' end processing. In fact, co-expression of PNLDC1 and TDRKH, another factor that has been shown to be important for the 3' end processing (Saxe *et al.* 2013), could efficiently process 50-nt ssRNAs loaded into MILI in HEK293T cells (Izumi *et al.* 2016). Further experiments are needed to confirm the role of PNLDC1 in 3' end processing in mice.

1.1.3.1 NEF-sp: a putative RNA exonuclease

NEF-sp is predicted to be an RNA exonuclease with 3' to 5' directionality. Its domain prediction revealed that it is composed by an N-terminal nuclease domain and two tandem RNA recognition motifs (RRMs). Analyses of the nuclease domain suggests that NEF-sp belongs to the superfamily of DEDD proteins which comprises RNases, the proofreading domains of some DNA polymerases as well as DNases (Y. H. Zuo *et al.* 2001). NEF-sp has in its nuclease domain four invariant acidic residues (D-E-D-D), distributed in three separate motifs, which are essential for nucleic acid degradation (Joyce *et al.* 1994). The presence of a fifth residue, a histidine in motif III, further classifies this protein as belonging to the DEDDh subfamily. RNA degradation is catalysed by a two-metal-ion mechanism in which the DEDDh residues are essential for coordination of the two divalent metals required for activity (Steitz *et al.* 1993). NEF-sp is also predicted to be a testis-specific nuclease, however its biological role was not described yet.

1.2 Aim of Project I

When this project was started the nuclease responsible for piRNAs 3' end trimming was still unknown. NEF-sp was predicted to be a nuclease and to be exclusively expressed in the testis. Therefore, we thought that NEF-sp could be a good candidate for having a role in piRNA maturation. The aim of this project was to investigate the functional role of NEF-sp. To do so, we aimed to:

- Characterize biochemically NEF-sp nuclease activity.
- Understand NEF-sp structural features and gain mechanistic insight into the catalytic activity.
- Determine its biological function by analysing NEF-sp knock-out mice.

PART II

Throughout my PhD I also had also the opportunity of developing a second project **"Characterization of METTL4 – a putative 6mA DNA methyltransferase"**. In the second part of the introduction I will focus in DNA methylation in mammals and its implications during development. I will explore the role 5-methylcitosine and I will also discuss the recently re-discovered N^6–methyladenine DNA mark.

2.1 DNA methylation

Cytosine nucleotides can be chemically modified by covalent attachment of a methyl group at the C5 position of the cytosine ring. This type of modification is the most well studied methylation mark in mammals. 5-methylcitosine (5mC) is largely restricted to CpG dinucleotides. Mammalian genome is particularly depleted of GC dinucleotides and its distribution is asymmetrical. Most genomic regions have a fairly low GC content and around 70-80% of these sites are methylated (Ehrlich *et al.* 1982). However, some regions of about several hundred to two thousand base pairs are highly enriched for GC dinucleotides and are generally unmethylated. These are called CpG islands (CGIs). Around 70% of CGIs are associated with promoter regions (Saxonov *et al.* 2006). Typically, 5mC DNA methylation is associated with transcriptional repression. However, the CpG ratio, the length and the GC content of CGIs can affect gene transcription depending of the presence/absence of the 5mC mark (Weber *et al.* 2007). Cytosine methylation in mammals has been shown to be essential for embryonic development, silencing of retrotransposons and establishment of genomic imprinting among others.

2.1.1 Protein machinery – DNA cytosine methyltransferases

The mammalian genome encodes for four catalytic active DNA methyltransferases (DNMTs) – DNMT1, DNMT3A, DNMT3B and DNMT3C – and a regulatory factor DNMT3L. All the four active proteins are able to methylate CG dinucleotides at the fifth carbon of cytosine residues, using AdoMet (S-adenosylmethionine) as a methyl group donner. They contain a methyltransferase catalytic domain (MTase) at the C-terminal region composed by ten motifs, six of which are quite conserved. Motifs I and X are known to bind to the AdoMet co-factor and motifs VII and IX recognize the target. The DNA methylation is catalysed by motifs IV, VI and VIII (Lauster *et al.* 1989, Posfai *et al.* 1989). Several differences between these proteins make them unique and essential for development.

DNMT1 was the first cytosine DNA methyltransferase to be identified in mammals (Bestor *et al.* 1988). DNMT1 is constitutively expressed in dividing cells (Bestor 1992) and associates with replication foci during S-phase (Leonhardt *et al.* 1992). This enzyme is responsible for the deposition of the 5mC mark in newly synthesised DNA guarantying that the DNA methylation landscapes are faithfully copied and

propagated between different cell generations. Therefore, DNMT1 is considered a maintenance methyltransferase. In mice, partial loss of DNMT1 results in embryonic lethality (Li *et al.* 1992). Homozygous embryos are not able to develop beyond E9.5 and die before reaching E11. 5mC levels were shown to be highly reduced in these mutants. Mouse embryonic stem cells lacking DNMT1 exhibit low levels of DNA methylation and proliferation after cell differentiation is compromised. *De novo* DNA methylation was not affected suggesting the existence of other DNA methyltransferase(s) (Lei *et al.* 1996).

Two other cytosine DNA methyltransferases, DNMT3A and DNMT3B, were later found in the mammalian genome (Okano *et al.* 1998). These two proteins were shown to be highly expressed in ES cells. However, their levels were low in differentiated cells and adult tissues. Recombinant DNMT3A and DNMT3B were shown to methylate CG dinucleotides without preference for hemi-methylated DNA. Studies using *Dnmt3A* and *Dnmt3B* double knock-out ES cells, demonstrated that integrating proviral DNA in these cells did not result in methylation at these sites. This result suggested that DNMT3A and DNMT3B are responsible for *de novo* DNA methylation (Okano *et al.* 1999). DNMT3A and DNMT3B were found to be expressed during early embryonic stages and to be essential for development. *Dnmt3A*$^{-/-}$ mice were normal at birth however could not develop beyond four weeks. *Dnmt3B*$^{-/-}$ mice exhibited a more striking phenotype and embryos died before E11.5. Double knock-out mice (*Dnmt3A*$^{-/-}$, *Dnmt3B*$^{-/-}$) also exhibit defects during development and morphogenesis stopped after gastrulation (Okano *et al.* 1999).

DNMT3L is another member of the DNA methyltransferase family (Aapola *et al.* 2000). This protein has a high sequence similarity with DNMT3A and DNMT3B, however the catalytic motifs essential for cytosine methylation are absent. Therefore, DNMT3L is an inactive DNA methyltransferase on its own (Hata *et al.* 2002). DNMT3L was shown to interacts with DNMT3A (Chedin *et al.* 2002) and DNMT3B (Hata *et al.* 2002, Suetake *et al.* 2004) stimulating its DNA methylation activity. *Dnmt3L*$^{-/-}$ mice are viable and do not display any developmental arrest. However, animals from both sexes are sterile (Bourc'his *et al.* 2001). Adult male knock-out mice display a total lack of germ cells. Methylation of differentially methylated regions (DMRs) was affected in testes of these mice (Kaneda *et al.* 2004). The oocytes of *Dnmt3L*$^{-/-}$ females were normal. The female sterility phenotype observed is a consequence of the absence of DNA methylation at imprinted genes. This leads to biallelic expression of genes that are not normally expressed from the maternal allele resulting in defects during development. Dnmt3a conditional knock-out exhibits similar phenotype to the DNMTL3$^{-/-}$ animals supporting the role of DNMT3L in regulating DNA methylation through interaction with DNMT3A (Kaneda *et al.* 2004).

DNMT3C was the latest cytosine methyltransferase to be discovered (Barau *et al.* 2016). Initially thought to be a duplication of *Dnmt3B* gene, it was later proven to be a functional paralog which is expressed exclusively in the male germline. This gene codifies for a protein with a domain organization similar to the other DNMT3 enzymes. DNMT3C was shown to be a *de novo* DNA methyltransferase (Barau *et al.* 2016). This enzyme is important for repression of young retrotransposon elements in the male germ line

by deposition of the 5mC methylation mark at the promoter region. Therefore, deletion of DNMT3C in mice results in male sterility. Despite of its essential role during spermatogenesis in mouse, Dnmt3C seems to be restricted only to the muroidea family not being found in humans for instance.

2.1.2 DNA methylation reprogramming during development

Cytosine methylation is a relatively static in somatic tissues and methylated CpGs are inherited during cell division. However, DNA methylation during mammalian development is a rather dynamic process. During embryonic development, epigenetic reprogramming occurs on a genome-wide scale. This involves DNA demethylation and histone remodelling at two defined phases: (i) immediately after fertilization until the preimplantation morula stage and (ii) during the establishment of primordial germ cells (PGCs) from embryonic day E10.5 to E13.5 (Monk *et al.* 1987) (Figure 1.4).

The first wave of DNA demethylation starts after fertilization in which the methylation patterns contributed by the paternal gametes are removed. This event is essential for the zygote to acquire the totipotent state necessary to develop a whole new organism. The genomic organization of the mammalian gametes at fertilization is dramatically different. Therefore, DNA demethylation is distinct and follows a different kinetics in the paternal and maternal genome. The paternal genome is actively demethylated and the process is completed before the first cell division (Oswald *et al.* 2000, Santos *et al.* 2002). 5mC demethylation occurs especially at retrotransposons (LINE-1) and genes such as *Oct4* and *Nanog*. However, some regions such as paternally imprinted genes (Olek *et al.* 1997) and IAP retrotransposons are quite resistant to this wave of demethylation (Lane *et al.* 2003). The maternal genome is passively demethylated from first cell cleavage up to the morula stage. The maintenance DNA methyltransferase is excluded from the nucleus due to the expression of an oocyte-specific form (DNMT1o) which lacks the NLS signal (Mertineit *et al.* 1998). This results in the dilution of the 5mC mark over the subsequent divisions. However, DNMT1 is transiently relocated to the nucleus during one cell cycle (at eight-cell stage). This was shown to be essential for maintenance of DNA methylation at the imprinted loci (Howell *et al.* 2001). Some studies also demonstrated that the maternal genome might also be actively demethylated (Guo *et al.* 2014).

De novo DNA methylation starts just before the blastocyst stage is reached (E3.5). This coincides with the first differentiation event which separates the embryonic lineage (inner cell mass (ICM)) from the trophectoderm (TE). DNA methylation levels are different between these two lineages: the ICM is hypermethylated whereas the 5mC levels in the trophectoderm are fairly low. These global DNA methylation differences between the embryo and placenta are maintained. The two DNA methyltransferases responsible for the *de novo* methylation events are DNMT3A and DNMT3B.

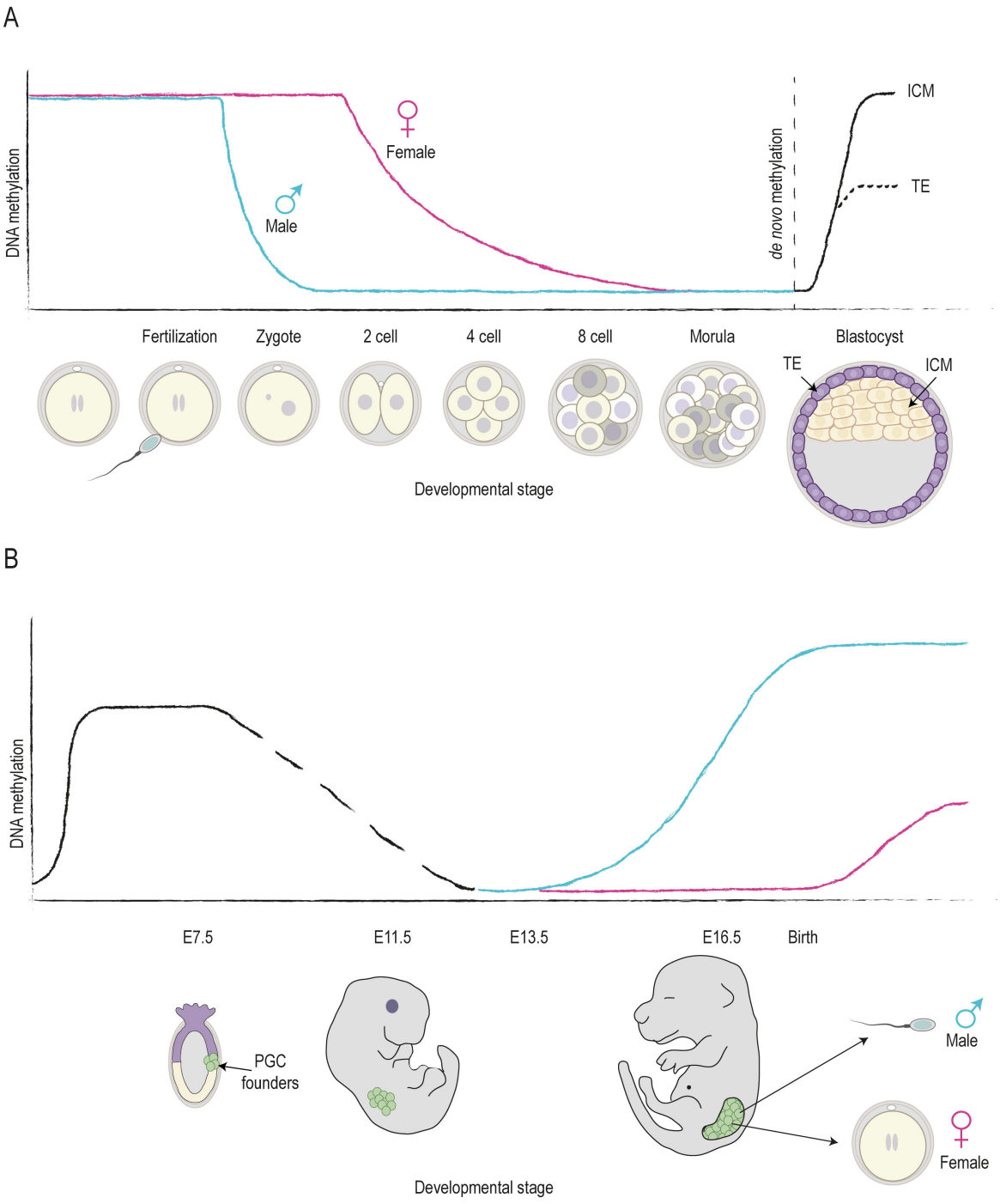


Figure 1.4 – DNA methylation reprogramming during development. (A) After fertilization, DNA methylation marks are erased from the parental gametes. Paternal genome is actively demethylated (shown in blue) followed by gradual passive loss of DNA marks from the maternal genome (shown in pink). DNA re-methylation starts once the blastocyst stage is reached. The two embryonic lineages: inner cell mass (ICM) and trophectoderm (TE) exhibit different DNA methylation levels. (B) PGCs (shown in green) suffer a second wave of DNA demethylation. DNA marks are subsequently re-established according to the gender of the new organism. Adapted from (Wu *et al.* 2010)

After blastocyst implantation (E4.5) a small subset of cells derived from the epiblast cells are selected to become the primordial germ cells (PGCs) (Ginsburg *et al.* 1990). This PGC specification occurs around the embryonic day E7.5 in response several factors such as bone morphogenetic protein 4 and 8 (BMP). These are produced by the extra-embryonic ectoderm (Lawson *et al.* 1999, Ying *et al.* 2000). BMP signalling induces the expression of two transcriptional regulators BLIMP1 (Ohinata *et al.* 2005) and PRDM14 (Yamaji *et al.* 2008). These are essential to supress the somatic programme and to re-acquire the expression of pluripotency factors. Although, epiblast cells exhibit pluripotency, they are starting to engage in a somatic cell fate. Therefore, migratory PGCs undergo a new wave of epigenetic reprogramming involving DNA demethylation, remodelling histone modifications, X-chromosome reactivation and erasure of genomic imprints. Upon arrival at the genital ridge at E11.5, the majority of the epigenetic marks are already erased (Hajkova *et al.* 2002). DNA demethylation reaches its peak around E13.5 and almost all genomic regions are highly unmethylated in both sexes. This second wave of epigenetic reprogramming is indeed more global than in the preimplantation embryos. Genomic imprints are erased and transposons demethylation is wider. Demethylation is rapidly completed and occurs in the presence of DNMT1. An active demethylation mechanism is thought to be involved (Hackett *et al.* 2013), however a combination of active and passive demethylation cannot be excluded.

After erasure of parental imprints, new imprints are established according to the gender of the new organism. Mouse female germ cells arrest at meiotic prophase I (around E13.0) and DNA methylation at imprinted regions is initiated after birth during oocyte growth. Retrotransposons are also supressed in the oocytes by DNA methylation and by small-RNA guided pathways (Watanabe *et al.* 2008). In male germ cells, DNA re-methylation starts at prospermatogonia arrested in the G1-phase of mitosis and it is completed before birth. The paternal methylation imprints are stablished in these cells from E14.5 to the new born stage (Li *et al.* 2004). Simultaneously, retrotransposon sequences are also methylated and inactivation of these elements involves the piRNA pathway (discussed in Part I). Both female and male DNA re-methylation is catalysed by DNMT3A and DNMT3B. DNMT3L is also essential for the establishment of the methylation patterns by regulating the activity of the *de novo* DNA methyltransferases.

2.1.3 The "sixth" base – N^6-methyladenine

5mC mark was thought to be the only type of DNA modification present in the mammalian genome and to have an implication during development. However, other types of DNA methylation have been emerging recently. One example is the N^6–methyladenine mark (6mA) which is being investigated for a putative implication in transcriptional control and development in mammals. This type of DNA modification has been widely studied in prokaryotes and shown to be implicated in virulence (Low *et al.* 2001), chromosome replication and DNA repair (Lu *et al.* 1994), regulation of transposable elements and transcription of certain genes. 6mA DNA methylation was also shown to be present in the genome of lower eukaryotes such as *Penicillium chrysogenum* (Rogers *et al.* 1986), *Chlamydomonas reinhardtii* (Hattman *et al.* 1978) and *Paramecium Aurelia* (Hattman 2005). However, the presence of 6mA DNA modification in the genome of higher eukaryotes has been controversial. The lack of proper techniques able to detect 6mA modification made its study challenging. Technological advancements in the past 6 years allowed drawing the first maps of 6mA distribution and prevalence. In fact, 6mA modification started to be heavily studied as a modification present in eukaryotic mRNA. Later, studies in *Chlamydomonas reinhardtii* (Fu *et al.* 2015), *Caenorhabditis elegans* (Greer *et al.* 2015) and *Drosophila melanogaster* (Zhang *et al.* 2015) confirmed the presence of 6mA mark in the genome of these organisms. This type of modification was also detected in vertebrates such as *Xenopus laevis* (Koziol *et al.* 2016) and in mouse (Wu *et al.* 2016).

2.1.3.1 *N^6-methyladenine* in eukaryotic messenger RNA

N^6–methyladenine (6mA) refers to the presence of a methyl group at the 6-nitrogen of the adenine ring. In mammals, around 0.1-0.4% of the adenine nucleotides of mRNA have the 6mA modification, corresponding to 3-5 6mA sites per mRNA. This type of modification was found to be enriched at stop codons, long exons and 3' UTR (Dominissini *et al.* 2012, Meyer *et al.* 2012). Some of the protein machinery involved in the deposition, erasure and reading of this mark has been already identified. A heterodimer of two methyltransferases METTL3-METTL14 is described as being responsible for the deposition of the 6mA in the RNA (Liu *et al.* 2014). WTAP a third protein is also present in "writers" complex and is thought to regulate the recruitment of the methyltransferase complex to mRNA targets (Ping *et al.* 2014b). The erasure of this RNA methylation mark has been attributed to two proteins: FTO and ALKBH5. FTO was shown to demethylate RNA *in vitro* and its knockdown/overexpression leads to an increase/decrease in the 6mA levels respectively (Jia *et al.* 2011). ALKBH5 was also shown to remove the 6mA mark from mRNA as well as other types of nuclear RNA. The knockdown of ALKBH5 in human cell lines leads to increase of 6mA levels and promotes RNA export to the cytoplasm (Zheng *et al.* 2013). The YTH domain-containing family was identified as a selective 6mA binding protein in mammalian cells. YTHDF2 was the first of its family to be characterized and was shown to bind 6mA RNA and promote its degradation by targeting mRNAs to the processing bodies (Wang *et al.* 2014). YTHDF1 was identified as another reader of the 6mA RNA mark and upon binding mRNA, translation

efficiency increases due to interaction with initiation factors (Wang *et al.* 2015). YTHDC3 was show to interact with YTHDF1 promoting mRNA translation and with YTHDF2 affecting mRNA decay (Shima *et al.* 2017). YTHDC1 is known to be important for mRNA splicing by recruiting the pre-mRNA splicing factor SRSF3 (Xiao *et al.* 2016). Besides, YTHDC1 also binds preferentially to 6mA residues on XIST being essential for XIST-mediated silencing (Patil *et al.* 2016). YTHDC2 is the remaining member of the YTH family which function is still a mystery.

2.1.3.2 N^6-methyladenine in *Chlamydomonas reinhardtii*

6mA DNA methylation has been previously shown to be present in the genome of *Chlamydomonas reinhardtii* (Hattman *et al.* 1978). However its genomic distribution has just been recently elucidated (Fu *et al.* 2015). High levels of 6mA modification, around 0.4 mol% of 6mA (6mA/A), were detected in *Chlamydomonas* nuclear DNA. These accumulated around the transcription start site (TSS) from -500 bp to +800 bp. However, 6mA mark was highly depleted at the TSS (0 bp position). This enrichment around the TSS seems to be predominant in actively transcribed genes. A preference for 6mA methylation at the ApT motif was also observed. The methylation mark was mainly detected in the linker DNA between adjacent nucleosomes. Therefore, 6mA might have a role in nucleosome position (Figure 1.5A). Besides 6mA DNA methylation, the 5mC mark is also present in *Chlamydomonas* genome. However, a correlation between these two DNA methylation marks was not found. These two types of modification seem to have distinct functions: 6mA mark is mainly associated with actively transcribed genes whereas 5mC is associated with transcriptional repression.

2.1.3.3 N^6-methyladenine in *Drosophila melanogaster*

The presence and functional role of 5mC DNA methylation in *Drosophila melanogaster* has always been controversial. From the three canonical DNMTs found in mammals, the fly genome encodes only for Dnmt2 (Lyko *et al.* 2011). Initial studies reported that Dnmt2 was responsible for 5mC DNA methylation (Phalke *et al.* 2009). However, *in vitro* analyses showed that Dnmt2 was responsible for methylation of tRNA (Goll *et al.* 2006). Whole genome bisulfite sequencing failed to reveal any evidence of 5mC DNA methylation in flies (Raddatz *et al.* 2013). A recent study reported the presence of 6mA DNA modification in *D. melanogaster* (Zhang *et al.* 2015). 6mA was shown to strongly accumulate in early embryos. Its levels decrease throughout development and reach the lowest level at the adult stage. This dynamic change coincides with expression of a DNA demethylase. *Drosophila* DNA 6mA demethylase (DMAD) was shown to be expressed at late embryonic stages. DMAD knock-out flies display developmental abnormalities and most of mutant animals were lethal at the pupa stage. Only a small population were able to further develop, however it died 3 days post-eclosion. Overall, the levels of 6mA modification were increased in these mutants. 6mA modification accumulated mainly in the gene body of transposable elements. This increase correlated with overexpression of transposons in the DMAD mutant ovaries (Figure 1.5B). In addition, nuclear extracts of DMAD mutant flies were shown to have 6mA

methylation activity. This suggests that a 6mA DNA methyltransferase might be present in *D. melanogaster*. CG14906 has been proposed to be the methyltransferase responsible for 6mA deposition (Heyn *et al.* 2015). However, further studies are needed to understand if CG14906 is involved in 6mA DNA methylation.

2.1.3.4 N^6-methyladenine in *Caenorhabditis elegans*

Similarly to *D. melanogaster*, 5mC DNA methylation was initially identified in *C. elegans (Klass et al. 1983)*. However, further experiments were not able to confirm the presence of this mark. Homologs of DNMT1 or DNMT3 were also not found in *C. elegans* leading to idea that DNA methylation is absent in this organism. However, 6mA DNA methylation mark was found to be present in *C. elegans* (Greer *et al.* 2015). The protein machinery involved in deposition/removal of 6mA mark was also described. N6-methyl adenine demethylase 1 (NMAD-1) was shown to be responsible for removing the 6mA methylation mark whereas DAMT-1 is the 6mA DNA methyltransferase (Figure 1.5C). A relationship between 6mA DNA modification and histone methylation was also established. In previous studies, it was shown that deletion of *spr-5* (orthologue of mammalian H3K4me2 demethylase LSD1/KDM1) leads to an increase of H3K4me2 levels and to a decrease of fertility over many generation (Katz *et al.* 2009). In the *spr-5* mutants, the levels of 6mA modification were increased in a trans-generational manner and correlate with increased infertility. Deletion of NMAD-1, and concomitant increase of 6mA levels, accelerated the infertility phenotype observed in the *spr-5* mutants. However, knockdown of *damt-1* led to a decreased in the H3K4me2 levels and infertility was also supressed in the *spr-5* mutant worms. The machinery involved in regulating 6mA mark and histone methylation seems to be linked and these two types of modification are important to transmit epigenetic information across generations.

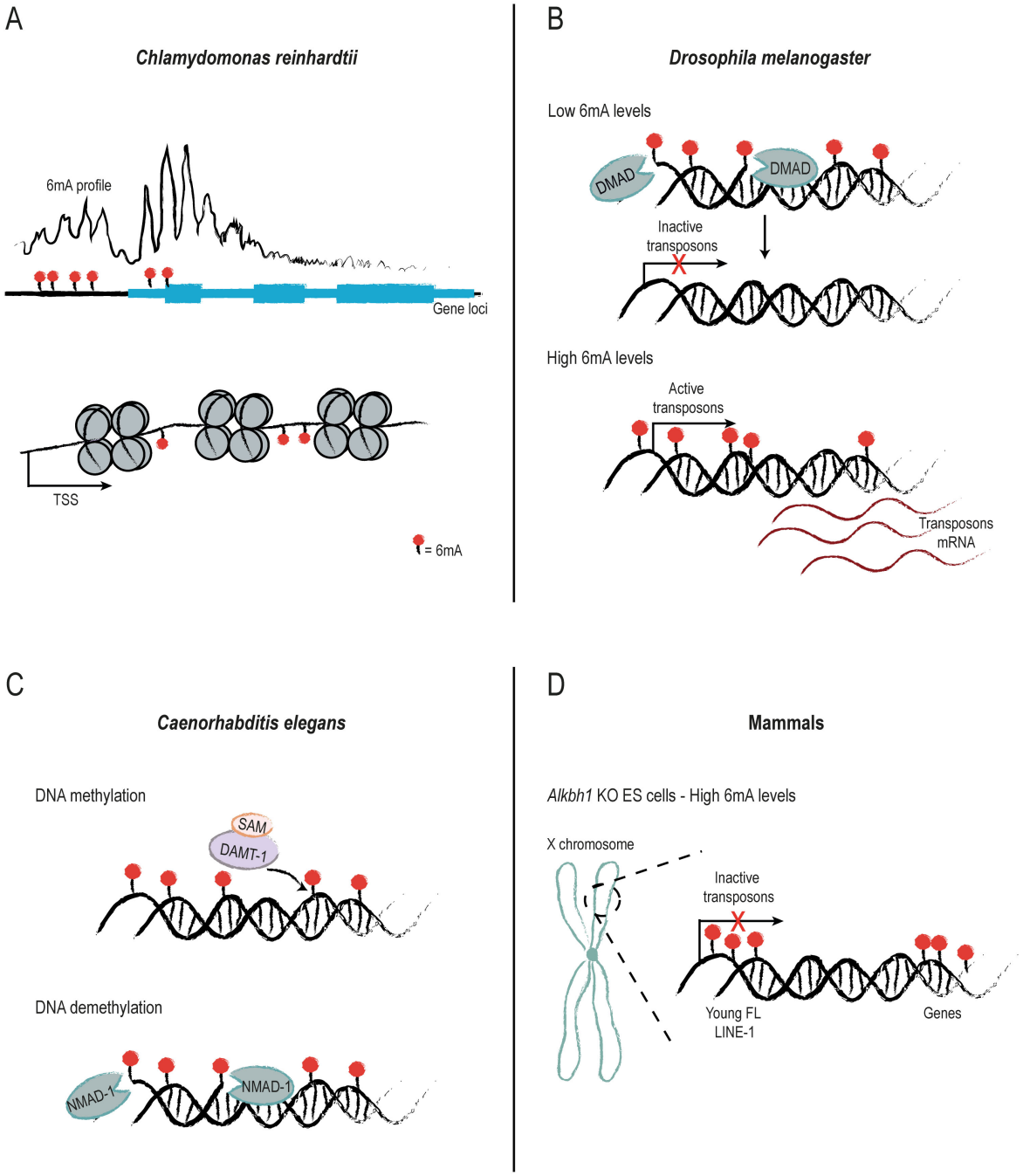


Figure 1.5 – N⁶–methyladenine DNA modification in different organisms. (A) In *Chlamydomonas reinhardtii*, 6mA is enriched upstream and downstream the TSS and marks actively transcribed genes. The methylation mark is specifically detected in the linker DNA between adjacent nucleosomes. (B) In *D. melanogaster*, 6mA modification is enriched at transposable elements and correlates with its activation. DMAD was identified as being responsible for removing the methylation mark. (C) DAMT-1 is responsible for deposition of 6mA modification in *Caenorhabditis elegans* whereas NMAD-1 removes the methylation mark. (D) In mammals, 6mA methylation is enriched at young full-length LINE-1 especially in chromosome X and its deposition correlates with inactivation of these elements.

2.1.3.5 N^6-methyladenine in mammals

Two different studies have shown that 6mA DNA methylation is present in the mouse genome (Koziol *et al.* 2016, Wu *et al.* 2016). In a study using embryonic stem cells (ES cells) the authors showed that 6mA modification is enrichment at intergenic regions (Wu *et al.* 2016). ALKBH1 was found to be the DNA demethylase involved in the removal of 6mA mark. *Alkbh1* knock-out ES cells shown higher levels of 6mA methylation and recombinant ALKBH1 removes the 6mA modification *in vitro*. RNA-seq of *Alkbh1* knock-out ES cells revealed that a substantial number of genes were downregulated indicating that 6mA DNA modification leads to gene silencing. A significant percentage of those downregulated genes were located in the X chromosome. Further analysis demonstrated that 6mA is specially enriched at the young full-length L1 elements. This accumulation correlates with transposon silencing as well as silencing of genes that are in the vicinity of these elements (Figure 1.5D). *Alkbh1* knockout ES cells did not display defects in differentiation and cell fate decision. However, a different study has shown that ALKBH1 is an RNA demethylase and it is involved in the removal of the methyl group from N^1-methyladenine (1mA) in tRNAs (Liu *et al.* 2016). Since 6mA DNA methylation was mainly study using *Alkbh1* knockout ES cells, its presence and biological relevance in mammals is not fully understood.

2.1.3.6 METTL4: a putative 6mA DNA methyltransferase

Protein machinery involved in the deposition, removal or recognition of the 6mA DNA mark is not completely characterized. Especially which protein(s) are involved in 6mA DNA deposition in mammals is unknown. Studies in *C. elegans* suggest that DAMT-1 is the 6mA DNA writer (Greer *et al.* 2015). This is the only 6mA DNA methyltransferase that has been characterized. Searches for orthologues of DAMT-1 in mammals and in flies identified METTL4 and CG14906 respectively as potential candidates. Domains predictions revealed that METTL4 has a methyltransferase domain (MTase) at C-terminus. A conserved DPPW motif characteristic of N^6-adenosine methyltransferases is also present (Iyer *et al.* 2016b). This motif is responsible for substrate recognition and catalytic activity. S-adenosylmethionine is used as the methyl donor for 6mA formation.

2.2 Aim of Project II

The aim of this second project was to investigate the functional role of METTL4. To do so, we tried to:

- Characterize biochemically METTL4 putative methyltransferase activity.
- Determine METTL4 structural features.
- Analyse its biological function in flies and mice by studying the consequences of its deletion in development and in the deposition of the 6mA mark.

2. MATERIAL AND METHODS

2.1 NEF-sp project

2.1.1 Multiple tissues reverse-transcription PCR

Total RNA from different mouse tissues and two human cells lines (HeLa and HEK293T) was extracted using TRIzol Reagent (Invitrogen cat. No. 15596026). Adult human testes total RNA was purchased (Clontech, cat no. 636533). Mouse tissues were collected and washed in ice cold 1X PBS followed by flash freezing and storage at -80 °C. The human cells lines were grown in a 6-cm dish in DMEM media. After reaching 80 % confluence, the media was removed and cells were scrapped off in 1 mL of ice cold 1X PBS. Cells were spun down at 4 °C and washed a second time in ice cold 1X PBS. The collected tissues and cell lines were lysed by adding 1 mL of TRIzol™ and homogenized with 30 strokes in a 1 mL glass dounce (homogenizer). After incubation for 5 minutes at room temperature, 0.2 mL of chloroform was added to the samples and incubated again for 5 minutes. The samples were centrifuged for 15 minutes at 12,000 × g at 4 °C and the aqueous phase containing the RNA was transferred to a new tube. RNA precipitation was done by adding 0.5 mL of isopropanol and incubating the samples for 10 minutes. After centrifugation for 10 minutes at 12,000 × g at 4 °C, the RNA pellet was washed with 75 % ethanol. The RNA was then air-dry for 5 minutes and afterwards resuspended in 20 µL DEPC water. RNA quality was assessed by 1 % agarose-formaldehyde gel electrophoresis: 1 g of agarose was dissolved in 72 mL of water by heating the suspension. After reaching 60 °C, 10 mL of 10 X MOPS running buffer (0.4 M MOPS pH 7.0, 0.1 M sodium acetate, 0.01 M EDTA) and 18 mL of 37 % formaldehyde were added to the gel. Gel was casted in appropriated cassettes. The total RNA (5 µg) mixed with formaldehyde loading buffer (1 mM EDTA pH 8.0, 0.25 % (w/v) bromophenol blue, 0.25 % (w/v) xylene cyanol, 50 % (v/v) glycerol and ethidium bromide), was loaded into the gel. Gel was ran in 1 X MOPS running buffer at 5 V/cm until the bromophenol blue reached 2/3 the length of the gel. After confirming the integrity of the isolated RNA we continued with following steps.

2 µg of total RNA was treated with DNase I (Thermo Fisher cat. No. EN0521) for one hour at 65 °C. After DNase treatment, RNA was re-extracted using phenol-chloroform extraction method followed by precipitation using 1 mL of 100 % ethanol, 1 µL of glycogen and 18 µL of 5 M NaCl. The RNA pellet was resuspended in DEPC water and quantified using the spectrophotometer (NanoDrop 1000, Thermo Scientific). First strand cDNA synthesis was done using SuperScript™ III Reverse Transcriptase (Invitrogen cat. No. 18080). 500 ng of total RNA were incubated with 1 µL of 10 mM dNTP Mix and 100 ng of random primers in a final volume of 13 µL for 5 minutes at 65 °C followed by incubation on ice for at least 1 minute. 4 µL of 5X First-Strand Buffer, 1 µL of 0.1 M DTT and 1 µL of SuperScript™ III RT were added to the samples and incubated at 25 °C for 5 minutes followed by incubation at 50 °C for 60 minutes and heat inactivation at 70 °C for 15 minutes. These cDNAs were used as a template for PCR amplification. PCR reactions were done by using 5 µL of 10 X Taq Buffer with KCl (100 mM Tris-HCl pH 8.8 at 25 °C, 500 mM KCl, 0.8% (v/v) Nonidet-P40) (Thermo Fisher cat. No. B38), 2.5 µL of 50

mM $MgCl_2$, 1 μL of 10 mM dNTPs Mix, 2.5 μL forward primer (10 μM), 2.5 μL reverse primer (10 μM), 2 μL of the obtained cDNAs and 1 μL Taq DNA polymerase in a final volume of 50 μL. PCR program consisted of an initial denaturation step at 95 °C for 5 min followed by x cycles (x=25 cycles for *Actin B*, x=35 cycles for *Mili* and *Nef-sp*) of 30 sec at 95 °C, 30 sec at 55 °C and 45 sec at 72 °C and a final step at 72 °C for 5 min. PCR products were resolved in 2 % agarose gel stained with ethidium bromide. The primers used in the PCR amplification are listed in the table 2.1.

Table 2.1 – Pair of primers used for detection of *Actin B*, *Mili* and *Nef-sp* expression in different mouse tissues and human cell lines.

PCR oligos	Sequence
mouse *Actin B*	Forward 5' – TAGGCACCAGGGTGTGATGG -3'
	Reverse 5' – CATGGCTGGGGTGTTGAAGG -3'
mouse *Mili*	Forward 5' – ATGGATCCTGTCAGGCCGTTG -3'
	Reverse 5' – CATGCCACGGAACATGGACAC -3'
mouse *Nef-sp*	Forward 5' – GATAGCAGTCCCCTCTTTGG -3'
	Reverse 5' – CAGCTTCTGTACATCTTTGAG -3'
human *Actin B*	Forward 5' – CTACAATGAGCTGCGTGTGGCTCC -3'
	Reverse 5' – CCAGGGAGGAGCTGGAAGCAGCC -3'
human *Mili*	Forward 5' – ATGGATCCTTTCCGACCATCG -3'
	Reverse 5' – GGCCTCGGAACATGGAGACC -3'
human *Nef-sp*	Forward 5' – ATGGAGCCAGAGAGGGAAGG -3'
	Reverse 5' – GCATACTTCAGCAATTCACAC -3'

2.1.2 Cloning and constructs design

The complementary DNA (cDNA) of human *NEF-sp/LOC81691* (Genbank Accession no. NM_030941) was amplified by reverse-transcription PCR (RT-PCR) using human adult testis total RNA (Clontech, cat No. 636533). The *Mus musculus* (house mouse) orthologue (Genbank Accession no. 2610020H08Rik) was similarly amplified from mouse adult testis total RNA. We identified two isoforms of the mouse *Nef-sp*: a full-length isoform composed of 784 aa (Genbank Accession no. KY853396), and a shorter C-terminal splicing variant of 696 aa (Genbank Accession no. KY853397). The sequences of the three proteins cloned are provided bellow.

Protein sequence of human NEF-sp (774 aa)

MEPEREGTERHPRKVRESRQAPNKLVGAAEAMKAGWDLEESQPEAKKARLSTILFTDNCEVTHDQLCELLKYAVLGKSNVPKPSWCQLFHQNHL
NNVVVFVLQGMSQLHFYRFYLEFGCLRKAFRHKFRLPPPSSDFLADVVGLQTEQRAGDLPKTMEGPLPSNAKAAINLQDDPIIQKYGSKKVGLT
RCLLTKEEMRTFHFPLQGFPDCENFLLTKCNGSIADNSPLFGLDCEMCLTSKGRELTRISLVAEGGCCVMDELVKPENKILDYLTSFSGITKKI
LNPVTTKLKDVQRQLKALLPPDAVLVGHSLDLDLRALKMIHPYVIDTSLLYVREQGRRFKLKFLAKVILGKDIQCPDRLGHDATEDARTILELA
RYFLKHGPKKIAELNLEALANHQEIQAAGQEPKNTAEVLQHPNTSVLECLDSVGQKLLFLTRETDAGELPSSRNCQTIKCLSNKEVLEQARVEI
PLFPFSIVQFSFKAFSPVLTEEMNKRMRIKWTEISTVYAGPFSKNCNLRALKRLFKSFGPVQSMTFVLETRQPHLCIQYEVLEAAQLAIESLDG
ILVDGICIKVQRPVTELTLDCDTLVNELEGDSENQGSIYLSGVSETFKEQLLQEPRLFLGLEAVILPKDLKSGKQKKYCFLKFKSFGSAQQALN
ILTGKDWKLKGRHALTPRHLHAWLRGLPPESTRLPGLRVVPPPFEQEALQTLKLDHPKIAAWRWSRKIGKLYNSLCPGTLCLILLPGTKSTHGS
LSGLGLMGIKEEEESAGPGLCS

Protein sequence of mouse NEF-sp long isoform (784 aa)

MEQEKEEANSRKRKEAPNSSASATERDGPPSHIQDREPQAKKARLSTILFTDNCEVTHRQLCELLKYAVLGKSAVPKPSWCQLSHQKQLNNVVV
FILKGMSQLHFYRFYLEFRFLRKTFRHKFSLPSPTSSFLFDIIGLQKKKSARSFRRTVEGPLISATLRSSIDLQNDPIIQKYGYKNVGLTKCLL
TKEEMKTFHFPLPGSPNYENFILTKYTGFITDSSPLFGLDCEVCLTSMGKELTRISLVTEGGYCLIDELVKPDLKILDYLTSFTGITKEILNPV
TTKLKDVQKLLRELLPPDAVLVGHCLDLDLRVLKIAEFNLEALAANQEQGDKEEAAHMSVLKCLESMGQKLLFLTQDINELSSYRNCQTIKCSS
NKKVLEQARVEVPLFPFNIVEFSFQPFPPLFAEEMKNSMKTKWTQMSTVYAGPFSKDCNVGALKKVFSSLGPVHSITLVLETYRPYFSIQYEVL
EAAQLALETMNGSLLEGSCIRVRRLVTELTLECDTLVRELEQDSENQGTIYVAGIGETFKEHLLEQSNLFPDLEAVILPKEVKSRKQKNYCFLK
FKTVNSAQVALEILKGKDWKLKGRNALTPRHLQAWLKDIHPEPAMPMGLRIVPPLMERHIFRTRKVNHPKIVAWRWSRKIEKLYHSLSPGTFCL
ILMPGTKNAFGSHPGLGLMKIKEEEESDTPGVRVSAGRHTFLTTQNGNFGSKANEKGRKRSISTYDMKIQKHQRKLSKKSTSSKTEDSLLLNPG
HLHWYHILEHQLYFTNRLVLTVSGVSTKLQVP

Protein sequence of mouse NEF-sp short isoform (696 aa)

MEQEKEEANSRKRKEAPNSSASATERDGPPSHIQDREPQAKKARLSTILFTDNCEVTHRQLCELLKYAVLGKSAVPKPSWCQLSHQKQLNNVVV
FILKGMSQLHFYRFYLEFRFLRKTFRHKFSLPSPTSSFLFDIIGLQKKKSARSFRRTVEGPLISATLRSSIDLQNDPIIQKYGYKNVGLTKCLL
TKEEMKTFHFPLPGSPNYENFILTKYTGFITDSSPLFGLDCEVCLTSMGKELTRISLVTEGGYCLIDELVKPDLKILDYLTSFTGITKEILNPV
TTKLKDVQKLLRELLPPDAVLVGHCLDLDLRVLKIAEFNLEALAANQEQGDKEEAAHMSVLKCLESMGQKLLFLTQDINELSSYRNCQTIKCSS
NKKVLEQARVEVPLFPFNIVEFSFQPFPPLFAEEMKNSMKTKWTQMSTVYAGPFSKDCNVGALKKVFSSLGPVHSITLVLETYRPYFSIQYEVL
EAAQLALETMNGSLLEGSCIRVRRLVTELTLECDTLVRELEQDSENQGTIYVAGIGETFKEHLLEQSNLFPDLEAVILPKEVKSRKQKNYCFLK
FKTVNSAQVALEILKGKDWKLKGRNALTPRHLQAWLKDIHPEPAMPMGLRIVPPLMERHIFRTRKVNHPKIVAWRWSRKIEKLYHSLSPGTFCL
ILMPGTKNAFGSHPGLGLMKIKEEEESDTPGVRVTPCC

In order to study NEF-sp in more detail, we cloned human and mouse NEF-sp in appropriated vectors for expression in different systems. The human and mouse NEF-sp were cloned into the SalI and NotI sites of the vector pCI-neo-N-HA (Pillai *et al.* 2005) suitable for mammalian expression. Mouse NEF-sp antigens used for antibody generation were expressed in bacteria (table 2.2). Antigens were cloned into the bacterial expression vector pETM-11 (6xHis-TEV-fusion; EMBL Protein Expression and Purification Core Facility) in the restriction sites NcoI and NotI. For the production of recombinant soluble proteins used in the nuclease assays and crystallization trials, different clones were made for bacteria and insect cells expression. For bacteria expression, the pETM-11-SUMO vector (EMBL Protein Expression and Purification Core Facility) was used to insert different constructs with the NcoI and NotI restriction sites. Protein expression with a C-terminal tag was done by cloning appropriate constructs into the XbaI and XhoI restriction sites of pETM-30 vector (EMBL Protein Expression and Purification Core Facility). Eukaryotic expression using insect cells was performed by cloning the constructs into the NheI and KpnI sites of the vector pACEBac2-SUMO (6xHis-Strep-SUMO-TEV tag). The constructs used in this study are listed in the table 2.2.

Table 2.2 – List of constructs used in NEF-sp study. Organism, construct boundaries, expression system, vector used for cloning, tag and construct name are indicated.

Organism	Protein	Expression host	Plasmid	Tag	Construct name
Homo sapiens	NEF-sp (1-774 aa)	Mammalian	pCI-neo-N-HA	N-HA	hNEF-sp
	NEF-sp (58-774 aa)				hNEF-sp$^{\Delta NLS}$
Mus musculus	NEF-sp (1-784 aa)				mNEF-sp
	NEF-sp (53-784 aa)				mNEF-sp$^{\Delta NLS}$
	NEF-sp (1-696 aa)				mNEF short isoform
Mus musculus	NEF-sp (1-200 aa)	*E. coli*	pETM-11	6xHis	Antigen 1
	NEF-sp (51-249 aa)				Antigen 2
	NEF-sp (604-784 aa)				Antigen 3
	NEF-sp (87-222 aa)				Antigen 4
	NEF-sp (132-258 aa)				Antigen 5
Homo sapiens	NEF-sp (1-774 aa)		pETM-11-SUMO	6xHis-SUMO-TEV	hNEF-sp
	NEF-sp (42-774 aa)				hNEF deletion 1
	NEF-sp (372-658 aa)				hNEF-A
	NEF-sp (147-724 aa)				hNEF-B
	NEF-sp (167-409 aa)				hNEF-C
	NEF-sp (190-576 aa)				hNEF-D
	NEF-sp (147-724 aa)			TEV-SUMO-6xHis (C-terminal tag)	hNEF-B2
	NEF-sp (1-774 aa)	Insect cells	pACEBac2-SUMO	6xHis-Strep-SUMO-TEV	hNEF-sp
	NEF-sp (42-774 aa)				hNEF deletion 1

2.1.3 Recombinant protein production and purification

Initial trials to obtain recombinant human NEF were performed using *E. coli* (BL21 strain). Several human NEF-sp constructs were cloned into pETM-11-SUMO (table 2.2). In order to express these constructs, 1 μL of each plasmid was transformed in *E. coli* and cultures were grown at 37 °C under selective resistance (kanamycin and chloramphenicol selection). Once the log phase (OD of 0.6-0.8) was reached, NEF-sp expression was induced by adding IPTG at a final concentration of 0.6 mM. After induction, cultures were grown overnight at 18 °C.

For production of recombinant proteins using the eukaryotic expression systems we used the following ovary-derived insect cell lines: Sf21 from the Fall Army worm *Spodoptera frugiperda* or High Five (Hi5) from the cabbage looper, *Trichoplusia ni*. NEF-sp constructs were cloned in pACEBac2-SUMO vector (table 2.2). Bacmid preparation for Sf21 transfection was done using electro-competent DH10*Multi*Bac*Turbo* cells as described previously (Bieniossek *et al.* 2012). 48 hours post transfection V0 virus was collected and 3.0 mL were added into 25 mL Sf21 ($0.5x10^6$ cells/mL) to produce V1. 48 hours after proliferation arrest, V1 was collected and used for large scale expression. Cells were harvest 72 hours post proliferation arrest.

All the human NEF-sp constructs expressed in bacteria and insect cells have an initial common purification step. Lysis of these constructs was made using the corresponding lysis buffer (table 2.3) supplemented with protease inhibitor (Roche; Complete EDTA-free). Cells were sonicated for 2 minutes (5 sec on; 15 sec off). The lysate was cleared by spinning down at 18000 x g for 45 min at 4 °C. The supernatant was collected and incubated with Ni-NTA beads. After incubation for 45 min at 4 °C in a rotating wheel, several wash steps were performed. Five step washes were performed using 10 mL of the following buffers (table 2.3) on a gravity flow column: wash buffer I, wash buffer II, wash buffer III followed by final washes using wash buffer II and wash buffer I. Elution of the Ni^{2+}-bound proteins was done by adding elution buffer containing 250 mM of Imidazole.

Table 2.3 – List of buffers used for purification of human NEF-sp constructs. Buffer name and their composition are indicated.

Buffer	Composition
Lysis Buffer	25 mM Tris pH 7.5, 500 mM NaCl, 5 mM of β-mercaptoethanol, 0.5 % Tween, 20 mM Imidazole, 5 mM $MgCl_2$, 10 % of glycerol
Wash Buffer I	25 mM Tris pH 7.5, 500 mM NaCl, 5 mM of β-mercaptoethanol, 0.1 % Tween, 20 mM Imidazole, 5 mM $MgCl_2$ and 10 % of glycerol
Wash Buffer II	25 mM Tris pH 7.5, 1.5 M NaCl, 5 mM of β-mercaptoethanol, 0.5 % Tween, 20 mM Imidazole, 5 mM $MgCl_2$ and 10 % of glycerol
Wash Buffer III	25 mM Tris pH 7.5, 500 mM NaCl, 5 mM of β-mercaptoethanol, 0.1 % Tween, 50 mM Imidazole, 5 mM $MgCl_2$ and 10 % of glycerol
Elution Buffer	25 mM Tris pH 7.5, 300 mM NaCl, 5 mM of β-mercaptoethanol, 0.1 % Tween, 250 mM Imidazole and 10 % of glycerol

Purity of proteins eluted from the Ni^{2+}-affinity chromatography were analysed by SDS-PAGE stained with Instant Blue (Expedeon cat. No. ISB1L). Most the constructs expressed in bacteria were either insoluble or poorly expressed (table 2.4). Others such as hNEF-B1, hNEF-B2 and hNEF-C were not suitable for our experiment purpose, either (discussed in section 3.2.4). Therefore, none of bacterial expressed constructs were further purified.

Table 2.4 – List of human NEF-sp constructs expressed in bacteria and insect cells.

Protein	Expression host	Tag	Construct name	Solubility
NEF-sp (1-774 aa)	*E. coli*	6xHis-SUMO-TEV	hNEF-FL	Insoluble/poorly expressed
NEF-sp (42-774 aa)			hNEF deletion 1	Insoluble/poorly expressed
NEF-sp (372-658 aa)			hNEF-A	Insoluble
NEF-sp (147-724 aa)			hNEF-B	Soluble
NEF-sp (167-409 aa)			hNEF-C	Soluble
NEF-sp (190-576 aa)			hNEF-D	Insoluble/poorly expressed
NEF-sp (147-724 aa)		TEV-SUMO-6xHis (C-terminal tag)	hNEF-B2	Soluble
NEF-sp (1-774 aa)	Insect cells	6xHis-SUMO-TEV	hNEF-FL	Soluble
NEF-sp (42-774 aa)			hNEF deletion 1	Soluble

Only hNEF-FL and hNEF deletion 1 expressed in insect cells were further purified. After Ni^{2+}-affinity purification, hNEF-FL was loaded into a Heparin column (HI-Trap; GE Healthcare). The purest fractions were concentrated and loaded into gel filtration column for the last polishing step (Superdex 200 10/300GL; GE healthcare) in the buffer 25 mM Tris pH 7.5, 500 mM NaCl, 5 mM of β-mercaptoethanol, 0.5 % Tween and 10 % of glycerol). Mono-dispersed fractions were concentrated to a concentration of 2 $mg.mL^{-1}$ and used in nuclease assays without removal of the 6xHis-Strep-SUMO tags.

2.1.4 Human NEF-sp crystallization trials

hNEF-sp deletion 1 expressed in insect cells was used for structural studies. After Nickel-NTA affinity purification, hNEF-sp deletion 1 was purified on a Heparin column (HI-Trap; GE Healthcare). The purest fractions were pulled together and the 6xHis-Strep-SUMO fusion tag was removed. The construct has a TEV site downstream the SUMO domain that allows TEV binding and cleavage. TEV cleavage was performed overnight at 4 °C in 2 L of dialysis buffer (25mM Tris pH 7.5, 300 mM NaCl, 5 mM of β-mercaptoethanol and 10 % of glycerol). After tag removal, hNEF deletion 1 was loaded into gel filtration chromatography column (Superdex 200 10/300GL; GE healthcare) and mono-dispersed fractions were concentrated. Crystallization trials were done using different hNEF deletion 1 concentrations (described in section 3.2.4). Initially, size exclusion chromatography of hNEF deletion 1 was done using 25 mM Tris pH 7.5, 500 mM NaCl, 5 mM of β-mercaptoethanol an 10 % of glycerol. However, we were not able to obtain hNEF-sp crystals in those conditions. We screened 96 different conditions (table 2.5) in order to find a suitable buffer. Buffer optimization was performed by a Thermo Fluor assay (provided by the HTX lab at EMBL Grenoble). This temperature-based assay assesses the thermal stability of a designated protein. A small amount of protein is incubated with SYPRO Orange dye, as temperature increases protein unfolds and the dye binds to hydrophobic patches/denatured protein. The increase in fluorescence is evaluated and the melting temperature can be calculated. The highest melting temperature of hNEF-sp deletion 1 was obtained in 100 mM Citric Acid pH 6.0 and 250 mM NaCl (more details in the section 3.2.4). New crystallization trials were done using hNEF deletion 1 purified in 25 mM Citric Acid pH 6.0, 250 mM NaCl, 5 mM β-mercaptoethanol and 10 % glycerol.

Table 2.5 – Composition of the 96-well buffer optimization screen.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Water	Citric Acid pH 4.0	Na Acetate pH 4.5	Citric Acid pH 5.0	MES pH 6.0	K Phosphate pH 6.0	Citric Acid pH 6.0	Bis-Tris pH 6.5	Na Cacodylate pH 6.5	Na Phosphate pH 7.0	K Phosphate pH 7.0	HEPES pH 7.0
B	MOPS pH 7.0	Am Acetate pH 7.3	Tris-HCl pH 7.5	Na Phosphate 7,5	Imidazol pH 8.0	HEPES ph 8.0	Tris-HCl pH 8.0	Tricine pH 8.0	Bicine pH 8.0	Bicine pH 8.5	Tris-HCl pH 8.5	Bicine pH 9.0
C	Water 250mM NaCl	Citric Acid 250mM NaCl pH 4.0	Na Acetate 250mM NaCl pH 4.5	Citric Acid 250mM NaCl pH 5.0	MES 250mM NaCl pH 6.0	K Phosphate 250mM NaCl pH 6.0	Citric Acid 250mM NaCl pH 6.0	Bis-Tris 250mM NaCl pH 6.5	Na Cacodylate 250mM NaCl pH 6.5	Na Phosphate 250mM NaCl pH 7.0	K Phosphate 250mM NaCl pH 7.0	HEPES 250mM NaCl pH 7.0
D	MOPS 250mM NaCl pH 7.0	Am Acetate 250mM NaCl pH 7.3	Tris-HCl 250mM NaCl pH 7.5	Na Phosphate 250mM NaCl 7,6	Imidazol 250mM NaCl pH 8.0	HEPES 250mM NaCl ph 8.0	Tris-HCl 250mM NaCl pH 8.0	Tricine 250mM NaCl pH 8.0	Bicine 250mM NaCl pH 8.0	Bicine 250mM NaCl pH 8.5	Tris-HCl 250mM NaCl pH 8.5	Bicine 250mM NaCl pH 9.0
E	Buffer Screen "A" pH 4.0	Buffer Screen "A" pH 4.78	Buffer Screen "A" pH 5.21	Buffer Screen "A" pH 5.62	Buffer Screen "A" pH 5.95	Buffer Screen "A" pH 6.23	Buffer Screen "A" pH 6.53	Buffer Screen "A" pH 6.81	Buffer Screen "A" pH 7.16	Buffer Screen "A" pH 7.80	Buffer Screen "A" pH 9.0	Buffer Screen "A" pH 10
F	Buffer Screen "B" pH 4.0	Buffer Screen "B" pH 4.22	Buffer Screen "B" pH 4.62	Buffer Screen "B" pH 5.06	Buffer Screen "B" pH 5.62	Buffer Screen "B" pH 6.49	Buffer Screen "B" pH 7.25	Buffer Screen "B" pH 7.76	Buffer Screen "B" pH 8.20	Buffer Screen "B" pH 8.75	Buffer Screen "B" pH 9.0	Buffer Screen "B" pH 10
G	10mM HEPES pH 7.5	50mM HEPES pH 7.5	100mM HEPES pH 7.5	250mM HEPES pH 7.5	10mM NaPO4 pH 7.5	50mM NaPO4 pH 7.5	100mM NaPO4 pH 7.5	200mM NaPO4 pH 7.5	10mM Tris-HCl pH 8.0	50mM Tris-HCl pH 8.0	100mM Tris-HCl pH 8.0	250mM Tris-HCl pH 8.0
H	50mM HEPES 50mM NaCl pH 7.5	50mM HEPES 125mM NaCl pH 7.5	50mM HEPES 250mM NaCl pH 7.5	50mM HEPES 500mM NaCl pH 7.5	50mM HEPES 750mM NaCl pH 7.5	50mM HEPES 1000mM NaCl pH 7.5	50mM Tris-HCl 50mM NaCl pH 8.0	50mM Tris-HCl 125mM NaCl pH 8.0	50mM Tris-HCl 250mM NaCL pH 8.0	50mM Tris-HCl 500mM NaCl pH 8.0	50mM Tris-HCl 750mM NaCl pH 8.0	50mM Tris-HCl 1000mM NaCl pH 8.0

Buffers from lane A, B, C, D, E, F are at 100mM

Buffer screens: A) Succinic Acid / Na PO4 / Glycine [2:7:7], B) Citric acid / CHES / HEPES [2:4:3]

2.1.5 Nuclease assay

ssRNA (Microsynh) and ssDNA substrates (Invitrogen) listed in table 2.6 were labelled at the 5' end with [γ-^{32}P] ATP and T4 polynucleotide kinase (Thermo Scientific cat. No. EK0031). The labelled probes were run on 15% denaturing urea-polyacrylamide gels. The band corresponding to the size of labelled nucleic acid was cut out from the gel and eluted in 0.3 M NaCl at 4 °C overnight. RNA/DNA substrates were extracted by phenol-chloroform extraction and precipitated using 100 % ethanol. The pellet was resuspended in 20 µL of RNase free water and 1 µL was used in each reaction.

RNA and DNA duplexes were generated as follow: 100 µM of the top strand (ssRNA4 or ssDNA1) was labelled at the 5' end with [γ-^{32}P] ATP and resolved in 20 % denaturing urea-polyacrylamide gels. The full-length labelled oligonucleotides were eluted in 600 µL of elution buffer (300 mM NaOac, 1 mM EDTA and 0.5 % SDS) overnight at 4 °C followed by precipitation using 1.8 mL of 100 % ethanol and 2 µL of glycogen. The pellets were resuspended in 16 µL of water and annealing reactions were performed using 2 µL of 100 µM unlabelled complementary strand and 2 µL of 10 x duplex annealing buffer (100 mM MOPS pH 6.5, 10 mM EDTA and 0.5 M KCl) followed by heating the samples to 95 °C and gradually cooled down to room temperature. The oligonucleotide duplexes were resolved on 15 % non-denaturing PAGE followed by gel elution and precipitation as described above.

For the nuclease reaction 1 µL of RNA substrate was mixed with 1 µM of purified recombinant human NEF-sp proteins in the assay buffer 25 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM DTT supplemented with 25 mM $MgCl_2$ or 2.5 mM $MnCl_2$ and 150 mM EDTA if required. The reaction was incubated at 37 °C for 1 hour. Protease K treatment was used to terminate the reaction followed by phenol-chloroform

RNA extraction and precipitation with ethanol. The samples were resolved by electrophoresis on 15% denaturing polyacrylamide gels and exposed overnight to Phosphor Storage screen (GE Healthcare). The screen was scanned using Typhon Scanner (GE Healthcare Life Sciences). The ssRNA1 was used for the nuclease assays in Figure 3.4C-F and Figure 3.5A. For Figure 3.5B we used ssRNA1, ssRNA2 and ssRNA3. The nuclease assay with 2-O-Met modification at 3' end was done using ssRNA2 (without modification) and ssRNA 2-O-Met_1 (Figure 3.5E). dsRNA probe was obtained by annealing ssRNA4 with ssRNA5 (Figure 3.5C) and dsDNA by annealing ssDNA1 with ssDNA2 (Figure 3.5D). The RNA and DNA oligonucleotides used in the nuclease assays are listed in table 2.6.

Table 2.6 – RNA and DNA oligonucleotides used for the characterization of human NEF-sp nuclease activity

RNA oligos	Sequence
ssRNA1	5'–GGCGAGAAAGCUAUCUGAGCACCUGUGUUCAUGUCAGCAU–3' (40nt)
ssRNA2	5'–UGACAUGAACACAGGUGCUCA–3' (21nt)
ssRNA3	5'–UGACAUGAAC–3' (10nt)
ssRNA4	5'– UGACAUGAACACAGGUGCUC–3' (20nt)
ssRNA5	5'–GAGCACCUGUGUUCAUGUCA–3' (20nt)
ssRNA 2-O-Met1	5'–UGACAUGAACACAGGUGCUCA–3' (21nt) same sequence as ssRNA2
ssDNA1	5'–CAGGTTTGACCAGTTCTGCCATAACACAGCAGC–3' (33nt)
ssDNA2	5'–GCTGCTGTGTTATGGCAGAACTGGTCAAACCTG–3' (33nt)

2.1.6 Mammalian cell culture and immunofluorescence

Human embryonic kidney (HEK) 293T, human epithelioid cervix carcinoma (HeLa) and mouse NIH/3T3 cell lines were maintained in DMEM media (Invitrogen catalogue no. 21969-035) supplemented with 10 % fetal bovine serum (Invitrogen; cat. No. 16000044), 1 % Penicillin: Streptomycin (Invitrogen; cat. No. 15140122) and 1 % of L-Glutamine (Invitrogen; cat. No. 25030123) referred as DMEM (+/+). The adherent cells were grown at 37 °C in temperature-controlled conditions with 5 % of CO_2. Confluent 75 cm^2 flasks were rapidly washed with 1X PBS at 37 °C and incubated with 1 mL trypsin (Life technology, cat. No. 25300-054) for 1 min allowing the cells to detach. The cells were resuspended in 10 mL of DMEM (+/+) media and 1 mL of the suspension was distributed in a 6-cm dish in a final volume of 4 mL. Cells were grown overnight prior transfection. For total cell extracts, HEK293T cells cultured in a 6-cm dish were quickly washed in DMEM serum-free media (-/-) and added 1.5 mL of DMEM (-/-). A mixture of 5 µg plasmid DNA and 15 µL X-tremeGENE™ HP (DNA Transfection Reagent, Roche; 06366236001) was prepared in 500 µL of DMEM (-/-) and incubated for 25 min. The transfection mixture was added to the 6-cm dishes and incubated at 37 °C for 3-5 h. Cells were supplemented with 2 mL of DMEM 2X (+/+) media and incubated at 37 °C for 24-48 h. Cells were collect in 1X PBS and spun down by centrifugation (800 x g). Lysis was done in 50 mM Tris-HCl pH 8.0,

150 mM NaCl, 5 mM $MgCl_2$, 1 % Triton X-100, 0.5 % Sodium deoxycholate, 1 mM DTT, 5 % glycerol and 1 x protease inhibitor followed by sonication 30 sec on-pulse.

For immunofluorescence analysis, HeLa and NIH/3T3 cell lines were grown on coverslips in 12-well plates and transfected with 1 μg of expression vectors following similar protocol used for HEK293T transfection. After 48 h transfection, cells were washed with 1X PBS and fixed using 4 % paraformaldehyde for 10 min. All the staining protocol was performed at room temperature. After two washes in 1X PBS, cells were incubated with 0.1 M Glycine for 4 min. Blocking was done in 0.2 % BSA, 1 % goat serum and 0.1 % Triton X-100 in 1X PBS for 5 min. Protein detection was done using primary anti-HA antibody or home-made antibodies which recognize the target proteins. Antibodies were diluted in blocking buffer in 1:200 (antibody:blocking buffer) ratio. Incubation of coverslips with the primary antibody was done for 90 min. The anti-mouse-IgG coupled Alexa Fluor 488 (Invitrogen) was used as secondary antibody. Secondary antibody was diluted in blocking buffer at 1:1000 ratio and incubated for 60 min. Afterwards, the coverslips were incubated with DAPI diluted in 1X PBS for 5 min in the dark. After three washes with 1X PBS the coverslips were mounted using ProLong® Gold Antifade Mountant with DAPI (Life Technologies cat. No. P-36931) on a glass slide. Cells were examined using an inverted confocal microscope Leica TCS SP2 AOBS.

2.1.7 Antigens and antibody purification

Five mouse NEF-sp antigens were produced in *E. coli* (*Rosetta* strain) (table 2.7). *E. coli* was transformed with the different constructs cloned in suitable vectors for bacterial expression containing an N-terminal 6xHis tag. Cultures were grown at 37 °C. When the log phase was reached, antigen expression was induced by adding IPTG at a final concentration of 0.6 mM. Cultures continued to grow overnight at 37 °C. To determine the antigen solubility 1 mL of culture was collected before and after induction. The bacteria pellets were resuspended in 1 mL of 1X PBS and lysed by sonication for 30 sec. After spinning down, the pellet fraction was resuspended in 8 M Urea. Both pellet and supernatant fractions were loaded in a SDS-PAGE gel stained with Instant Blue in order to determine antigen solubility. All the expressed antigens were insoluble.

Table 2.7 – List of antigens expressed in bacteria. Construct boundaries, expression system, tag, construct name and solubility are indicated.

Protein	Expression host	Tag	Construct name	Solubility
NEF-sp (1-200 aa)	*E. coli*	6xHis-TEV	Antigen 1	Insoluble
NEF-sp (51-249 aa)			Antigen 2	
NEF-sp (604-784 aa)			Antigen 3	
NEF-sp (87-222 aa)			Antigen 4	Not expressed
NEF-sp (132-258 aa)			Antigen 5	Insoluble

We purified 5 antigens from the inclusion bodies. 1 L of cell culture was spun down at 5000 rpm for 15 min at 4 °C. The pellet was resuspended in 10 mL of 1X PBS supplemented with 1 mM of PMSF and sonicated for 2 min: 5 sec on, 15 sec off. The lysate was spun down at 15000 x g for 40 min at 4 °C. The pellet was resuspended in 10 mL of Resuspension Buffer II (0.1 M Tris-HCl pH 7 and 20 mM EDTA). A new round of centrifugation was done at 15000 x g for 20 min at 4 °C followed by resuspension of the pellet in 10 mL of High Salt Buffer (60 mM EDTA and 1.5 M NaCl) with 6 % Triton X-100. The pellet was again collected by centrifuging at 15000 x g for 20 min at 4 °C and resuspended in 10 mL of High Salt Buffer. The suspension was pelleted again by centrifugation at 15000 x g for 20 min at 4 °C and the pellets were resuspended in 5 mL of 8 M urea and 50 mM Tris-HCl pH 8.0 to extract protein (antigen) from inclusion bodies. The antigens were further purified using Nickel-NTA affinity purification and eluted in 50 mM Tris-HCl pH 8.0, 8 M urea and 250 mM Imidazole. Chosen antigens were immunized in a rabbit (EMBL Heidelberg Animal Facility) and several bleeds containing the polyclonal antibodies were obtained (see section 3.2.6).

Specific antibodies were affinity-purified using a matrix prepared with the antigens. Purified antigens in buffer (50 mM Tris-HCl, 8 M urea) were loaded into a large well (remove spacers between individual wells) of a 10 % SDS-PAGE gel. Proteins were transferred to a nitrocellulose membrane using transfer buffer (25 mM Tris, 190 mM glycine and 20 % methanol) overnight at 5 mA. The membrane was then stained with Ponceau S and the strip corresponding to the antigen was cut out. After washing in wash buffer (20 mM Tris Base, 137 mM NaCl, pH 7.6, 5 % dry milk powder) for 1 hour at room temperature, the membrane was incubated overnight at 4°C in 10 mL of wash buffer containing 1 mL of antisera. The strip was washed 3 times with 15 mL of wash buffer. The washed strip was then placed in a 50 mL Falcon tube with the protein side up and bound antibodies were eluted with 500 µL of low pH elution buffer (0.1 M glycine pH 2.8, 150 mM NaCl) by passing the solution several times over the strip (a total of 1 min incubation time). The collected solution is then immediately neutralized with a pre-determined amount of 0.5 M Tris-HCl pH 8.0, 150 mM NaCl. Elution step was repeated twice and both elution fractions were pooled together and concentrated using Amicon ultra Centrifugal filters (0.5 mL, ultracel with 10 kDa cut-off; Reference no. UFC501096) to a maximum of 1 $mg.mL^{-1}$ concentration.

Other antibodies used in this study:

Mouse α-HA monoclonal was a kind gift from Marc Buhler, FMI, Basel, Switzerland. α-His (Amersham; 27-4710-01), and α-HA affinity matrix (Roche; cat. no 11815016001) for immunoprecipitations were commercially purchased. For immunofluorescence studies the following secondary antibodies were used: anti-mouse-IgG coupled to Alexa Fluor 488 (Invitrogen). For western blot analysis the following secondary antibodies conjugated to Horse Radish Peroxidase were used: anti-rabbit IgG HRP-linked antibody (GE Healthcare; NA934), anti-mouse IgG HRP-linked (GE Healthcare; NA931) and anti-rat IgG HRP-linked (GeneScript; A00167)

2.1.8 Generation of *Nef-sp* mouse mutant

The coding sequence in *Nef-sp* locus of the mouse genome was disrupted using the RNA-guided Cas9 endonuclease [Cas9 nickase (Cas9n)] and a pair of guide RNAs (gRNAs) (Mouse Biology Program, University of California, Davis). The *Cas9n* mRNA and the *in vitro* transcribed gRNAs (gRNA L1 and R4) were used for injection into C57B1/6J host embryos for mouse generation using one-cell stage injection. Founder animals were genotyped by PCR from tail genomic DNA and sequenced to identify modifications in the *Nef-sp* locus (Figure 2.1). From the different N*ef-sp* founders identified, we decided to retain only one of the allele which has a 23bp deletion. The founder female was crossed with wildtype C57Bl/6J Rj (Janvier labs, France) male. The resulting F1 animals were genotyped to confirm germline transmission. The 23bp deletion generates a frameshift mutation resulting in a premature stop codon upstream the nuclease domain. This will be referred as *Nef-sp* knock-out mouse.

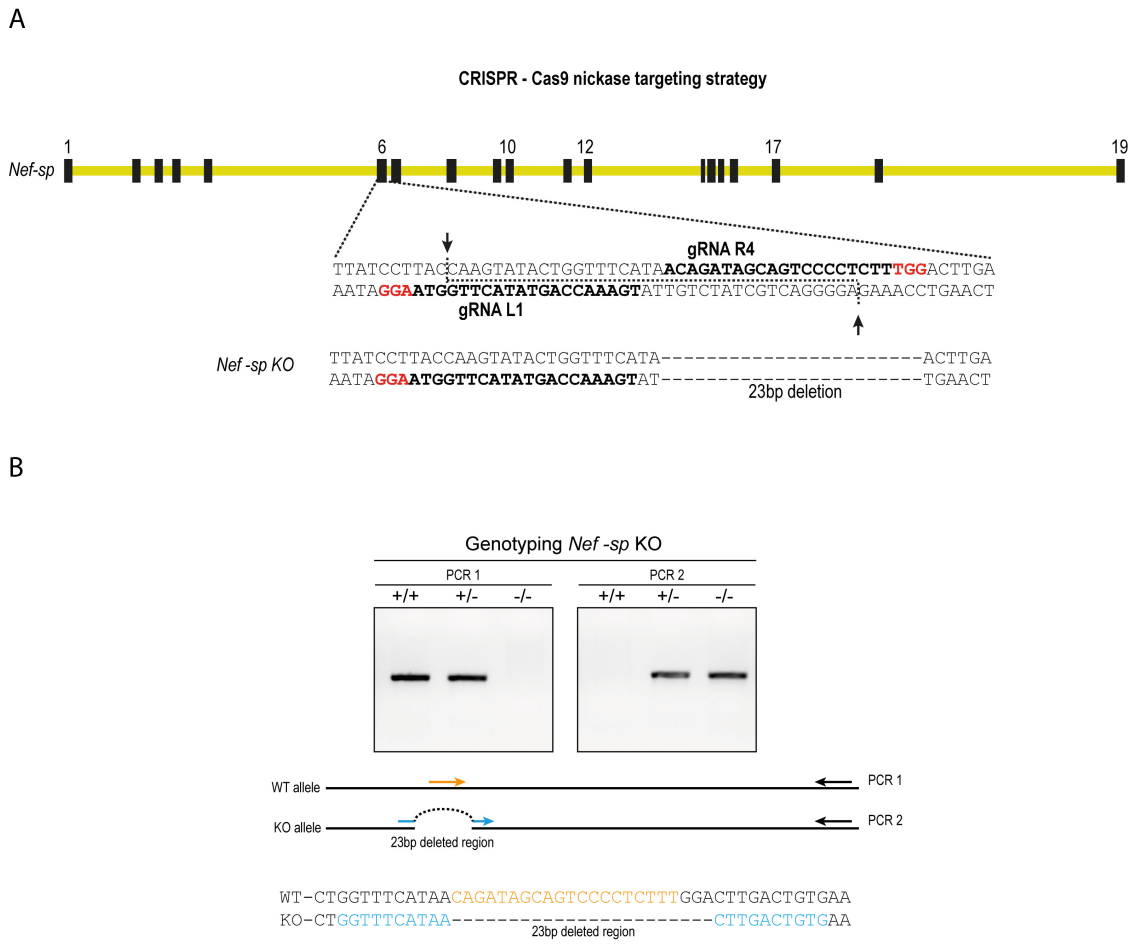


Figure 2.1 – Generation of *Nef-sp* knock-out mice using the CRISPR-Cas9 system. (A) gRNAs used to create the knock-out mouse are indicated in bold. An allele with a 23bp deletion was used to establish the *Nef-sp* knock-out line. (B) Genotyping of *Nef-sp* KO mice. Ethidium bromide stained gel with the products obtained from PCR 1 and PCR 2. Primer SS330 (represented by orange arrow) only anneals in the wild-type allele and primer SS323 (represented by blue arrow) only anneals in the knock-out allele. The reverse primer SS336 is represented by the black arrow.

Tail genomic DNA isolation and genotyping PCR conditions: Tails were digested in buffer (50 mM Tris-HCl pH 8.0, 100 mM EDTA, 100 mM NaCl, 1 % SDS with 25 µg of Proteinase K) at 55 °C overnight. DNA was precipitated with isopropanol, washed in 70 % (v/v) ethanol and resuspended in 10 mM Tris-HCl pH 8.0. To identify homozygous animals two PCRs were performed: PCR 1: using primers SS330 (represented in orange) and SS336. PCR 2: using primers SS323 (represented in blue) and SS336 (Figure 2.1B). Genotyping primers: SS330 (5'–CAGATAGCAGTCCCCTCTTT–3'), SS323 (5'–GGTTTCATAACTTGACTGTG–3'), SS336 (5'–GGATTACGGAAACTCAAAGTG–3'). The PCR reactions were performed in 1X Taq Buffer (Fermentas), 200 µM dNTPs, 0.25 µM primers, 2.5 µM of $MgCl_2$, 2 µL of DNA template and 0.5 µL of Taq DNA polymerase in a final volume of 25 µL. PCR conditions were: 95 °C for 5 min, then 40 cycles of 30 sec at 95 °C, 30 sec at 60 °C and 30 sec at 72 °C and a final step at 72 °C for 5 min.

2.1.9 Total RNA Libraries and Bioinformatics

Total testes RNA was isolated with TRIzol Reagent (Invitrogen cat. No. 15596026). Strand-specific RNA-seq libraries were prepared using TruSeq Stranded Total RNA Sample Preparation Kit (Illumina) after removal of abundant ribosomal RNAs with Ribo-Zero. Libraries were sequenced with the Illumina HiSeq 2000 platform (EMBL Heidelberg Gene Core facility) for 50 cycles. Reads were sorted into individual libraries based on the barcodes, the 3' adapter sequences were removed and mapped to the mouse genome (mm9). The software used for processing the data (genomic coordinates) from the raw data files are in-house tools developed by the Sachidanandam Lab (Olson *et al.* 2008). Only reads perfectly matching the genome were kept for further analysis. The reads were divided into groups based on their annotation (repeat sense, repeat antisense, gene exon sense, gene exon antisense, gene intron sense, gene intron antisense, ncRNA sense, ncRNA antisense, miRNA sense, miRNA antisense, none) and their counts were compared between *Nef-sp* $^{-/-}$ and *Nef-sp*$^{+/-}$. DESeq2 bioconductor package (Love *et al.* 2014) was used to search for differentially expressed transcripts. Adjusted p-value 0.05 was used as a threshold for statistical significance.

2.2 METTL4 project

2.2.1 Clones and constructs

The coding sequence of mouse *Mettl4* (Genbank Accession number: NM_176917.4), human *Mettl4* (Genbank Accession number: NP_073751.3), *Bombyx mori Mettl4* and *Drosophila melanogaster Mettl4* were cloned by RT-PCR. The total RNA used as a template for cDNA synthesizes was obtained from the different sources using TRIzol Reagent (described in section 2.1.1). Mouse total RNA was obtained from adult mouse testis. HEK293T total RNA and adult human brain total RNA (Clontech, cat No. 636530) were used to clone human *Mettl4*. Due to the difficulty in obtaining the full length cDNA by RT-PCR the human *Mettl4* was cloned as two fragments and joined together by overlap PCR. *B. mori Mettl4* and *D. melanogaster Mettl4* were cloned by using total RNA extracted from BmN4 cell line and OSCs (ovarian somatic cells) respectively.

Mouse METTL4 (mMETTL4) – Isoform 1 (471 aa)

MSVVHHLPPGWLLDHLSFINKVNYQLCQHQESFCSKNNPTSSVYMDSLQLDPGSPFGAPAMCFAPDFTTVSGNDDEGSCEVITEKYVFRSELFN
VTKPYIVPAVHKERQQSNKNENLVTDYKQEVSVSVGKKRKRCIAFNQGELDAMEYHTKIRELILDGSSKLIQEGLRSGFLYPLVEKQDGSSGCI
TLPLDACNLSELCEMAKHLPSLNEMELQTLQLMGDDVSVIELDLSSQIIENNSSFSKMITLMGQKYLLPPQSSFLLSDISCMQPLLNCGKTFDA
IVIDPPWENKSVKRSNRYSSLSPQQIKRMPIPKLAAADCLIVTWVTNRQKHLCFVKEELYPSWSVEVVAEWYWVKITNSGEFVFPLDSPHKKPY
ECLVLGRVKEKTPLALRNPDVRIPPVPDQKLIVSVPCVLHSHKPPLTEVLRDYIKPGGQCLELFARNLQPGWMSWGNEVLKFQHMDYFIALESG
C

Mouse METTL4 – Isoform 2 (450 aa)

MSVVHHLPPGWLLDHLSFINKVNYQLCQHQESFCSKNNPTSSVYMDSLQLDPGSPFGAPAMCFAPDFTTVSGNDDEGSCEVITEKYVFRSELFN
VTKPYIVPAVHKERQQSNKNENLVTDYKQEVSVSVGKIRELILDGSSKLIQEGLRSGFLYPLVEKQDGSSGCITLPLDACNLSELCEMAKHLPS
LNEMELQTLQLMGDDVSVIELDLSSQIIENNSSFSKMITLMGQKYLLPPQSSFLLSDISCMQPLLNCGKTFDAIVIDPPWENKSVKRSNRYSSL
SPQQIKRMPIPKLAAADCLIVTWVTNRQKHLCFVKEELYPSWSVEVVAEWYWVKITNSGEFVFPLDSPHKKPYECLVLGRVKEKTPLALRNPDV
RIPPVPDQKLIVSVPCVLHSHKPPLTEVLRDYIKPGGQCLELFARNLQPGWMSWGNEVLKFQHMDYFIALESGC

Human METTL4 (hMETTL4) – Isoform 1 (472 aa)

MSVVHQLSAGWLLDHLSFINKINYQLHQHHEPCCRKKEFTTSVHFESLQMDSVSSSGVCAAFIASDSSTKPENDDGGNYEMFTRKFVFRPELFD
VTKPYITPAVHKECQQSNEKEDLMNGVKKEISISIIGKKRKRCVVFNQGELDAMEYHTKIRELILDGSLQLIQEGLKSGFLYPLFEKQDKGSKP
ITLPLDACSLSELCEMAKHLPSLNEMEHQTLQLVEEDTSVTEQDLFLRVVENNSSFTKVITLMGQKYLLPPKSSFLLSDISCMQPLLNYRKTFD
VIVIDPPWQNKSVKRSNRYSYLSPLQIQQIPIPKLAAPNCLLVTWVTNRQKHLRFIKEELYPSWSVEVVAEWHWVKITNSGEFVFPLDSPHKKP
YEGLILGRVQEKTALPLRNADVNVLPIPDHKLIVSVPCTLHSHKPPLAEVLKDYIKPDGEYLELFARNLQPGWTSWGNEVLKFQHVDYFIAVES
GS

Human METTL4 (hMETTL4) – Isoform 2 (422 aa)

MSVVHQLSAGWLLDHLSFINKINYQLHQHHEPCCRKKEFTTSVHFESLQMDSVSSSGVCAAFIASDSSTKPENDDGGNYEMFTRKFVFRPELFD
VTKPYITPAVHKECQQSNEKEDLMNGVKKEISISIIGKDKGSKPITLPLDACSLSELCEMAKHLPSLNEMEHQTLQLVEEDTSVTEQDLFLRVV
ENNSSFTKVITLMGQKYLLPPKSSFLLSDISCMQPLLNYRKTFDVIVIDPPWQNKSVKRSNRYSYLSPLQIQQIPIPKLAAPNCLLVTWVTNRQ
KHLRFIKEELYPSWSVEVVAEWHWVKITNSGEFVFPLDSPHKKPYEGLILGRVQEKTALPLRNADVNVLPIPDHKLIVSVPCTLHSHKPPLAEV
LKDYIKPDGEYLELFARNLQPGWTSWGNEVLKFQHVDYFIAVESGS

***Bombyx mori* METTL4 – (347 aa)**

MSLVIKQGNCCLIDHWEFIAKLYKDVLFEENYQSYTLSAKLFSITSEVRTRKRKKVYEELSIETSKIKRMYEEFIAKVPNSIKERISQKYNILD
TSSVRDFSASLFQSTIFDHHSINGGSTSDIPLKCKIQNEYFLIPPNSRFYYGSVNEECHKLDSGIFDLVIADPPWWNRYIRRLKNANEQLSYEM
MYNEDIAAIPLSNLLSSNCLVAVWCTNSPANIQAVKNLIFRNWGVRYVASWHWLKVAVDLSPICPFGTGSTKQPYEMLIIGKVGSVAPIPDGQL
IVSIPSALHSHKPPLLDLLKPYINKEQPRILELFARYLLPNTTSVGYEPLKWQHISLYETVDISE

Drosophila melanogaster METTL4 – (359 aa)

MLKLQKKTEDSKFAVFLDHKTLINEAYDEFKLKSELFQFHAKKTDKGIEEDKTRKRKRKAGVEDASSLEDLHLVNEYLELLSKPVEPEDSSPMK
RHWEDGYNVPQLHGANESGRMQRFLRVDGSRGVYLIPNQSRFFNHNVDNLPALLHQLLPAYDLIVLDPPWRNKYIRRLKRAKPELGYSMLSNEQ
LSHIPLSKLTHPRSLVAIWCTNSTLHQLALEQQLLPSWNLRLLHKLRWYKLSTDHELIAPPQSDLTQKQPYEMLYVACRSDASENYGKDIQQTE
LIFSVPSIVHSHKPPLLSWLREHLLLDKDQLEPNCLELFARYLHPHFTSIGLEVLKLMDERLYEVRKVEHCNQEEVN

For mammalian cell culture expression, mouse and human METTL4 were cloned into the EcoRI and NotI sites of the vector pCI-neo (N-HA-fusion tag). Initial expression studies with N-terminal tagged proteins revealed that the fusion protein was poorly expressed. However, C-terminal tagged proteins were robustly expressed and therefore all mouse and human proteins used in this study were tagged at C-terminal. Constructs were cloned in pCI-neo using NheI and NotI restrictions sites. To generate rabbit polyclonal antibodies, coding sequence for the region (1-145 aa) of the mouse METTL4 protein was cloned into the pETM-11 bacterial expression vector (6xHis-TEV-fusion tag; EMBL Protein Expression and Purification Core Facility) in the restriction sites NcoI and NotI. For bacterial expression of mouse METTL4 with N-terminal 6xHis-SUMO tag, constructs were cloned in the NcoI and NotI restriction sites of pETM-11-SUMO vector (EMBL Protein Expression and Purification Core Facility). Expression with C-terminal tag was done by cloning in the XbaI and XhoI restriction sites of pETM-30 vector (EMBL Protein Expression and Purification Core Facility). Eukaryotic expression using insect cells was performed by cloning into the NheI and KpnI sites of the vector pACEBac2-SUMO. All the constructs used in this study are listed in the table 2.8.

Table 2.8 – List of constructs used in METTL4 study.

Organism	Protein	Expression host	Plasmid	Tag	Construct name
Homo sapiens	METTL4 (1-472 aa)	Mammalian	pCI-neo-N-HA	C-terminal N-HA	hMETTL4
Mus musculus	METTL4 (1-471 aa)				mMETTL4
Mus musculus	METTL4 (1-471 aa)	Bacteria	pETM-11-SUMO	6xHis-SUMO-TEV (N-term)	mMETTL4
D. melanogaster	METTL4 (1-359 aa)				DmMETTL4
Mus musculus	METTL4 (1-471 aa)	Insect cells	pACEBac2-SUMO	6xHis-Strep- SUMO-TEV (N-term)	mMETTL4
Bombyx mori	METTL4 (1-347 aa)				BmMETTL4
Mus musculus	METTL4 (1-471 aa)	Bacteria	pETM-30	TEV-SUMO-6xHis (C-term)	Mouse METTL4
		Insect cells	pACEBac2-SUMO	TEV-SUMO-6xHis (C-term)	Mouse METTL4
	METTL4 (143-471 aa)	Bacteria	pETM-30	TEV-SUMO-6xHis (C-term)	mMETTL4-A
	METTL4 (279-471 aa)				mMETTL4-B
	METTL4 (143-300 aa)				mMETTL4-C
	METTL4 (46-471 aa)				mMETTL4-D
	METTL4 (3-187 aa)				mMETTL4-E
	METTL4 (1-145 aa)		pETM-11	6xHis (N-term)	N-terminal antigen

2.2.2 Recombinant mouse METTL4 production and purification

Production of METTL4 from bacteria or insect cells was done similarly to the production of recombinant human NEF-sp (described in section 2.1.3). Expression trials to produce recombinant *D. melanogaster* and mouse METTL4 were initially made using *E. coli* (BL21 strain). All the constructs, either C-terminal or N-terminal tagged, were poorly expressed in bacteria (discussed in section 4.2.1). Therefore, we tried to express full-length *B. mori* and mouse METTL4 in insect cells (table 2.9). Mouse METTL4 with a C-terminal tag was the only soluble construct.

Table 2.9 – List of constructs expressed in bacteria and insect cells. Construct boundaries, expression system, tag, construct name and solubility are indicated.

Organism	Protein	Expression host	Tag	Solubility
Mus musculus	METTL4 (1-471 aa)	*E. coli*	6xHis-SUMO-TEV (N-term)	Insoluble
D. melanogaster	METTL4 (1-359 aa)			Insoluble
Mus musculus	METTL4 (1-471 aa)	Insect cells	6xHis-Strep- SUMO-TEV (N-term)	Insoluble
Bombyx mori	METTL4 (1-347 aa)			Insoluble
Mus musculus	METTL4 (1-471 aa)	*E. coli*	TEV-SUMO-6xHis (C-term)	Insoluble
		Insect cells	TEV-SUMO-6xHis (C-term)	Soluble

Purification of mouse METTL4 was based in the protocol used to purify human NEF-sp (described in section 2.1.3). The full-length mouse METTL4 (1-471 aa) and the point mutant (D287A) were purified by Ni^{+2}-affinity chromatography in Lysis Buffer (table 2.3) supplemented with protease inhibitor (Roche; Complete EDTA-free). Washing and elution steps were done as described in section 2.1.3. After Ni-NTA elution, two different purification protocols were followed depending on the final purpose of the purified protein. Methylation assays were done with METTL4 tagged at C-terminal with TEV-SUMO-6xHis. After Ni-NTA, the protein was further purified on a HiTrap™ Q HP column (GE Healthcare; 17-1153-01) followed by a HiTrap™ Heparin HP (GE Healthcare; 17-0406-01). The purest fractions were mono-dispersed by gel filtration chromatography (Superdex 75 10/300 GL; GE healthcare) in a final buffer of 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5 % Glycerol and 5 mM of β-mercaptoethanol. For structural studies, the C-terminal tag was removed by TEV cleavage. After Ni-NTA, the eluted fractions were pooled together in a 10 kDa dialysis membrane supplemented with 500 µL of TEV (10 $mg.mL^{-1}$). The tag was cleaved overnight at 4°C in dialysis buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 % Glycerol and 5 mM of β-mercaptoethanol). The protein was further purified using an HiTrap™ Heparin HP (GE Healthcare; 17-0406-01) and mono-dispersed fractions were collected by gel filtration chromatography (Superdex 75 10/300 GL; GE healthcare) in a final buffer of 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5 % Glycerol and 5 mM of β-mercaptoethanol.

2.2.3 *In vitro* methylation and 6mA-IP-seq

In vitro methylation reactions (100 µL) were done by mixing 1000 pmols of the dsDNA library (prepared as described in section 2.2.5), 10 µg of mouse METTL4 (WT or D287A mutant) and 2 µL of S-adenosyl methionine (stock of 32 mM; NEB). Reactions were incubated for 5 hours at 37 °C. A control reaction was done using 80 U of Dam DNA methyltransferase (NEB). After incubating, reactions were treated with proteinase K and extracted with phenol-chloroform. The purified DNAs were resuspended in 50 µL of wash buffer (10 mM Tris-Cl, pH 8.0, 150 mM NaCl, 0.05 % NP-40) and incubated with 1.5 µg of anti-6mA antibody (Synaptic Systems; Cat. no. 202 003) for 1 hour at 20°C on a Thermomixer (Eppendorf). Low-bind tubes were used to prevent loss of nucleic acids by stickiness to tube walls. Required amount (50 µL volume of the product = 1.5 mg beads) of Dynabeads Protein G (Thermo Fisher Scientific) were washed/resuspended in wash buffer and added to each tube to get a total volume of 100 µL. After incubation for 30-45 min, beads were washed five times with wash buffer and transferred to a new tube. Beads were subject to Proteinase K treatment and bound DNAs were extracted with phenol-chloroform. Isolated nucleic acids were resuspended in 6 µL water and used for library preparation.

2.2.4 *In vitro* methylation assays

In vitro methylation assays were done by incubating ^{32}P-end-labelled dsDNA oligonucleotides (table 2.10) with purified recombinant mouse METTL4. One of the single strands (10 pmols) within a double-stranded DNA (dsDNA) is end-labelled with ^{32}P-γ-ATP (Perkin Elmer; NEG002A001MC) using polynucleotide kinase (PNK; Thermo Fisher). Labelled oligonucleotide is purified by gel elution, dissolved in 50 µL and stored at -20°C until further use. For preparation of dsDNA, 0.8 pmols of labelled ssDNA was annealed with 10 pmols of the cold partner strand. *In vitro* methylation reactions containing 0.08 pmols of labelled dsDNA, 1-5 µg of recombinant mouse METTL4 protein (WT or D287A mutant) or 16 U of Dam methyltransferase (NEB) and 80 µM of SAM were set up in the reaction buffer (50 mM Tris-HCl, 5 mM β-mercaptoethanol, 10 mM EDTA, pH 7.5). Incubations were done at 37 °C for 2-5 hours. Nucleic acids were extracted by phenol-extraction and precipitated. These were then subjected to restriction digest (20 µL volume) by either DpnI or DpnII (NEB) for 2 hours at 37 °C. DpnI cuts only when its recognition sequence GA^{m}TC is 6mA-methylated, and DpnII cuts only when its recognition sequence GATC is unmethylated. After digestion, the samples were directly resolved by 18% urea-PAGE and the gel was fixed for 10 min (40 % methanol, 10 % acetic acid and 3 % glycerol), then dried at 80 °C for 3 hours. Dried gels were exposed to Storage Phosphor screens (GE Health; reference no. 28-9564-75) and scanned with a Typhoon FLA 9500 scanner (GE Health; reference no. 29-1885-90). The dsDNA RPOligo 17 was used for the methylation assays in Figure 4.3C.

For direct examination of methylation activity of mouse METTL4, recombinant protein was incubated with the radioactively-labelled methyl donor S-adenosyl-L-[methyl-^{14}C] methionine (Perkin Elmer; NEC363010UC) and various ssRNA, ssDNA or dsDNA substrates (table 2.10). We used the purified

6mA RNA methyltransferase human METTL3-METTL14 complex and 6mA DNA methyltransferase *E.coli* Dam (NEB) as positive controls. Reactions of 50 µL were set up with the following components: 500 pmol of nucleic acid substrate, 5 µg of purified enzyme (METTL proteins) or 40 U of Dam, 10 U of RNasin (Thermo Ficher Scientific), 0.1 µCi of ^{14}C-SAM, 50 mM Tris-HCl, pH 7.5, 100 mM KCl, 5 mM $MgCl_2$ and 2 mM DTT. Reactions were incubated for 8 h at 37 °C and extracted with phenol-chloroform followed by precipitation with ethanol. Nucleic acids were resolved by 18% urea-PAGE, dried, exposed to Storage Phosphor screens (GE Health; reference no. 28-9564-75) and scanned with a Typhoon FLA 9500 scanner (GE Health; reference no. 29-1885-90).

Table 2.10 – List of oligonucleotides used in the *in vitro* methylation assays. Consensus sequences are marked in red.

Name	Sequence
DNA	
RPOligo 17	Sense: 5'-CTCGGCATCAGATCGGAGTCTTACTACAGGCACC-3'
	Antisense: 5'-GGTGCCTGTAGTAAGACTCCGATCTGATGCCGAG-3'
SS497	Sense: 5'-GGCTGGAGGAGTCAG-3'
	Antisense: 5'-CTGACTCCTCCAGCC-3'
SS499	Sense: 5'-GGCTGCATGAGTCAG-3'
	Antisense: 5'-CTGACTCATGCAGCC-3'
SS501	Sense: 5'-GGCTGGATCAGTCAG-3'
	Antisense: 5'-CTGACTGATCCAGCC-3'
SS503	Sense: 5'-GCTGTTAGAAAGTCA-3'
	Antisense: 5'-TGACTTTCTAACAGC-3'
Reporter: dsDNA fragment with all the NNANN combinations	5'ACCACACTCGTTCGTTTTCCACAGGCTACAACATGATACTTTAACGC AAGGTAGGCTATTTAAAAATATTGAATAGTATCAAAGCGGAATCAAA AAGATCTGACCTGAGGAAATCTAAGGCTAAGACAGATGATTACACTC TAATCGAATCCATCGCAACGGAGCATAATACATCGAAGTCAAGAAGA GTCTAACTTAGCCCACCTGATAACACCGGATTTAAGCCTACCCGAGTA CAGTGTAGATTAATATACCCAAGTTAACGGCAATGTAGTATATAGAA AACGAAGTTAAGCTAGCGTACGATAGCGGAACGGAGGAAATGCGAT CGCACGTCAAGTAAATTTAAATTATTCCAAAGCATCAAAACATAGGCA ACATAAGTTAACTGGACGTTATCACAATGGAAGAGACCCTAGGTCAG AGGACAGCACCTTAGTTAAAGGTAGCACACGGAAACCCAGAGCACAA GAGCGGAAATCAGTTCAGCACATTACAGCCCATTACAAGTCATCCTAC GTTACCTCACTAAAATCCAGGAAAAATAATGGAACTGTAAGGTACAT AATACCAACGGACTGCACTACAAATTATGTAAGACAATAGCATGTCAT TCTAGTGTAACCTACTAAATTGAACACCACTGCAGGAAACCATATCCA AACGGAGATGAGTAGACTAAATATGAATGAATGTCAATATATGATAC ATGAGCTGACACCAATAGAAGATAACTCACCCTAGAACATAAAACTG TAATTTATAACAACTTAGACCATGGTACCGTACTTAACATAAACCAAA GCGAGCAGAAACGACGCGACTGAACGTCAGGCAATTTGACGCGATT CAACGGTAAAAGACACCAGTCGAAAGTATTATAAGAAAGTCAAATGA AGCTCAACTTACGCAATCTGACTTAACCGCATTGAAATACACATGAAC TCATGGAAAGCGAACGCATAAAAAGCGACCAGATAGCAGACTATCCC AAGGTATGTCATATAATCGGATAGAAGACTATAAGATTTCACACGAG GCGAGAAAACGATAAATTAGGCAAGGGAATTTGATCCCATACGATA GGAGTTCAAATCACGAAAGGGGATGATAGAGAATCAGAACTGAGAC CACGCCAGCAGACGCTATGCAAGCTTACATAATTAGAGGAGAATGGA CCATAGTCGACAGAAGGCAACTGCAGTTGAAGGCAGCTGAAACAACC GTATAAAAGATGATGCTAAAAGAGAGAACCAAACAGGATCGCAAACC ACAATACGTTACTCAATGAGATGCGATGTTGACCCATAATCGCTCCTC -3'
RNA	

RPRNA19	5'-GCGUCUUUACGGUGCUUAAAACAAAACAAAACAAAACAAA-3'
RPRNA20	5'-AAACAAAACAAAACAAAACAAAAUAGCACCGUAAAGACGC-3'
RPRNA28	5'-GGGAGCGUUCCGUCUGCUAUUAUCACG-3'
RPRNA29	5'-GAUAAGCUGUAGGAGGUUCUUCUAGUU-3'
RPRNA30	5'-GUGAUCGUCGUACACAAUGAAGAUUUA-3'
RPRNA31	5'-GCGCGAUCGAUCGUUGCUGACUGAGCC-3'
RPRNA36	5'-UGACAUGAACACAGGUGCUC-3'
MET1	5'-UACACUCGAUCUGGACUAAAGCUGCUC-3'
MET2	5'-UACACUCGAUCUGGAUUAAAGCUGCUC-3'

2.2.5 Bind-N-Seq: high-throughput analysis of *in vitro* protein-DNA interactions

Bind-N-Seq experiments were done by using a library of double-stranded DNA (dsDNA) oligonucleotides carrying a central 21 nt randomized sequence (Microsynth, CH). The flanking sequences had information necessary for amplification using primers suitable for sequencing in the Illumina platform. Forward sequencing primer is highlighted in yellow and barcode reading primer in green. The randomized 21 nt sequence is highlighted in red.

5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**21Ns**AGATCGGAAGA
GCACACGTCTGAACTCCAGTCACxxxxxxATCTCGTATGCCGTCTTCTGCTTG3'

The library was purchased in the form of single-stranded DNA. An aliquot of 1000 pmols of the library (enough for two binding reactions) was annealed with an equimolar amount of a reverse primer by denaturation at 90 °C for 2 min and allowed to cool down gradually to room temperature in a Thermomixer (Eppendorf). A 50 µL reaction containing the annealed oligonucleotides, 20 U of DNA Pol I (Thermo Scientific) and 5 µL of dNTP mix (10 mM stock) was extended for 1 hour at 37 °C to create the double-stranded DNA (dsDNA) library. A 5 µL aliquot of the reaction was run on a 2% agarose gel and compared with a similar amount of non-extended annealed oligonucleotides to confirm an increase in intensity with ethidium bromide staining.

Approximately 500 pmols of dsDNA library was incubated with 25 µg of recombinant mouse METTL4 protein (WT and D287A mutant) and incubated at 25 °C for 5 hours on a Thermomixer (Eppendorf). When untagged protein was used, complexes were isolated by incubating with anti-mouse METTL4 antibodies for 2 hours at 4°C and afterwards recovered with Dynabeads Protein G (Thermo Fisher Scientific). Beads were washed five times with washing buffer (10 mM Tris-Cl, pH 8.0, 150 mM NaCl, 0.05 % NP-40). Bound DNAs were purified by Proteinase K digestion for 1 hour at 55 °C, followed by phenol-chloroform extraction. Tagged mMETTL4-TEV-SUMO-6xHis proteins were incubated with the

dsDNA library for 5 hours at 20 °C in binding buffer (50 mM Tris, 150 mM NaCl, 5% Glycerol, 5 mM β-mercaptoethanol, 20 mM Imidazole, pH 8.0, 0.05 % NP-40). Complexes were then washed five times with wash buffer (50 mM Tris pH 8.0, 150 mM NaCl, 5% Glycerol, 5 mM β-mercaptoethanol, 40 mM Imidazole, , 0.05 % NP-40), with a change of tube in the last wash. Proteins were eluted three times by incubation in 100 µL of elution buffer (50 mM Tris, 150 mM NaCl, 5% Glycerol, 5 mM β-mercaptoethanol, 260 mM Imidazole, pH 8.0). Pooled elutions were treated with Proteinase K and bound DNAs were extracted with phenol-chloroform. DNA extracted from above procedures were dissolved in 6 µL water and used directly for barcoded library preparation suitable for Illumina sequencing.

2.2.6 Electrophoretic mobility shift assay (EMSA)

Increasing amounts of purified mouse METTL4 were incubated with dsDNA probed of 32 bp, 23 bp and 13 bp separately. ssDNA probes of 15 bp were also used (table 2.11). Binding was done in 20 mM HEPES pH 7.4, 150 mM NaCl, 1 mM DTT and 5% glycerol. Samples were resolved in 6 % native-PAGE gels. Gel was stained with ethidium bromide followed by staining with Instant Blue.

Table 2.11 – List of oligonucleotides used in EMSA

DNA probes	Sequence
dsDNA 32 bp	Sense: 5'-TGGAAGCCTTTTCAAGGAAACGAAAGTGAACT-3'
	Antisense: 5'-AAGTTCACTTTCGTTTCCTTGAAAAGGCTTCC-3'
dsDNA 23 bp	Sense: 5'-TTTCAAGGAAACGAAAGTGAAC-3'
	Antisense: 5'-AGTTCACTTTCGTTTCCTTGAA-3'
dsDNA 13 bp	Sense: 5'-TAAACGAAAGTGA-3'
	Antisense: 5'-ATCACTTTCGTTT-3'
ssDNA 1	5'–GGCTGGAGGAGTCAG-3'
ssDNA 2	5'–GGCTGCATGAGTCAG-3'
ssDNA 3	5'–GGCTGGATCAGTCAG-3'
ssDNA 4	5'–GCTGTTAGAAAGTCA-3'

2.2.7 Limited proteolysis of recombinant mouse METTL4

To identify well-defined structural domains by limited proteolysis, we digested mouse METTL4 recombinant protein with Trypsin (Roche, cat. No. 3708985001) and Chymotrypsin (Roche, cat. No. 11418467001). A 100 µL reaction containing 150 µg of METTL4 and 150 ng of proteases in the reaction buffer 50 mM Tris-HCl pH 7.5, 200 mM NaCl, 10 % Glycerol and 5 mM of β-mercaptoethanol were incubated at 25°C and removed aliquots of the reactions at different time-points 0, 5, 30 and 60 min. Reactions were stopped by adding SDS-PAGE loading dye. Products were separated by SDS-PAGE and stained with Coomassie Blue, indicated fragments (Figure 4.6 of section 4.2.5) were identified by acid hydrolysis and mass spectrometry (EMBL Proteomics Core Facility, EMBL Heidelberg).

Table 2.12 – List of fragments identified by trypsin digestion. The different constructs were expressed in bacteria. Shown are the construct boundaries, the expression system, their tag and solubility.

Organism	Protein	Expression host	Tag	Solubility
Mus musculus	METTL4 (143-471 aa)	*E. coli*	TEV-SUMO-6xHis (C-term)	Insoluble
	METTL4 (279-471 aa)			Insoluble
	METTL4 (143-300 aa)			Soluble
	METTL4 (46-471 aa)			Insoluble
	METTL4 (3-187 aa)			Insoluble

2.2.8 Antigen purification and antibody production

METTL4 antigen was produced in the *E. coli* (BL21 strain). *E. coli* was transformed with a bacterial expression vector containing an N-terminal 6xHis tag fused to mouse METTL4 (1-145 aa). Cultures were grown at 37 °C until O.D. 0.6 was reached. Antigen expression was induced by adding IPTG at a final concentration of 0.6 mM and grown overnight at 37 °C. After large scale expression, we verified that mouse METTL4 antigen was insoluble. We used the same purification methodology as described in the section 2.1.7

2.2.9 Cell culture and immunofluorescence

All the experiments done using Human embryonic kidney (HEK) 293T and human epithelioid cervix carcinoma (HeLa) cell lines were performed as described previously (section 2.1.6). Transfection in human cell lines was done using mouse METTL4 and human METTL4 cloned in pCI-neo with an N-HA C-terminal tag.

2.2.10 Generation of *Mettl4* knock-out flies

To generate knock-out lines for the *Drosophila melanogaster* orthologue of Mettl4 (DmMettl4; CG14906) we targeted its gene locus on the right arm of chromosome 3 using the RNA-guided bacterial endonuclease Cas9. The DmMettl4 protein is encoded within a single large exon. Two independent guide RNAs (gRNAs) predicted to generate a ~50 bp deletion in the methyltransferase domain of DmMettl4 were cloned in pCFD4 plasmid (Addgene; 49411) and injected into embryos from the nos-Cas9.P aka CFD2 with the genotype (y[1] M{w[+mC]=nos-Cas9.P}ZH-2A w[*]) (BestGene, USA). Once adult, crosses were set up in groups consisting of a single potential founder male and five wildtype females. F1 females were sacrificed for genomic PCR genotyping to identify transfer of the desired deletion. When confirmed, the remaining F1 males were crossed individually with wild-type females and the parent being sacrificed to identify the different KO lines. Genotyping was done using one single fly smashed in 50 µL of 10 mM Tris-HCl pH 8.2, 1 mM EDTA, 25 mM NaCl and 200 µg.mL^{-1} Proteinase K and incubated at 37 °C for 30 min followed by inactivation at 95 °C for 2 min. PCR reactions were performed in 1X Taq Buffer (Fermentas), 200 µM dNTPs, 0.25 µM primers, 2.5 µM of $MgCl_2$, 5 µL of DNA template and 1 µL of

Taq DNA polymerase in a final volume of 50 µL using as forward primer 5'-CCCGTGGCGTGTACTTGATTCC-3' and reverse primer 5'-GGGCAGCAGCTGCTGCTCCAGGG-3'. Initial denaturation step at 95 °C for 5 min followed by 35 cycles of 30 sec at 95 °C, 30 sec at 60 °C and 30 sec at 72 °C and a final step at 72 °C for 5 min. Four independent founder lines were established (# 6, 7, 18, 27) and they all showed out-of-frame deletions ranging from 52 to 74 bp, which are predicted to delete parts of the methyltransferase domain, including the catalytic residues DPPW. These were used to assess viability and fertility of the mutant flies.

2.2.11 Generation of *Mettl4* knock-out and knock-in (DPPW → APPW) mice

The *Mettl4* locus in the mouse genome was targeted using a single guide RNA (gRNA) and Cas9 endonuclease (Transgenic Mouse Facility, University of Geneva). The gRNA template for *in vitro* transcription was created by PCR using Phusion polymerase (Thermo Fischer; F-530S) in a final volume of 100 µL (20 µL 5X Phusion HF buffer, 67 µL ddH_2O, 2 µL 10mM dNTPs, 5 µL 10 µM Forward Primer, 5 µL 10 µM Reverse primer and 1 µL Phusion DNA polymerase). Samples were amplified under the following conditions: 98 °C for 30 sec followed by 35 cycles of 10 sec at 98 °C, 30 sec at 60 °C and 15 sec at 72 °C and a final step at 72 °C for 10 min.

Forward primer:

5'GAAATTAATACGACTCACTATAGGAATTGTGATTGATCCACCATGTTTTAGAGCTAGAAA TAGC'3

Reverse primer

3 CAAAATCTCGATCTTTATCGTTCAATTTTATTCCGATCAGGCAATAGTTGAACTTTTTCA CCGTGGCTCAGCCACGAAAA 5

PCR products were gel purified using QIAquick Gel Extraction Kit (Qiagen; 28706) followed by phenol-chloroform extraction. The PCR products were used as a template for *in vitro* transcription using the T7 MEGAshortscript kit (Ambion, AM1354). The reactions were assembled in 20 µL final volume (2 µL T7 10X Reaction Buffer, 2 µL T7 ATP Solution (75 mM), 2 µL T7 CTP Solution (75 mM), 2µL T7 GTP Solution (75 mM), 2µL T7 UTP Solution (75 mM), 8 µL Template DNA, 2 µL T7 Enzyme Mix) and incubated at 37 °C for 4 hours followed by digestion of the DNA template (1 µl turbo DNAse for 15 min at 37 °C). The RNA was purified by phenol-chloroform extraction followed by precipitation with ethanol.

We targeted the genomic sequence (exon 4) that encodes for the SAM-binding motif of METTL4. The designed gRNA overlaps with the NcoI restriction site, which may be disrupted by Cas9 cleavage and subsequent repair, allowing easy identification of the mutant alleles. We also wished to create a point mutant knock-in allele where the SAM binding motif DPPW is mutated to APPW. We designed a 167 nt ssDNA carrying the required DNA mutation and silent codon mutations in the proline residues to

abolish the NcoI restriction site, (GAT CCA CCA TGG → GCT CCT CCG TGG; DPPW → APPW) and flanking homology arms (highlighted in blue).

Sequence of ssDNA oligo used for homology repair

TCATCACATTACAGAGGTTGTTTCTGAAATTGTGCTTTTGATTTCAGGTGGTAAGACATTT
GATGCAATTGTGATTGCTCCTCCGTGGGAGAACAAATCGGTTAAAAGAAGTAACAGGTAT
GTGGGATCACTTACCTGTCAGATTATATTCTAGAACTGCAGTTTCT

A mixture containing the gRNA, Cas9 mRNA (A29378; Thermo Fisher Scientific) and the ssDNA oligo (Integrated DNA Technologies, USA) for homology-mediated repair was injected into one-cell mouse embryos of B6D2F1 (mixed background of 50% C57BL6/50% DBA). Embryos were implanted into foster mothers NMRI. The black/grey coat colour of the embryo-derived mice is distinguishable from the white coat colour of the mothers. Founder animals were genotyped by PCR using genomic DNA obtained from ear punches used for animal numbering. The ear punch was digested in 100 µL of acidic solution (10 mM NaOH and 0.1 mM EDTA) for 90 min at 95 °C followed by neutralization with 100 µL of neutralization solution (20 mM Tris-HCl pH 8.0 and 0.1 mM EDTA). PCR was performed as described below and the PCR products were purified using QIAquick PCR Purification Kit (Qiagen, cat. no 28106) followed by digestion with NcoI-HF (NEB, cat. no R3193S). The digested fragments were resolved in 2 % agarose gel. Only the wild-type allele carrying an intact restriction site is expected to be digested, while any mutations introduced will show up as undigested DNA. We identified one founder (female) with a 2-nt out-of-frame deletion that is predicted to lead to premature termination of translation within the methyltransferase domain, and will be referred to as the *Mettl4* knock-out allele. Sequence analysis indicates that although the NcoI site is left intact in the mutant allele, it is likely that formation of hybrids with the wildtype allele renders the DNA non-cleavable. Another founder (male) with the desired knock-in mutations in the SAM binding motif and disruption of the NcoI restriction site was identified. We will refer to this allele as the *Mettl4* knock-in allele (Figure S2B). The founder animals were crossed with wildtype C57Bl/6J Rj males (Janvier labs, France). The resulting F1 animals were genotyped to confirm germline transmission. Mice carrying the following alleles were used for further experiments: *Mettl4* knock-out (KO) allele with 2 bp out-of-frame deletion in methyltransferase domain. *Mettl4* knock-in (KI) allele with the D287A point mutation in the SAM-binding motif.

PCR primers used: To identify the F1 animals, the forward primer (5'-GAAGAATTCTGCGTATTCAG-3') located 171 bp upstream the NcoI restriction site and the reserve primer (5'-GCCTTGAATAGTTCAGGCCTC-3') located 267 bp downstream the NcoI restriction site were used. PCR reactions: The reactions were performed in 1X Taq Buffer (Fermentas), 200 µM dNTPs, 0.25 µM primers, 2.5 µM of $MgCl_2$, 5 µL of DNA template and 1 µL of Taq DNA polymerase in a final volume of 50 µL. PCR program: Initial denaturation step at 95 °C for 5min followed by 35 cycles of 30 sec at 95 °C, 30 sec at 55 °C and 30 sec at 72 °C and a final step at 72 °C for 5min.

3. Characterization of the mammalian RNA Exonuclease 5/NEF-sp as a testis-specific 3′→5′ exoribonuclease

Résumé

Caractérisation d'ARN Exonucléase 5/NEF-sp: une exoribonucléase 3'→ 5' spécifique du testicule

Les études biochimiques et génétiques ont révélé que de nombreuses protéines jouent un rôle important dans la voie des ARNpi (ARN interagissant avec Piwi). Néanmoins, certaines caractéristiques mécanistes ne sont pas encore complètement comprises. D'autre part, certaines protéines impliquées dans la voie des ARNpi pourraient ne pas encore être caractérisées. Lorsque ce projet a commencé, le traitement 3' des ARNpi n'avait pas été complètement identifié. Par contre la directionnalité 3' à 5' avait été observée avec une dépendance au Mg^{2+}. Toutefois, la protéine responsable du raccourcissement des ARNpi à leur longueur mature n'était toujours pas connue.

NEF-sp devrait être une exo nucléase d'ARN avec une directionnalité de 3' à 5'. Sa prédiction de domaine a révélé que cela se compose d'un domaine de nucléase N-terminal et de deux motifs de reconnaissance d'ARN (RRM). NEF-sp est également prévue comme une nucléase spécifique aux testicules, mais son rôle biologique n'a pas été caractérisé auparavant. Ce projet s'est concentré sur une hypothèse initiale selon laquelle NEF-sp pourrait potentiellement agir comme une nucléase impliquée dans la voie des ARNpi.

Globalement, nous avons démontré que NEF-sp humaine est une exoribonucléase avec une directionnalité 3'→ 5'qui est active sur les substrats de l'ARN simple brin. Cette protéine est exprimée exclusivement dans les testicules des souris. En plus, NEF-sp pourrait fonctionner dans le compartiment nucléaire. Notre analyse de la souris Nef-sp mutante n'a révélé aucun phénotype évident. Nous avons observé que les homozygotes des deux sexes sont viables et présentent une fertilité normale. De plus, NEF-sp ne semble pas jouer un rôle dans la voie des ARNpi. Il est possible que le recadrage du/des substrat(s) inconnu(s) ne soit pas important pour la viabilité et fertilité en raison de la compensation potentielle par d'autres nucléases. Néanmoins, notre étude fournit une caractérisation biochimique et génétique de l'exoribonucléase NEF-sp des mammifères.

3.1 Introduction

Genetic and biochemical studies have identified more than twenty different proteins involved in piRNA pathway. However, some of the mechanistic features are still not fully understood. When this project started, piRNA 3' end processing was not completely characterized. In mice, pre-piRNAs are 30-40 nt in length and therefore their 3' end needs to be shortened. Trimming was observed before in BmN4 cells and an Mg^{2+}-dependent 3' to 5' exonuclease was implicated in this process (Kawaoka *et al.* 2011). However, the protein responsible for piRNA maturation was still not known. Two different studies demonstrated that *Bombyx mori* Papi and mouse orthologue Tdrdkh are involved in pre-piRNA trimming. Upon deletion of these proteins an accumulation of longer piRNAs was observed (Honda *et al.* 2013, Saxe *et al.* 2013). However, BmPapi and Tdrdkh lack any recognizable nuclease domain. Therefore, it was proposed that these proteins influence 3' processing by interacting with the nuclease responsible for shortening piRNAs 3' ends (commonly referred as Trimmer).

Our attempts to identify the Trimmer by biochemical approaches were unsuccessful (as done previously (Reuter *et al.* 2009)). Genetics studies in flies were also unable to identify the nuclease involved in this process (Handler *et al.* 2013). Therefore, we decided to search for a nuclease which, as most of the piRNA factors, is exclusively expressed in the testis. This approach led us to the identification of a nuclease called NEF-sp. Besides being a testis-specific factor, NEF-sp shows similarity to SDN1, a nuclease involved in degradation of small non-coding RNAs in plants. In *Arabidopsis thaliana*, SDN1 activity is crucial to maintain miRNA steady-state levels and for proper plant development (Ramachandran *et al.* 2008). Therefore, we thought that NEF-sp could be the nuclease involved in piRNA maturation.

NEF-sp is predicted to be an RNA exonuclease with 3' to 5' directionality. This protein is composed by three domains: a nuclease domain at N-terminal and two C-terminal RNA binding motifs. NEF-sp belongs to the superfamily of DEDD nucleases. A conserved histidine (H310) in motif III further classifies this protein as belonging to the DEDDh subfamily (Y. H. Zuo *et al.* 2001). Four catalytic residues (DEDD) are essential for NEF-sp nuclease activity. These are clustered in three different motifs: the first aspartic acid (D232) and glutamic acid (E234) residues are present in motif I, the second aspartic acid (D259) in motif II and the final aspartic acid (D315) and a fifth residue (H310) in motif three (Figure 1A). Nucleic acid degradation is catalysed by a two-metal-ion mechanism (Steitz *et al.* 1993). The catalytic residues are responsible for coordination of two metal ions essential for enzymatic activity. Metal A is thought to facilitate the nucleophilic attack by promoting the formation of hydroxide ion. Metal B stabilizes the intermediate state and acts as a Lewis acid facilitating the leaving of the oxyanion group (Figure 1B). The divalent metals most commonly used by the two-metal-ion mechanism are Mg^{2+}, Mn^{2+} or Zn^{2+} .

NEF-sp has two RNA recognition motifs (RRMs) at C-terminal. RRMs are one of the most common RNA binding motifs and typically these are composed by 80-90 residues. Usually two short consensus sequences: RP1 [(K/R)-G-(F/Y)-(G/A)-F-V-X-(F/Y)] and RP2 [(L/I)-(F/Y)-(V/I) (G/K)-(N/G)-(L/M)] are responsible for RNA binding. However, these two canonical consensus sequences are not present in NEF-sp. Structural studies demonstrated that RRMs have a βαββαβ topology consisting of four-stranded antiparallel β-sheet arranged in the order $\beta_4\beta_1\beta_3\beta_2$ and two α-helices (α_1 and α_2) packed against the β-sheet (Figure 1C) (Nagai *et al.* 1990). In general, these type of domains play an important role in recognizing specific RNA sequences (Kenan *et al.* 1991).

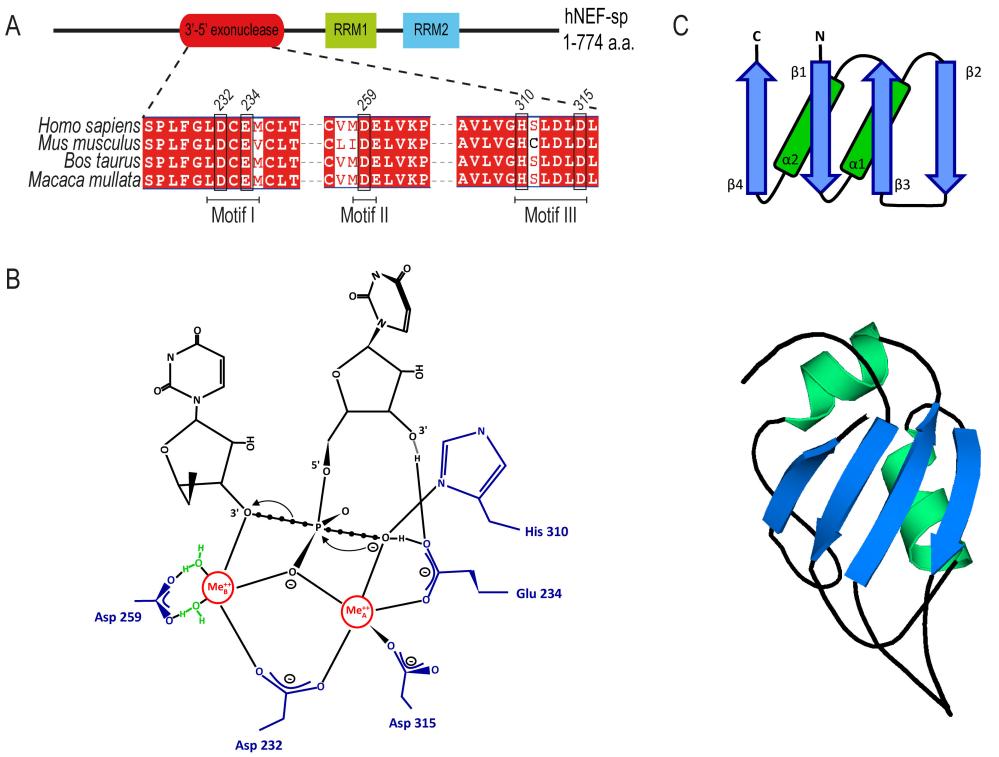


Figure 3.1 – Domain architecture of NEF-sp and representation of nuclease catalytic pocket. (A) Representation of the three motifs of NEF-sp nuclease domain that contain the catalytic resides (black rectangles). (B) Putative catalytic pocked of NEF-sp and RNA degradation by a two-metal-ion mechanism. NEF-sp catalytic residues are represented in blue and the metal ions are represented in red. Attacking water molecules are represented in green. (adapted from (Steitz *et al.* 1993)). (C) Topology diagram of a RRM domain (top panel). Structure of the RRM domain of U1 A protein (bottom panel) (adapted from (Nagai *et al.* 1990, Kenan *et al.* 1991)).

Some studies demonstrated that RRMs can also be responsible for protein-protein interactions. For instance, the RRM of UPF3b, a protein involved in nonsense-mediated mRNA decay (NMD), does not bind to an RNA target and instead interacts with another NMD factor UPF2 (Kadlec *et al.* 2004). Some proteins also have tandemly linked RRMs similarly to the predicted domain organization of NEF-sp. Several studies demonstrated that two tandem RNA recognition motifs function together and that this interaction increases the stringency of target recognition (Deo *et al.* 1999).

Although NEF-sp is predicted to be a testis-specific nuclease, there is no experimental data confirming this hypothesis. Its biological function is also unknown: does NEF-sp play a role in piRNA pathway? Is it essential for spermatogenesis? Throughout this project we tried to elucidate some of these aspects by characterizing biochemically and structurally human NEF-sp. Analysis of its biological relevance was also done by deleting NEF-sp in mice.

3.2 Results

3.2.1 NEF-sp is a testis-specific protein

Searching for potential nucleases led to the identification of NEF-sp/LOC81691 (Accession: NP_001185982) in testis transcriptomes. To confirm its presence in human and mouse testes we performed reverse transcription-PCR (RT-PCR). Total RNA was extracted from adult mouse testis while total human testis RNA was purchased from Clontech (cat no. 636533) (see section 2.1.1 of Material and Methods). A pair of specific primers was used to detect *Nef-sp* expression. We also used a pair of primers to detect a testis specific factor (*Mili*) (Kuramochi-Miyagawa et al. 2001) and a housekeeping gene (*G3pdh*). From the results obtained by RT-PCR we could observe that *Nef-sp* is expressed in mouse and human testes (Figure 3.2A).

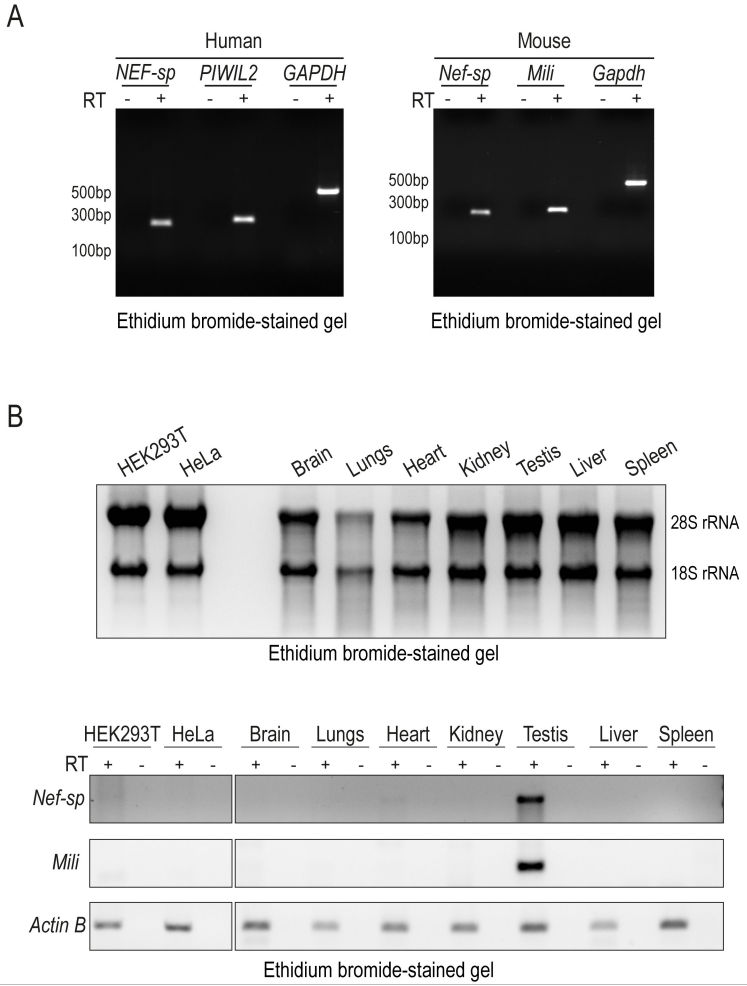


Figure 3.2 – Expression profile of *Nef-sp* in different mouse tissues and human cell lines. (A) *Nef-sp* transcript was detected in total RNA of human and mouse testes by RT-PCR. *Mili* was used as a testis-specific gene and *Gapdh* as ubiquitously expressed gene. (B) Agarose gel electrophoresis of total RNA isolated from human cell lines and mouse tissues (top panel) stained with ethidium bromide. RT-PCR of *Nef-sp*, *Mili* and *Actin B* using the total RNA extracted from the different sources (lower panel).

To analyse if *Nef-sp* is expressed exclusively in the testis, we extracted total RNA from seven mouse tissues and two human cell lines. RNA integrity was evaluated by denaturing agarose gel. We could observe the presence of two bands corresponding to the 28S and 18S ribosomal RNAs (rRNAs). Intensity of the 28S rRNA band was almost twice the intensity of the 18S rRNA band, indicating that the RNA was intact (Figure 3.2B top panel). We used this RNA to assess the expression of *Nef-sp* in multiple tissues. As previously, a pair of primers specific for *Nef-sp* and two controls: *Mili* (testis specific factor) and *Actin B* (housekeeping gene) were used. We did not detect the transcript in HeLa or HEK293T cells (Figure 3.2B lower panel). RT-PCR using total RNA from different mouse tissues indicated exclusive expression of *Nef-sp* in testes.

3.2.2 Recombinant expression of human NEF-sp

Recombinantly expressed proteins are a useful tool for biochemical and structural studies. For that reason, we tried to express recombinant human NEF-sp. The first trials were performed using *Escherichia coli* due to the fact of being a rapid and simple expression system. Human NEF-sp was cloned into pETM11-SUMO, a vector suitable for protein production in bacteria. N-terminus was tagged with 6xHis-SUMO to aid NEF-sp expression, solubility and purification (see section 2.1.3 of Material and Methods). hNEF-sp was poorly expressed in bacteria (Figure 3.3A) and even after enrichment by Ni-NTA affinity resin its levels were low and similar to other contaminates. Several attempts to make it more pure such as using different affinity columns were unsuccessful. Therefore, we decided to design a new construct (hNEF deletion 1), in which the first 41 N-terminal residues predicted as unfolded were deleted (Figure 3.6A). Similarly to the full-length protein, hNEF deletion 1 expression yield was low in *E. coli* (Figure 3.3B).

After unsuccessful attempts to express human NEF-sp in bacteria, we decided to change for a eukaryotic expression system. We used Baculovirus-insect cell system which should aid in the production of eukaryotic proteins due the presence of specific chaperones and post-translational modification machinery. Full-length human NEF-sp was expressed with an N-terminal 6xHis-Strep-SUMO fusion tag. After lysis in a suitable buffer, human NEF-sp was purified by Ni-NTA affinity resin (see section 2.1.3 of Material and Methods). The expression yield obtained using insect cells was significantly higher (Figure 3.3C). We tried to remove the 6xHis-Strep-SUMO fusion tag by overnight cleavage using TEV protease. The cleavage efficiency was around 50% (Figure 3.3D) and the downstream purifications were more difficult having two protein species (tagged-hNEF-sp and untagged-hNEF-sp) with similar proprieties. Therefore, we decided to maintain the 6xHis-Strep-SUMO tag and further purify human NEF-sp using a HiTrap Heparin HP column (Figure 3.3E). The most pure fractions were collected and mono-dispersed by size exclusion chromatography (Figure 3.3F). After the three chromatographic steps we could observe a single band corresponding to 6xHis-Strep-SUMO-hNEF-sp (Figure 3.3F).

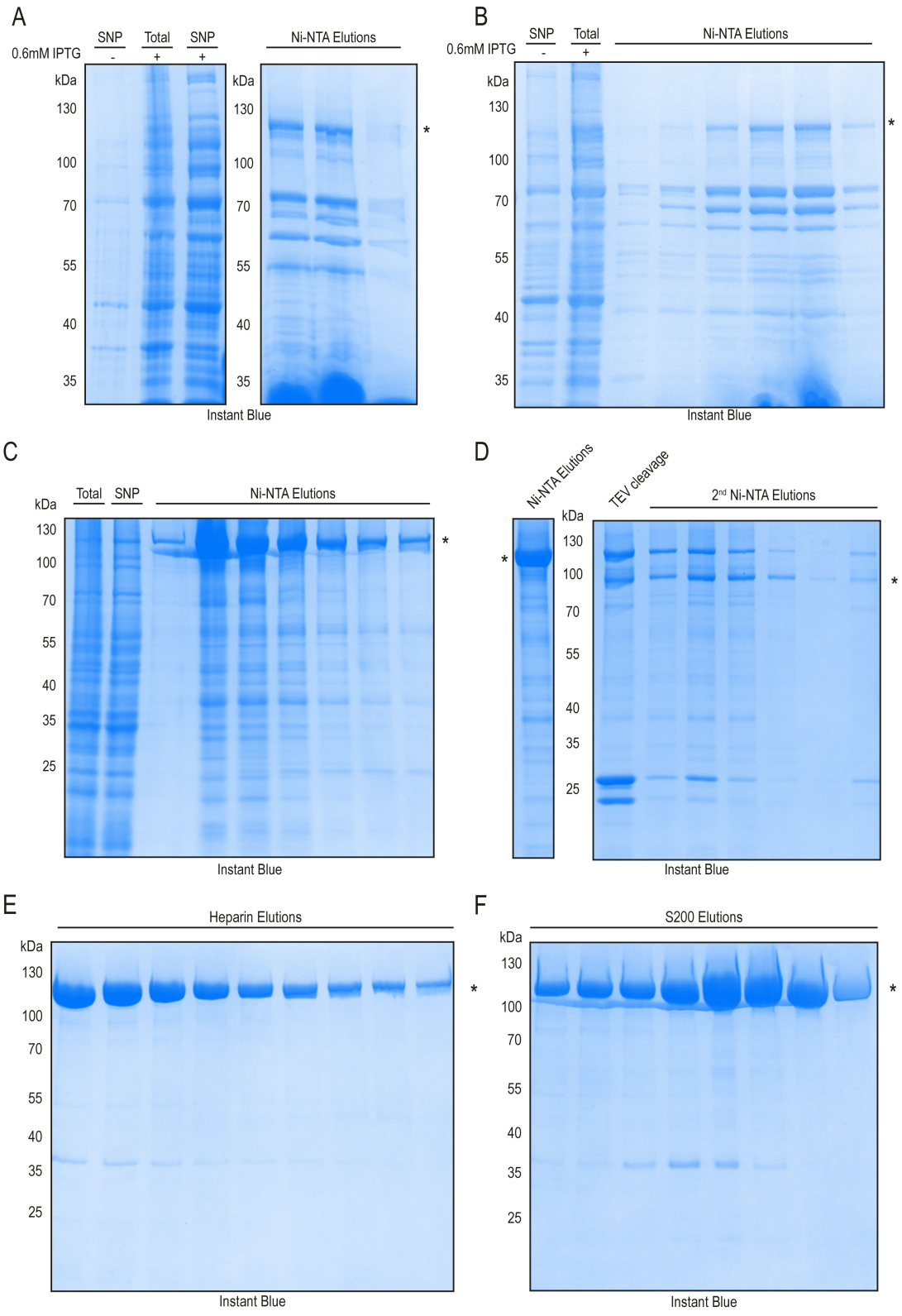


Figure 3.3 – Recombinant expression and purification of human NEF-sp. Expression trials of hNEF-sp in *E. coli* and purification by Ni-NTA affinity resin of: (A) full-length and (B) hNEF-sp deletion 1. Human NEF-sp fused to 6xHis-Strep-SUMO tag was expressed in insect cells and purified by: (C) Nickel-NTA affinity resin, (E) HiTrap Heparin HP column and the purest fractions were monodispersed by (F) size exclusion chromatography. (D) Attempts to cleave the N-terminal 6xHis-Strep-SUMO tag were unsuccessful. Samples were resolved by SDS-PAGE and stained with Instant Blue. *asterisk indicates the band corresponding to human NEF-sp.

3.2.3 human NEF-sp is a Mg-dependent 3′→5′ exonuclease

Having successfully established the purification protocol, we tested hNEF-sp nuclease activity. Based on sequence homology NEF-sp is a member of the DEDD superfamily of nucleases (Figure 3.4A). Degradation of nucleic acids is catalysed by a two-metal-ion mechanism in which the DEDD residues are essential for coordination of the two divalent metals ions required for activity. Mutations in any of these residues abolish the nuclease activity of members of this family. We made a mutation of a single residue present in motif III (D313A). This aspartic acid residue does not belong to the predicted catalytic tetrad. Full-length hNEF-sp and hNEF-sp^{Mut} were purified using the conditions established before (section 32.2). hNEF-sp^{Mut} behaves identically to wildtype protein during size-exclusion chromatography indicating lack of any gross conformational changes (Figure 3.4B).

In vitro nuclease assays consisted in incubating purified hNEF-sp with a 5'-end radioactively labelled single-stranded RNA (ssRNA) in specific buffer conditions for 1 hour at 37 °C. After incubation, the reaction was stopped by adding formamide loading buffer and heated at 72 °C for 2 min. Samples were resolved in a 15% denaturing urea-polyacrylamide gels, exposed overnight to Phosphor Storage screen and scanned using the Typhon (see section 2.1.4 of Material and Methods). Initial trials were done testing five divalent metal ions (Mg^{2+}, Mn^{2+}, Ca^{2+}, Zn^{2+} and Co^{2+}) at a concentration of 2.5 mM (Figure 3.4C). A 40 nt single stranded RNA was used as substrate (lane1). RNA degradation was only observed in the presence of Mg^{2+} (lane 3) or Mn^{2+} (lane 4). Incubation of the protein without addition of any metal ion also showed slight RNA degradation (lane 2), however in a much lower extent. This suggested that some contaminants contributing for the background noise might be presence. RNA degradation in the presence of Ca^{2+} (lane 5), Zn^{2+} (lane 6) or Co^{2+} (lane 7), was not as robust as in the presence of Mg^{2+} or Mn^{2+}. Similarly to lane 2, the presence of some contaminant might have contributed for RNA degradation observed in the presence of these three metal ions (Ca^{2+}, Zn^{2+} or Co^{2+}) rather than from hNEF-sp activity.

To further explore hNEF-sp dependency of Mg^{2+} and Mn^{2+}, we repeated the assay in the presence of chelating agent EDTA (Figure 3.4D). We observed that the RNA substrate remained intact in the presence of only Mg^{2+} (lane 2 and 3), Mn^{2+} (lane 4 and 5) or hNEF-sp (lane 6 and 12). However, the substrate was degraded when the RNA probe was incubated with hNEF-sp in the presence of 2.5 mM of Mg^{2+} (lane 7 and 8) or 2.5 mM Mn^{2+} (lane 9 and 10). Incubation with 10 mM EDTA was not sufficient to abolish hNEF-sp activity (lane 11) suggesting that the concentration of chelating agent should be higher. We also observed that hNEF-sp nuclease activity was more robust in the presence of Mn^{2+} than in the presence of Mg^{2+}. We decided to test if different Mg^{2+} concentrations would enhance hNEF-sp activity (Figure 3.4E). Increasing the Mg^{2+} concentration to 5 mM (lane 4) resulted in a similar activity to the one observed in the previous assays. However, the increase to 25 mM (lane 5) resulted in a more robust activity, similar to the one observed in the presence of 2.5 mM Mn^{2+}. An Mg^{2+} concentration of 50 mM

(lane 6) has some inhibitory effect and 100 mM (lane 7) and 150 mM (lane 8) completely abolished hNEF-sp activity.

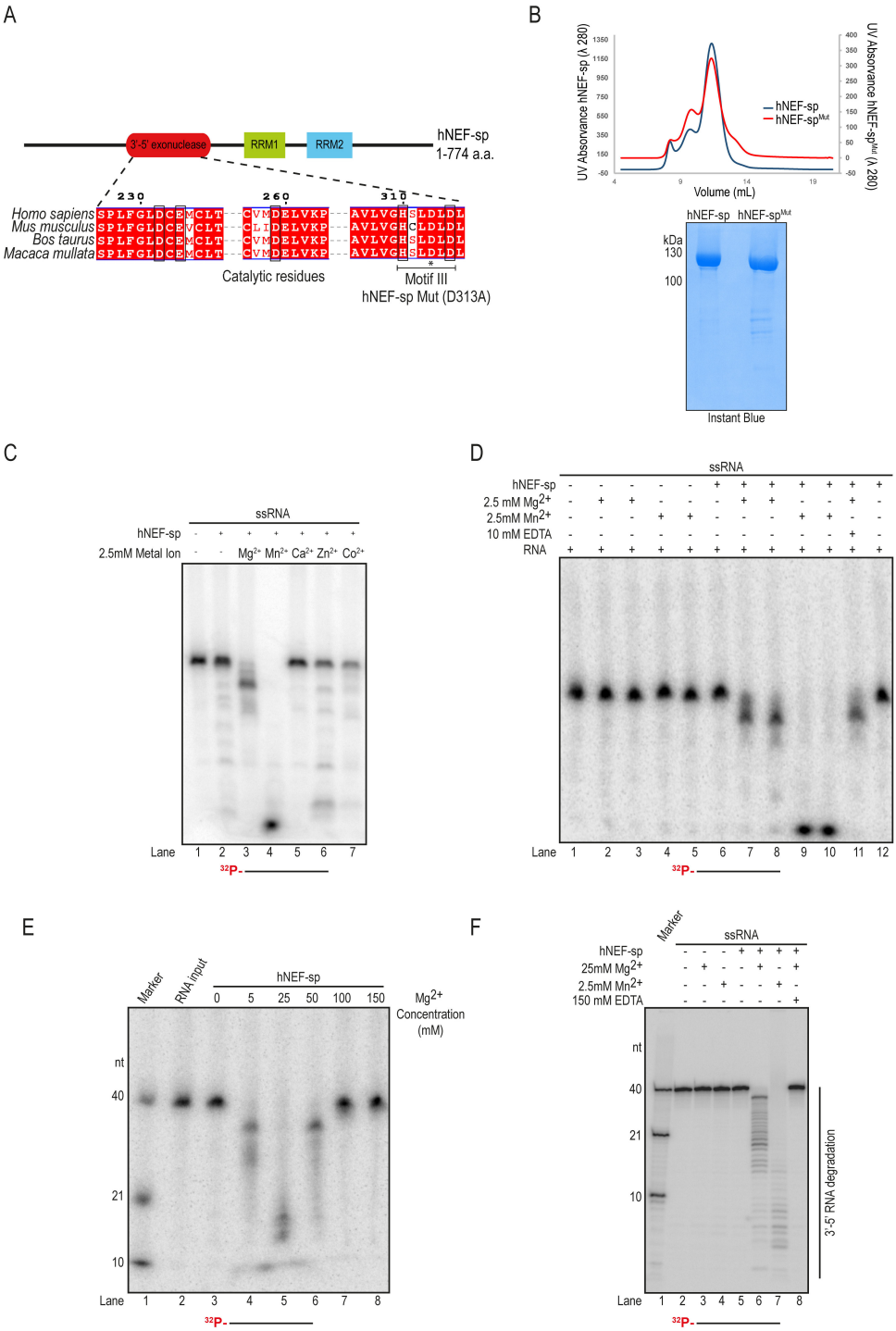


Figure 3.4 – Optimization of human NEF-sp nuclease assay. (A) Schematic representation of the domain architecture of hNEF-sp and alignment of the exonuclease domain of NEF-sp homologs. The four catalytic residues important for activity are indicated (black rectangles). The mutation (D313A) introduced in the motif III is also represented (*). (B) SDS-PAGE gel after purification of hNEF-sp and hNEF-sp^{Mut}. Size exclusion chromatography profiles for the two proteins are shown. (C) Nuclease assay testing hNEF-sp activity in the presence of different metal ions. (D) Activity test using Mg^{2+} or Mn^{2+} and the chelating agent EDTA. (E) Titration of different Mg^{2+} concentrations. (F) Nuclease assay using the conditions established.

Having most of the conditions optimized we decided to improve the results quality by introducing some extra steps in the protocol. Instead of loading the samples directly on the Urea-PAGE gels, which might affect the gel quality due to the presence of protein, salts and other contaminates, we started by doing Proteinase K treatment followed by phenol-chloroform extraction and RNA precipitation. After resuspending the precipitated RNA in water and adding formamide loading dye, samples were loaded in the Urea-PAGE gels and handled as described previously (see section 2.1.4 of Material and Methods). We observed that the quality of the results was indeed improved. We were able to detect a clear degradation pattern in the presence of Mg^{2+} (lane 6) or Mn^{2+} (lane 7) (Figure 3.4F). In addition, the degradation observed (RNA ladder) indicates that hNEF-sp has 3' to 5' end directionality. The increase of EDTA concentration to 150 mM was sufficient to completely abolish hNEF-sp activity by sequestration of the metal ion Mg^{2+} (lane 8).

NEF-sp has four catalytic residues that are essential for nuclease activity. To confirm that the nuclease activity observed was specific from hNEF-sp, we created a point mutation in one of the aspartic acid residues of motif III (D313A). We tested hNEF-sp^{Mut} ability to degrade RNA. As previously, we observed that hNEF-sp (lane 3) is able of trimming the RNA substrate (Figure 3.5A). However, we did not observe any RNA degradation (lane 4) after incubating the substrate with hNEF-sp^{Mut}. These results further strength that the nuclease activity observed is specific from hNEF-sp activity. To further characterize hNEF-sp, we tested if a minimal RNA length is required to be processed (Figure 3.5B). We incubated hNEF-sp with three single stranded RNA oligonucleotides: 40 nt (lane 3), 21 nt (lane 4) or 10 nt RNA (lane 5). The degradation efficiency was similar, independently of RNA length. The shortest processed product observed was 8-7 nt (lane 5) (Figure 3.5B).

We also evaluated hNEF-sp activity towards double stranded RNA oligonucleotides. A 20 nt dsRNA was used as substrate and the ssRNA sense oligonucleotide as control (Figure 3.5C). We observed that hNEF-sp is able to degrade ssRNA substrates (lane 2) as shown previously. However, dsRNA substrates could not be cleaved by hNEF-sp (lane 4). DNA oligonucleotides were also tested as a putative hNEF-sp substrate (Figure 3.5D). Single stranded DNA oligonucleotides (lane 2) were degraded while the double stranded remained intact (lane 4). We observed some degradation of double-stranded oligonucleotides for the higher size substrates (Figure 3.5C and 3.5D). This result might be due to technical difficulties in purifying RNA and DNA duplexes from the gel. Therefore, some of the isolated substrates might have 1 to 2 nucleotide overhangs that can be cleaved by hNEF-sp. To complete hNEF-sp nuclease activity characterization, we tested if the presence of 3' end modifications could protect the RNA from degradation as is proposed for the 2'-*O*-methyl modification (Huang *et al.* 2009). We observed that the substrate carrying this modification remained intact (lane 4) when compared with the same RNA with no 3' end modifications (lane 2) (Figure 3.5E).

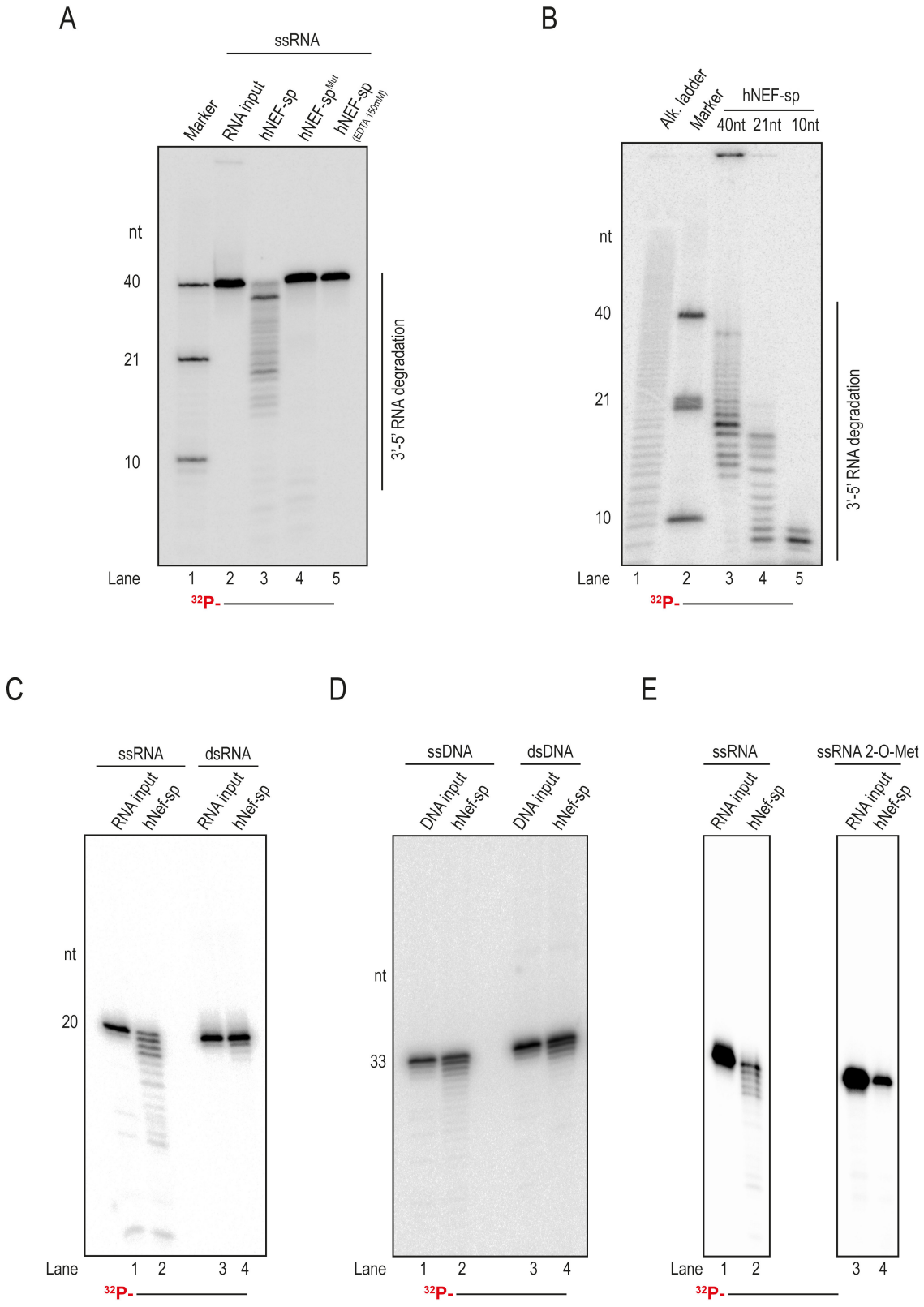


Figure 3.5 – human NEF-sp nuclease activity in different conditions. (A) Nuclease assays using hNEF-sp and hNEF-sp^{Mut}. Activity of hNEF-sp using different substrates: (B) different RNA length, (C) dsRNA, (D) ssDNA and dsDNA and (E) ssRNA with 3' end modifications.

3.2.4 Crystallization trials of human NEF-sp

Crystal structures can provide a great insight on protein function and in understanding its mechanisms of action. Domain predictions reveal that human NEF-sp is composed by an N-terminal nuclease domain followed by two C-terminal tandem RNA-recognition motifs (Figure 3.6A). As shown previously full-length human NEF was highly expressed in insect cells. However removal of the fusion tag was quite inefficient (Figure 3.3D). This might indicate that TEV binding site might be hindering by the protein folding. We used a bioinformatic tool to assess the folding of hNEF-sp (Prilusky *et al.* 2005). Analysis of the folding prediction revealed that the first 42 residues at N-terminal and that some residues belonging to the second RRM might be unfolded (Figure 3.6A). We were interested in obtaining the crystal structure of human NEF with the three domains (nuclease domains and the two RRMs). Therefore, we expressed in insect cells a construct in which the first 41 N-terminal residues were deleted (hNEF-sp deletion 1). Similar purification strategy as described in the section 3.2.2 was followed. The N-terminal 6xHis-Strep-SUMO tag was successful removed by TEV cleavage overnight (see section 2.1.4 of Material and Methods). After all the purification steps we confirmed that protein purity was suitable for structural studies (Figure 3.6B). We set 6 standard crystallization screens previously optimized by the High Throughput Crystallization (HTX) lab at EMBL Grenoble. The screens were made using the sitting drop method and two protein concentrations were tested, 4.0 $mg.mL^{-1}$ and 8.0 $mg.mL^{-1}$. No hits were obtained.

We decided to optimize certain conditions. It was previously described that 50 % of the proteins with a melting temperature of 45 °C or above have a higher chance of being crystallized (Dupeux *et al.* 2011). We analysed the thermal stability of our protein. hNEF-sp deletion 1 in the purification buffer initially used (25 mM Tris pH 7.5, 250 mM NaCl, 5 mM of β-mercaptoethanol, 10 % of glycerol) had a melting temperature of 35 °C (Figure 3.6C red curve). For that reason, we tried to improve protein stability by optimizing the buffer conditions. We screened 96 different conditions ranging from: buffer concentration of 0.01 M to 0.25 M, pH 4 to 10 and NaCl concentration of 0 to 1 M (table 2.5 of Material and Methods). The highest melting temperature obtained was 42 °C (Figure 3.6C green curve) and the buffer composition was 100 mM Citric Acid pH 6.0 and 250 mM NaCl. A new crystallization trial was set up using hNEF-sp deletion 1 purified in the new buffer condition (25 mM Citric Acid pH 6.0, 250 mM NaCl, 5 mM β-mercaptoethanol and 10 % glycerol). Higher protein concentrations (10 $mg.mL^{-1}$ and 15 $mg.mL^{-1}$) were also used. The melting temperature in these new conditions was re-evaluated and we verified that it had increased to 47 °C.

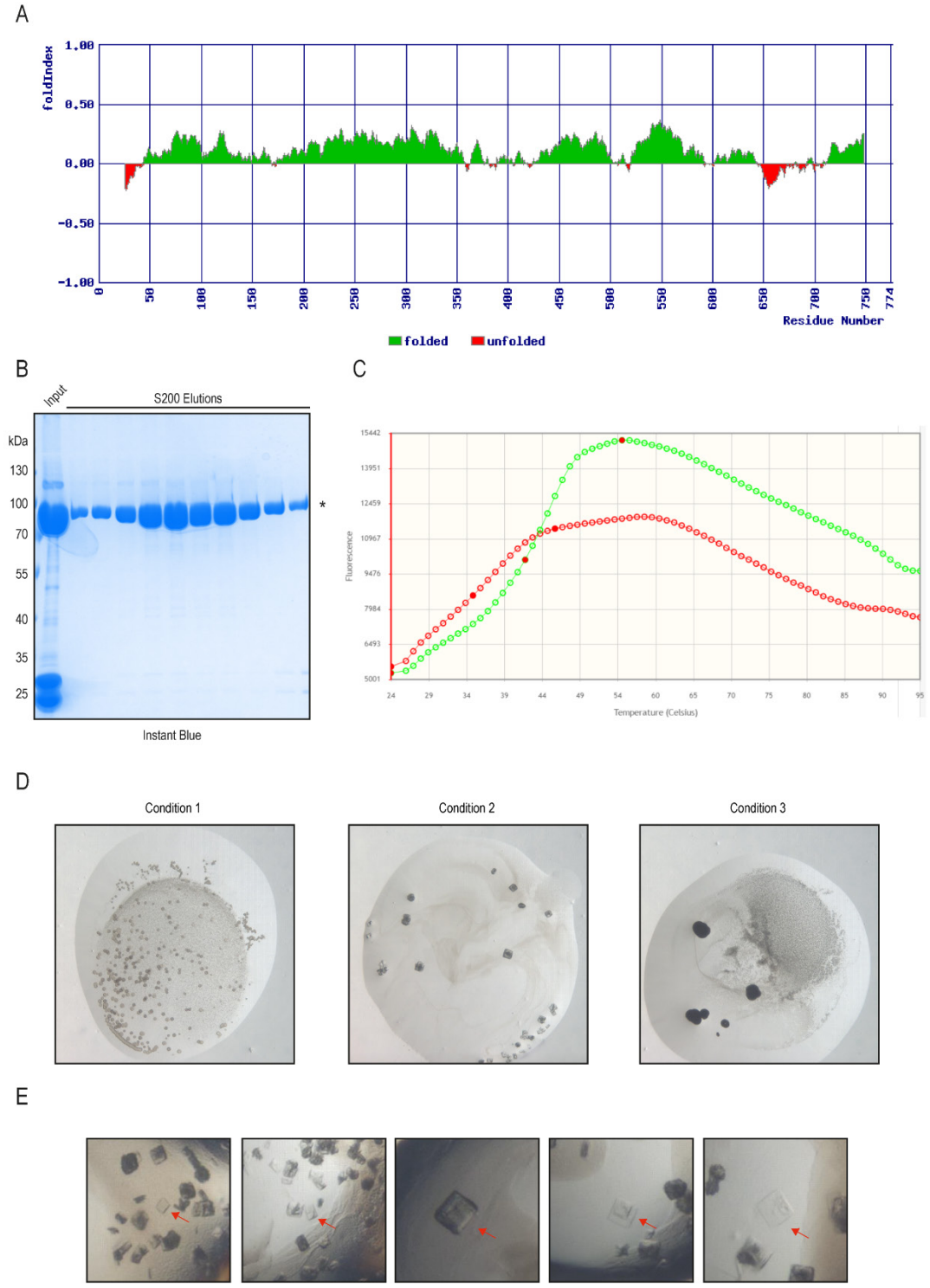


Figure 3.6 –human NEF-sp crystallization trials. (A) human NEF-sp folding prediction. (B) SDS-PAGE gel stained with Instant Blue after size exclusion chromatography of hNEF-sp deletion 1 (*asterisk indicates the band corresponding to human NEF-sp). (C) Melting temperature of hNEF-sp deletion 1 before buffer optimization (red curve) and in the best buffer condition (green curve). (D) Putative crystal hits obtained from the standard crystallization screens. (E) Crystals harvested for diffraction test after optimizing the growing conditions. Red arrows indicate the crystals precisely harvested.

From the crystallization screening using hNEF-sp deletion 1 purified in the new conditions, we obtained hits in three different conditions: condition 1 (0.2 M calcium chloride, 0.1 M HEPES sodium salt pH 7.5 and 28 % (v/v) PEG 400), condition 2 (0.1 M cadmium chloride, 0.1 M sodium acetate pH 4.6 and 30 %(v/v) PEG 400) and condition 3 (0.1 M Tris pH 7.0, 0.2 M Calcium acetate and 20% (w/v) Polyethylene glycol 3000) (Figure 3.6D). We observed some liquid-liquid phase separation in the first condition. While in the second condition some crystals appeared. In the third condition, precipitation was visible in the crystallization drops. We decided to refine the drop conditions by making small changes in precipitant concentration, buffer concentration and salt concentration. We obtained some hits and the most promising crystals were harvested and tested for diffraction in the beamline at the European Synchrotron Radiation Facility (ESRF Grenoble, France). Diffraction pattern revealed that crystals were salt.

hNEF-sp has three domains that may putatively bind to nucleic acids (nuclease domain and the two RRMs). RNA oligonucleotides may help in the crystallization of NEF-sp by creating a stable RNP (ribonucleoprotein) complex. Therefore, we incubated hNEF-sp deletion 1 with an 8 nt poly-U RNA at a ratio of 1 μM protein: 1.2 μM RNA. Drops were set up at a protein concentration of 10 $mg.mL^{-1}$. No hits were obtained from this crystallization trial.

Since we were unsuccessful in crystallizing hNEF-sp deletion 1, we tried to identify a more stable domain by limited proteolysis. We tested trypsin and chymotrypsin at three ratios (1:10, 1:100 and 1:1000 protein:protease) for 45 min at 4 °C. We observed that digestion was more robust with trypsin (Figure 3.7A) and wider size ranges were obtained. Therefore, we decided to use trypsin in a 1:100 ratio and digestion at 4 °C was tested at four time points: 45 min, 60 min, 90 min and overnight. The digestion patterns obtained in the three first time points were similar. After digestion overnight, we obtained three main fragments, two of about 25 – 40 kDa and a larger fragment higher than 55 kDa (Figure 3.7B). We decided to identify these bands due to the fact of being highly resistant to trypsin and therefore more probable of representing a stable construct ideal for crystallization. Fragments were analysed by LC-TOF mass spectrometry (IBS Grenoble). The most significant fragments are schematically represented (Figure 3.7C). To test the expression and solubility of the new constructs, we overexpressed the proteins with an N-terminal 6xHis-SUMO tag in *E. coli*. We checked if any of these were soluble (in the supernatant) or if instead they accumulate in the inclusion bodies (pellet). From the four constructs tested, hNEF-A and hNEF-D clearly accumulated in the inclusion bodies. hNEF-B and hNEF-C seem to be present in cytoplasmic fraction and pellet (Figure 3.7D). We performed Ni-NTA affinity purification to check if we could purify the different constructs from the soluble fraction. We confirmed that hNEF-A and hNEF-D expression levels are low and not suitable for structural studies (Figure 3.7E). However, the expression level of hNEF-B and hNEF-C were high (Figure 3.7E). Since hNEF-B contains the nuclease domain and the two RRMs, we decided to focus in this construct. To further purify hNEF-B, we tried to cleave the 6xHis-SUMO fusion tag by TEV protease digestion. We were unsuccessful in removing the tag, probably

because the TEV binding site was hindered inside of the protein (Figure 3.7E). We tried another construct where the 6xHis-SUMO fusion tag was placed at C-terminal (hNEF-B2). hNEF-B2 was well expressed, however the tag could not be removed by TEV cleavage (Figure 3.7F). After all the unsuccessful approaches we decided to drop the crystallization of human NEF-sp.

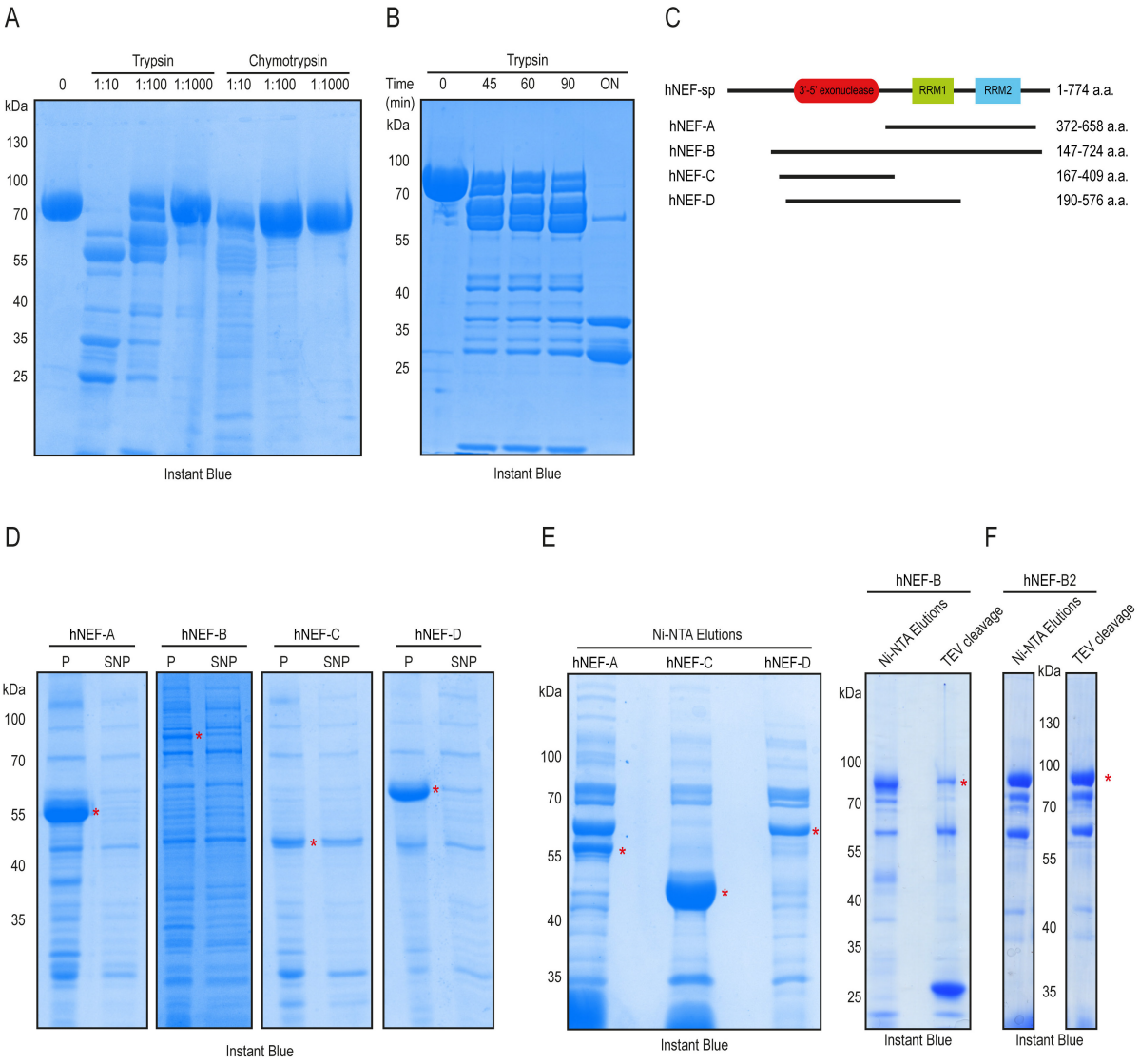


Figure 3.7 – Limited proteolysis of human NEF-sp deletion 1. (A) Limited proteolysis using trypsin and chymotrypsin. (B) Time course digestion using trypsin. (C) Schematic representation of the fragments obtained by overnight trypsin digestion and identified by LC-TOF mass spectrometry. (D) Expression test of the four constructs identified. (E) Ni-NTA purification of hNEF-A, hNEF-C hNEF-D. hNEF-B was also purified by Ni-NTA purification followed by TEV cleavage overnight as well as (F) hNEF-B2. *red asterisk indicates the band corresponding to the different human NEF-sp constructs.

3.2.5 Human NEF-sp is a nuclear protein

Subcellular localization is important for proper protein function and interaction with putative partners and targets. Since many specific processes take place in a particular cellular localization, knowing where NEF-sp localizes could help us to understand its function. For instance, piRNA factors can be found in different granules depending on stage of spermatogenesis. MILI localizes in the inter-mitochondrial cement together with other piRNAs factors in the initial stages of spermatogenesis. However, in late meiotic spermatocytes and haploid spermatids, MILI together with MIWI, localizes in the chromatoid body. MIWI2 shuttles between the nucleus and the piP-bodies (Aravin *et al.* 2009).

To study human NEF-sp and mouse NEF-sp localization we took advantage of human HeLa cell cultures available in the lab. We transiently expressed HA-tagged NEF-sp in HeLa cells. Protein was detected by immunofluorescence using anti-HA antibody and anti-mouse-IgG coupled Alexa Fluor 488 as a secondary antibody (see section 2.1.5 of Material and Methods).

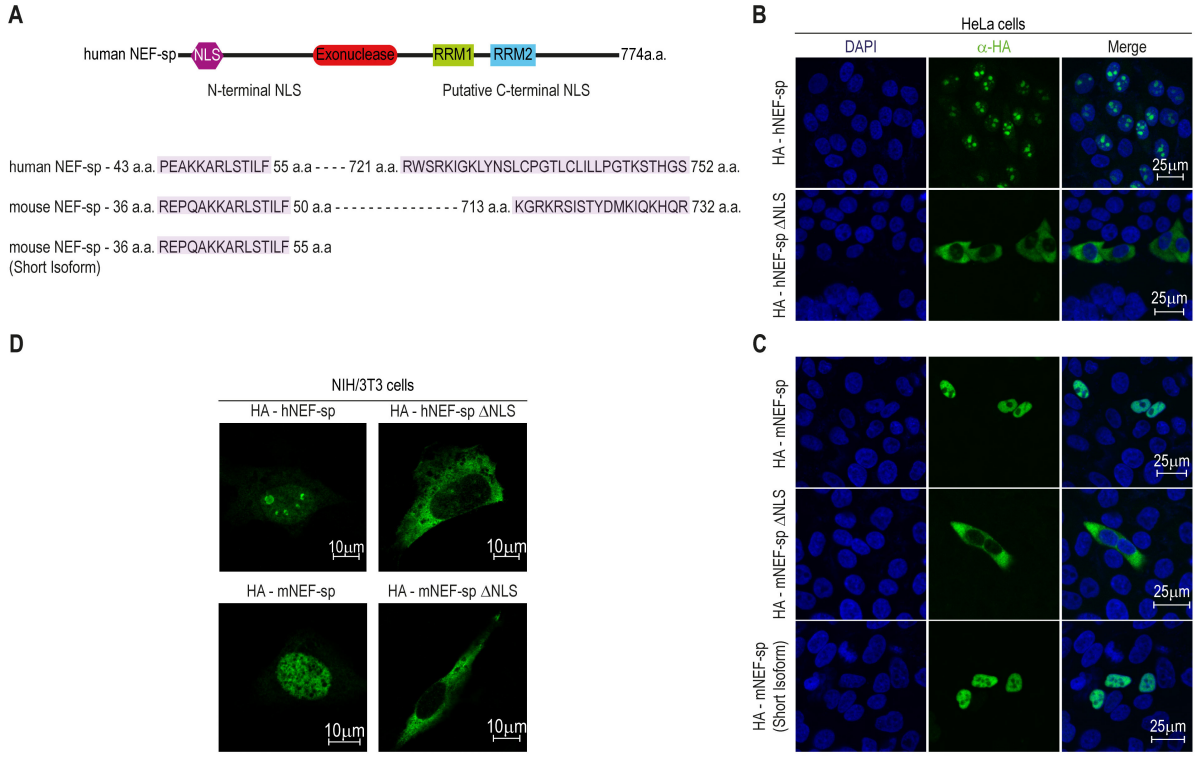


Figure 3.8 – Localization of human and mouse NEF-sp in HeLa cell cultures. (A) Domain organization of human NEF-sp and predicted nuclear localization signals. (B) Immunofluorescence analysis of HeLa cell expressing HA-tagged NEF-sp. DNA was detected by DAPI staining (blue) and the transfected proteins using α-HA antibody and α-mouse-IgG coupled Alexa Fluor 488 (green). Merged signals are also represented. Scale bar is shown in µm. (C) Immunofluorescence analysis of NIH/3T3 cells expressing HA-tagged NEF-sp. Transfected proteins were detected using α-HA antibody and α-mouse-IgG coupled Alexa Fluor 488 (green). Scale bar is shown in µm.

hNEF-sp efficiently accumulated in the nucleolus of transfected cells (Figure 3.8B). mNEF-sp also showed nuclear localization, however it was clearly excluded from the nucleoli. This might be due to an incompatibility with species-specific factors that facilitate efficient nucleolar accumulation. Therefore, we tested localization of mNEF-sp in mouse NIH3T3 cells. Mouse protein still accumulated in the nucleus while being excluded from the nucleolus pointing to species-specific differences (Figure 3.8C).

Since NEF-sp localizes in the nuclear compartment, we tried to identify a putative nuclear localization signal (NLS) using an appropriated software (Kosugi *et al.* 2009). Computational search using hNEF-sp and mNEF-sp sequences revealed the presence of two potential NLS, one at N-terminus and another at C-terminus (Figure 3.8A). The putative N-terminal NLS was the one predicted with highest score. We deleted this domain and analysed if localization of NEF-sp would be compromised. Two constructs were created: hNEF-sp$^{\Delta NLS}$ (58–774 aa) and mNEF-sp$^{\Delta NLS}$ (53–784 aa). Both were transfected in the HeLa cells and detected by immunofluorescence as described before. We observed that deletion of the potential N-terminal NLS excluded hNEF-sp and mNEF-sp from the nucleus and these were retained in the cytoplasm (Figure 3.8B). The short isoform of mNEF-sp lacking the predicted C-terminal NLS was still efficiently imported into the nucleus, supporting the observation that the N-terminal signal is the one critical for nuclear accumulation. Our attempts to examine endogenous NEF-sp in mouse testes were unsuccessful (discussed in the section 3.2.6).

3.2.6 α-mNEF-sp antibody production and validation

To aid the analysis of the *in vivo* role of NEF-sp, we raised polyclonal antibodies against mouse NEF-sp. We expressed five constructs containing a unique sequence of mNEF-sp to guarantee specific recognition of our protein (Figure 3.9A). The different constructs were expressed in *E. coli* with an N-terminal 6xHis fusion tag. Antigen solubility was tested and all the constructs were insoluble except antigen 4 that seemed not to be expressed (Figure 3.9B). We purified the antigens from the inclusion bodies under denaturing conditions (see section 2.1.6 of Material and Methods). Since the levels of purity obtained were not the most suitable for antibody production (Figure 3.9C), we further purify them by Ni-NTA affinity purification (Figure 3.9D). After this purification step we obtained highly pure antigens. The identity of the bands was analysed by mass spectrometry analysis at Proteomics Core Facility, EMBL Heidelberg. The band observed in antigen 4 purification corresponds to antigen 5. This might probably be due to cross contamination during the purification process. The same contamination was observed in antigen 2. Therefore, we decide to send for injection only antigen 3 and antigen 5 due to its highest amount and purity. These allowed us to produce antibodies against the N-terminal and the C-terminal region of mouse NEF-sp. Antigens were injected in two different rabbits. Serum from immunized animals was collected along the process. For experimental purposes, we used mainly the final bleed since it has the highest amount of antibody. The two antibodies will be referred as α-mNEF-3 and α-mNEF-5 produced with antigen 3 and antigen 5 respectively.

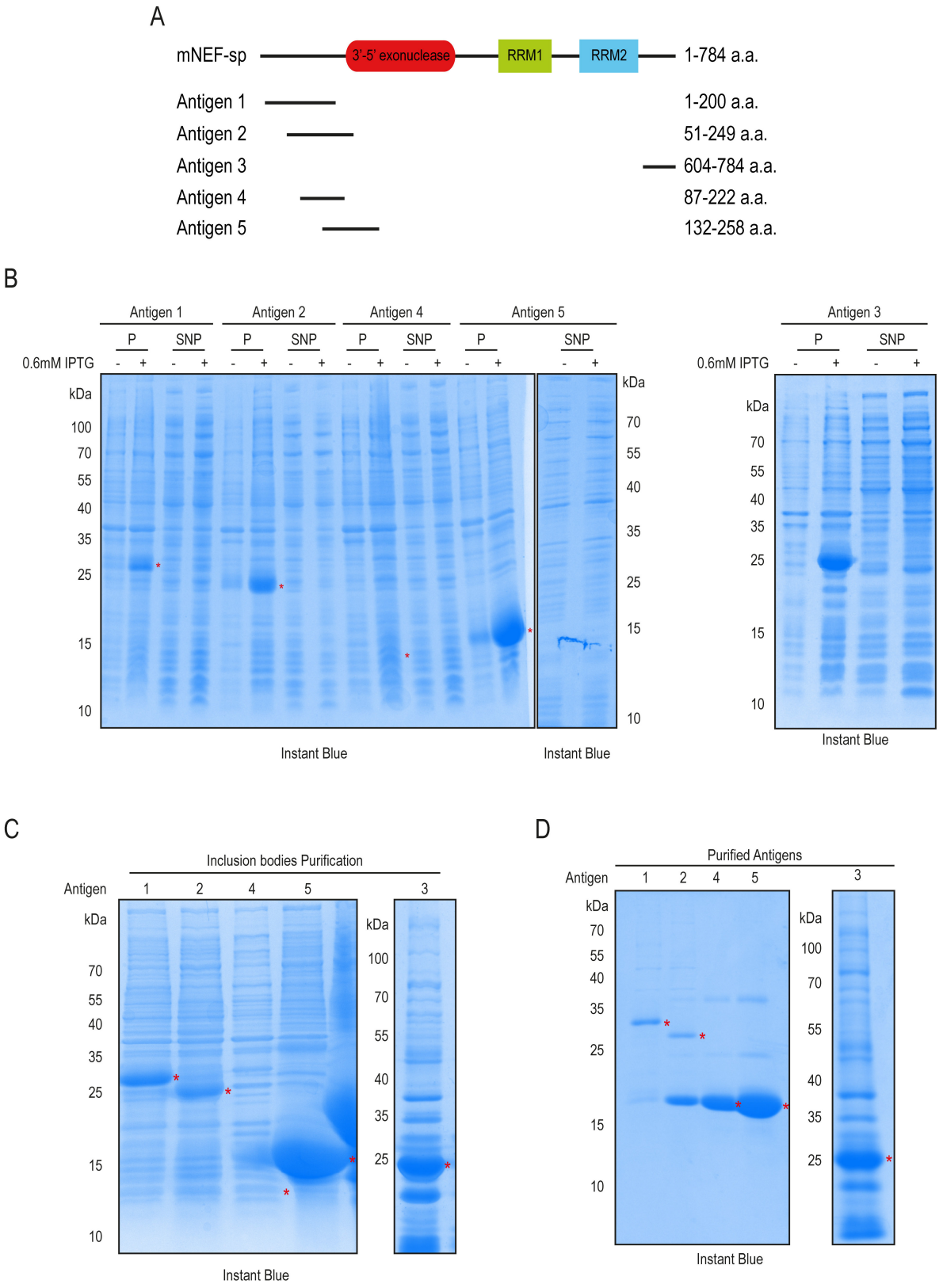


Figure 3.9 – Antigen expression and purification for production of α-mNEF-sp antibody. (A) Cartoon representing the five antigens. (B) Analysis by SDS-PAGE of the solubility of the antigens expressed in *E. coli*. (C) Purification from the inclusion bodies. (D) Ni-NTA purification under denaturing conditions. *red asterisk indicates the band corresponding to the different antigens.

The ability of α-mNEF-3 and α-mNEF-5 to recognize mNEF-sp was tested. The antibodies were used to detect HA-mNEF-sp transfected in HEK293T cells by western blot. Both α-mNEF-3 and α-mNEF-5 antibodies were able to recognize mNEF-sp in HEK293T transfected cells lysate (Figure 3.10A and Figure 3.10B). α-HA antibody was used as a control. Afterwards, we tested if the raised antibodies could recognize the protein in its native environment. As shown in section 3.2.1 mNEF-sp is exclusively expressed in testis. We did a western blot using total testis lysate and probed with α-mNEF-3 and α-mNEF-5 antibodies. We were not able to detect mNEF-sp using α-mNEF-5 antibody (Figure3.10B). However, α-mNEF-3 antibody detected a band that migrates similarly to what could potentially correspond to mNEF-sp (89 kDa) (Figure 3.10C). We decided to confirm this result by taking advantage of having *Nef-sp*$^{-/-}$ testis (mouse knock-out is described in section 3.2.7). We observed the same pattern in the three different conditions, *Nef-sp*$^{+/+}$, *Nef-sp*$^{+/-}$ and *Nef-sp*$^{-/-}$ indicating that the bands detected do not correspond to mNEF-sp and are background signal (Figure 3.10D). We hypothesized that the levels of mNEF-sp might be low and not detectable by western blot of total testis lysate. Therefore, we performed immunoprecipitation α-mNEF-3. We could not detect any protein in the size corresponding to mNEF-sp (Figure 3.10E).

We also tested if α-mNEF-3 antibody could be used for immunofluorescence. HeLa cells transfected with mNEF-sp and mNEF-sp$^{\Delta NLS}$ were probed using α-mNEF-3 antibody and anti-rabbit-IgG coupled Alexa Fluor 488 as a secondary antibody. We could observe that α-mNEF-3 antibody could recognize mNEF-sp and mNEF-sp$^{\Delta NLS}$ in transfected cells. As shown before, the full length construct localizes in the nucleus and the construct with NLS deletion accumulates in the cytoplasm. However, the antibody also detects some background in the control HeLa cells (Figure 3.9F). The α-mNEF-3 antibody was tested in testis section of adult *Nef-sp*$^{+/+}$ and *Nef-sp*$^{-/-}$ mice (Figure 3.9G). Signal was detected in wild-type and knock-out testis. Since NEF-sp is not expressed in the knock-out mouse, the signal detected is background. Therefore, α-mNEF-3 antibody is not suitable to detect mNEF-sp in its native environment. We also tried to detect endogenous NEF-sp using α-mNEF-3 and α-mNEF-5 purified from the serum. However, the results obtained were similar to the ones shown above.

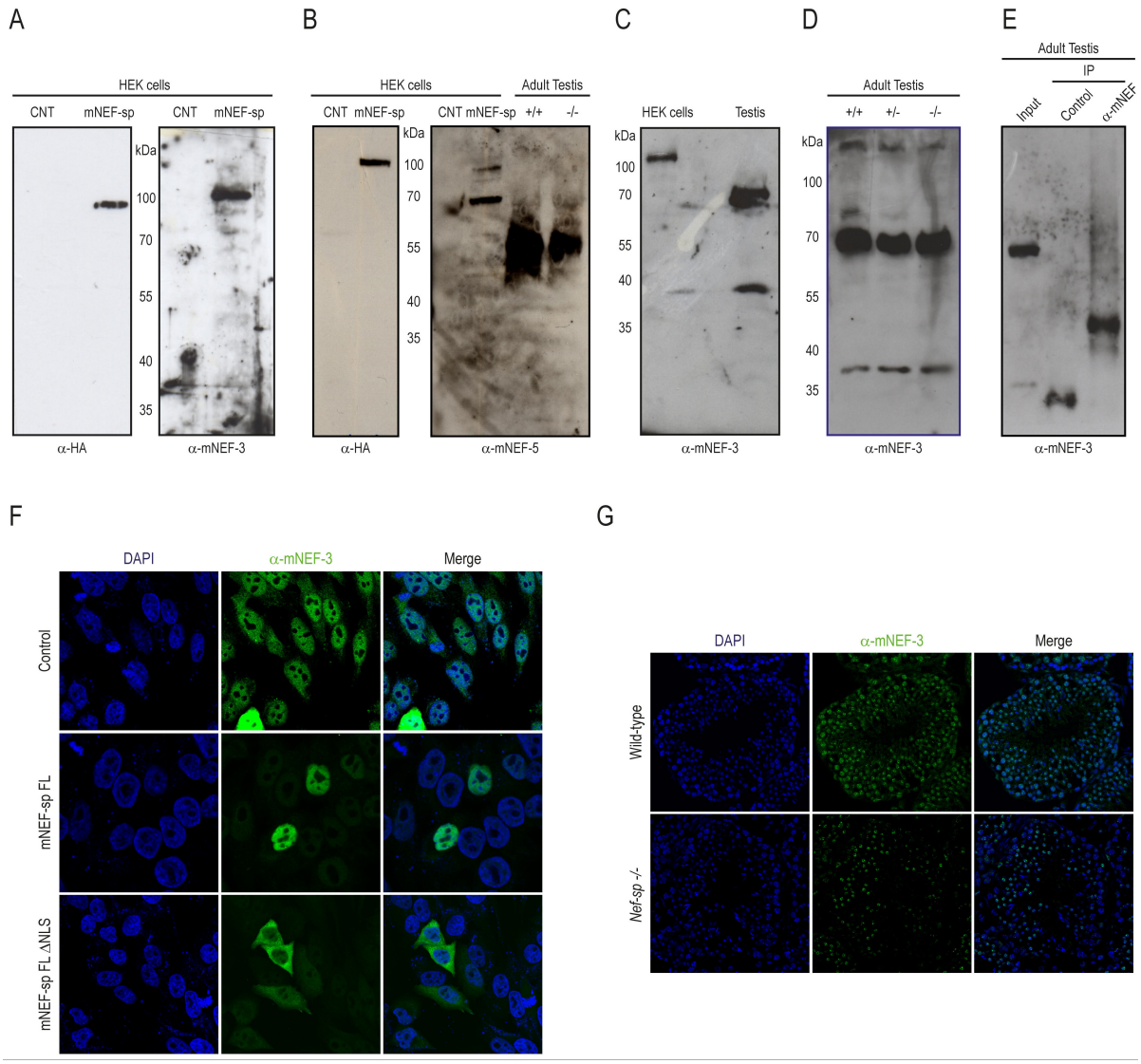


Figure 3.10 – Validation of the two antibodies produced: α-mNEF-3 and α-mNEF-5. Western blot testing the produced antibodies in different conditions: (A) HEK293T cells transfected with HA-mNEF-sp. α-HA was used as a control. (B) HEK293T transfected with HA-mNEF-sp and *Nef-sp*$^{+/+}$and *Nef-sp*$^{-/-}$ adult testis lysate (C) Adult testis total lysate and (D) *Nef-sp*$^{+/+}$, *Nef-sp*$^{+/-}$ and *Nef-sp*$^{-/-}$ adult testis lysate. (E) Immunoprecipitation of NEF-sp from adult testis using α-mNEF-sp-3 antibody and detection by western blot. (F) Immunofluorescence of HeLa cells transfected with mNEF-sp and mNEF-sp$^{\Delta NLS}$ and probed with α-mNEF-3. (G) Immunohistochemistry of cryosections from adult mouse testis of *Nef-sp*$^{+/-}$ and *Nef-sp*$^{-/-}$. α-Rabbit antibody coupled with Alexa 488 (green signal) was used as a secondary antibody. The nucleus was detected by staining DNA with DAPI (blue signal). Merged signals are also represented.

3.2.7 *Nef-sp*$^{-/-}$ mouse generation and phenotype characterization

To understand the *in vivo* function of NEF-sp, we created a knock-out mouse using the CRISPR-Cas9 system. A pair of guide RNAs (gRNAs) was used to target exon 6 of *Nef-sp* loci in single-cell mouse embryos (Figure 3.11A-B). We obtained four different alleles in which *Nef-sp* was successfully targeted (see section 2.1.7 of Material and Methods). We decided to use the founder carrying a deletion of 23 base-pairs in *Nef-sp* locus. This deletion generates a premature stop codon (D220*) upstream of the nuclease domain (Figure 3.11C). The mRNA transcribed from this locus will be recognized by the nonsense-mediated mRNA decay (NMD) machinery and the aberrant transcript will be target for degradation. Therefore, expression of NEF-sp will be abolished. We will refer this allele as *Nef-sp* knock-out allele (*Nef-sp*$^{-}$).

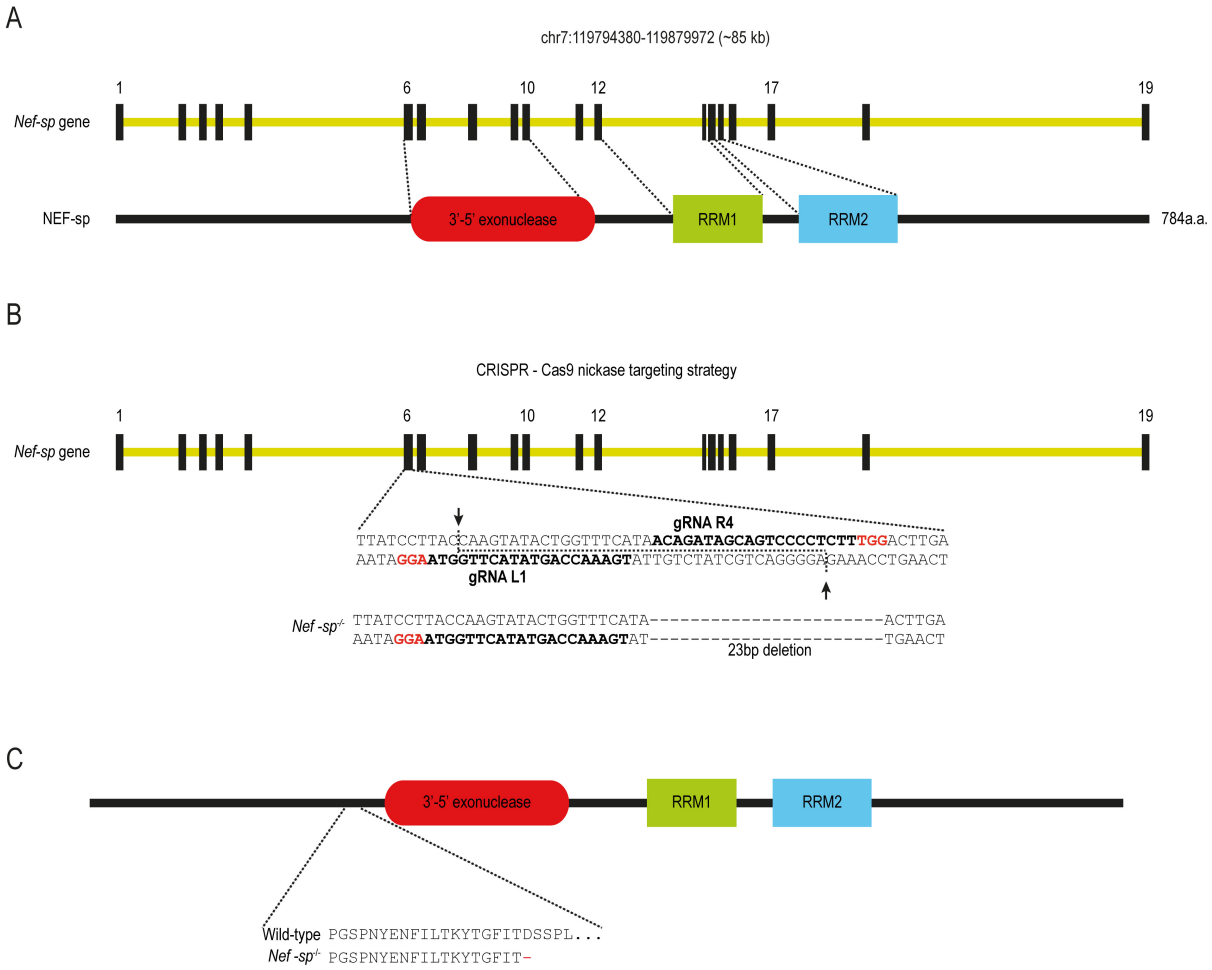


Figure 3.11 – Generation of *Nef-sp* knock-out mouse using the CRISPR-Cas9 system. (A) Representation of the different *Nef-sp* exons and protein domain architecture. (B) Representation of the gRNAs used for generation of *Nef-sp*$^{-/-}$ knock-out mouse and the 23 bp deletion obtained and used throughout this study. (C) Protein prediction after deletion of the 23 bp in *Nef-sp* locus and representation of the premature stop codon (red -) upstream the nuclease domain.

Our initial hypothesis was that NEF-sp might have a role in the piRNA pathway. Therefore, we decided to analyse some features that are common to most of the piRNA factors knock-out mice. In mice, deletion of genes codifying for proteins involved in the piRNA pathway mainly results in male infertility and transposon activation. We decided to test the fertility of *Nef-sp*$^{-/-}$ animals. We observed that homozygous animals of both sexes are viable and display normal fertility. We collected testes from adult (P90) *Nef-sp*$^{+/-}$ and *Nef-sp*$^{-/-}$ and compared their size (Figure 3.12A). Several studies demonstrated that deletion of piRNAs factor results in reduced testes size. However, *Nef-sp*$^{-/-}$ testis had similar size to the heterozygous mouse. We also evaluated transposon activation in the knock-out testis by detecting LINE-1 ORF1 by western blot. *Nef-sp*$^{-/-}$ did not reveal any LINE-1 activation (Figure 3.12 B). The levels observed are comparable to the wild-type and therefore they are not significant. We also tried to detect LINE-1 activation by immunofluorescence. However, we could not detect any LINE-1 expression in cryosections of *Nef-sp*$^{-/-}$ testis (Figure 3.12 C). Adult testes of Vasa DQ mice known to have LINE-1 activation were used as a positive control (Wenda *et al.* 2017).

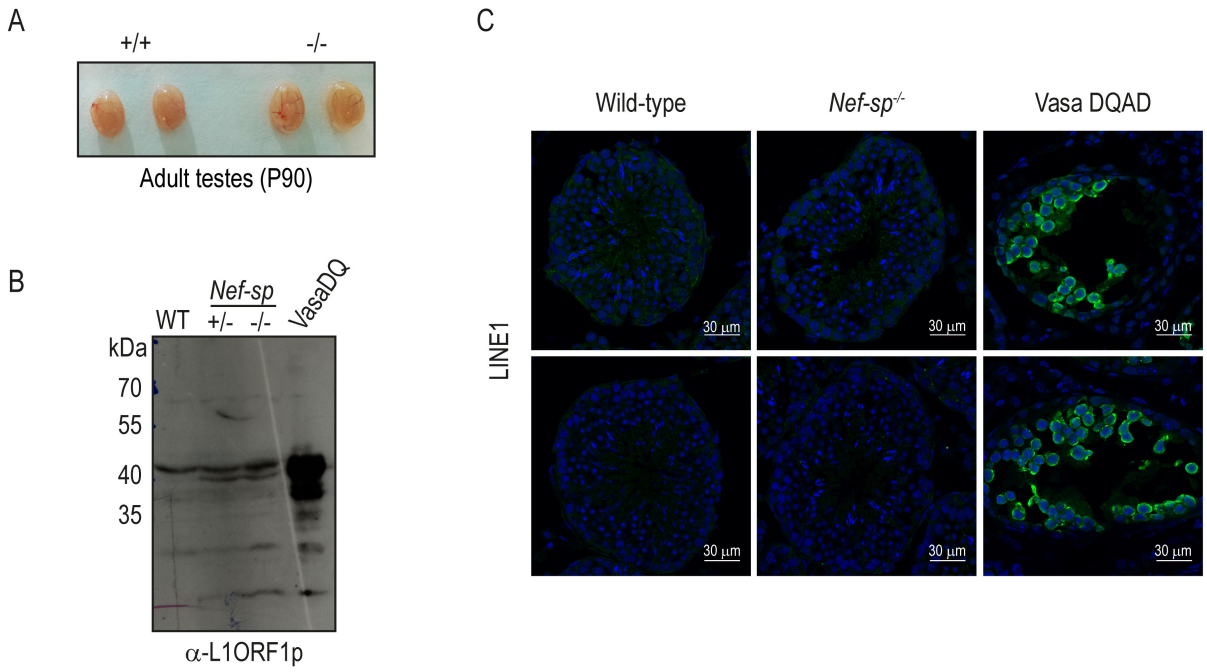
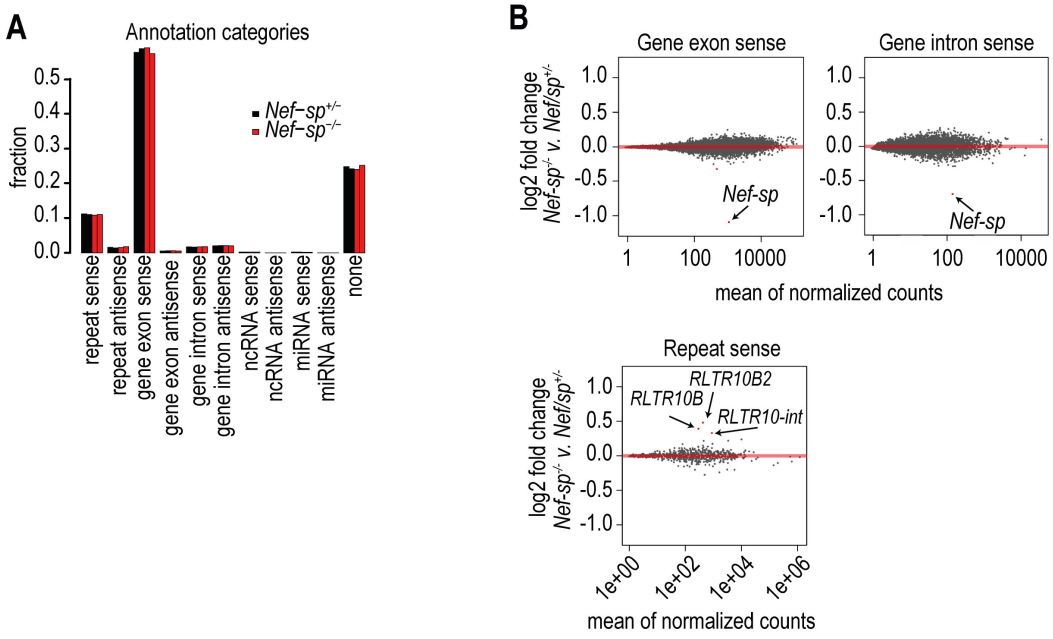


Figure 3.12 – Characterization of the adult testis of *Nef-sp* knock-out mouse. (A) Adult testis (P90) of wild-type and *Nef-sp*$^{-/-}$ mice. (B) Detection of LINE-1 by western blot in adult testis lysate. (C) Immunofluorescence for detection of LINE-1 in adult testis sections of wild-type and *Nef-sp*$^{-/-}$ mice. Vasa DQ testis lysate was used as a positive control (Wenda *et al.* 2017).

Our results have demonstrated that NEF-sp is an active 3' - 5' exonuclease (described in section 3.2.3). Therefore, we decided to analyse if NEF-sp deletion had any effect in testes transcriptome. We deep sequenced testicular RNAs depleted of ribosomal RNA from wildtype and homozygous knock-out mice. Analysis of the transcripts revealed unchanged levels in the different annotation categories (Figure 3.13A). The only transcript that was significantly affected in the mutant was *Nef-sp* transcript (Figure 3.13B). *Nef-sp* mRNA was 25 % lower in the *Nef-sp*$^{-/-}$ testis. To be certain that the 23 base-pairs deletion is also present at transcript level, we extracted total RNA from adult *Nef-sp*$^{-/-}$ testis and after RT-PCR we sequenced the region flanking the putative deletion in exon 6. We confirmed that the 23 bp deletion is present in *Nef-sp*$^{-/-}$ transcripts. Besides the changes in *Nef-sp* transcript, we also observed a slight increase of repeat-associated transcripts in the knock-out testis. RLTR10B, RLTR10B2 and RLTR10-int transposon sequences were significantly increased in the *Nef-sp*$^{-/-}$ mice (Figure 3.13B-C). No other changes were detected in the sequenced transcriptomes.



C

repeat sense

	baseMean	log2FoldChange	p value	p value - adj
RLTR10B2.LTR.ERVK	422.75	0.48	2.67E-07	2.67E-07
RLTR10B.LTR.ERVK	293.35	0.39	4.15E-05	4.15E-05
RLTR10-int.LTR.ERVK	881.78	0.33	1.37E-04	1.37E-04

gene exon sense

	baseMean	log2FoldChange	p value	p value - adj
2610020H08Rik	1078.44	-1.09	1.81E-59	2.25E-55
Klk1b21	469.60	-0.33	7.96E-07	4.94E-03

gene intron sense

	baseMean	log2FoldChange	p value	p value - adj
2610020H08Rik	140.58	-0.70	8.79E-06	5.81E-02

Figure 3.13 – Transcriptome analysis of adult *Nef-sp* $^{-/-}$ testis. (A) Transcriptome analysis of adult tesits of *Nef-sp*$^{+/-}$ and *Nef-sp*$^{-/-}$ mice. (B) Comparison of read counts of sense-oriented gene exonic and intronic reads. The significantly changed transcripts are indicated. (C) Table showing fold-changes in read counts of sense-oriented repeat elements and *Nef-sp* transcript (2610020H08Rik corresponds to *Nef-sp*).

3.3 Discussion and perspectives

Throughout this project we have examined the biochemical properties of an uncharacterized mammalian nuclease called NEF-sp. We have shown that *Nef-sp* transcript is exclusively expressed in mouse testis.

Our *in vitro* experiments have demonstrated that recombinant human NEF-sp is able to degrade RNA. NEF-sp belongs to the superfamily of DEDD nucleases, which requires divalent cations for activity (Y. Zuo *et al.* 2001). Members of this family degrades RNA/DNA by a two-metal-ion mechanism (Steitz *et al.* 1993). Accordingly, we have shown that NEF-sp activity is dependent on divalent metals Mg^{2+} or Mn^{2+}. We also observed that RNA degradation was more robust in the presence of Mn^{2+} than in the presence of Mg^{2+}. This might be due to the relaxed coordination properties of Mn^{2+} compared with the stringent coordination geometry requirement of Mg^{2+} (Yang *et al.* 2006). The four catalytic residues (DEDD) are known to be essential for nuclease activity of members of this family (Y. Zuo *et al.* 2001). Mutation of an aspartic acid residue to an alanine residue (D313A) abolished the enzyme activity, confirming that the nuclease activity we have observed is inherent to hNEF-sp. The elution profile obtained from gel filtration chromatography of hNEF-sp^{Mut} is comparable to the wild-type. Therefore, mutation of D313A did not grossly affect the structural stability of the protein. However, this aspartic acid residue (D313) that we mutated does not belong to the canonical catalytic tetrad that is predicted based on sequence alignments. Instead, D313 is located in the motif III close to the predicted catalytic aspartic acid residue D315 (the last D in DEDD motif). Given the loss of catalytic activity in the mutant protein, we propose that NEF-sp catalytic pocket might be slightly different from other DEDD nucleases and D313 might also participate in the coordination of divalent metals. Caf1, a nuclease involved in poly(A) tail degradation, belongs to the DEDD superfamily and contains the canonical residues essential for activity (Shimizu-Yoshida *et al.* 1999, Viswanathan *et al.* 2004). However, the catalytic tetrad of the yeast homolog Pop2 is non-conserved (S-E-D-Q). Nevertheless, Pop2 is still an active nuclease and the serine residue was shown to be essential for nuclease activity (Daugeron *et al.* 2001, Thore *et al.* 2003). Similarly, NEF-sp might also have a non-canonical catalytic pocket. Another possibility is that the H-x(2)-D sequence might be present in motif III instead of the typical H-x(4)-D sequence (Y. Zuo *et al.* 2001). Alternatively, the mutation we made (D313) might have caused sufficient alteration to the structure to affect coordination geometry of the catalytic residues for the metal ions or and, affect the cleavage reaction. Since we did not obtain hNEF-sp crystal structure, we cannot confirm that D313 is indeed one of the catalytic residues.

We further examined the specificity of hNEF-sp activity by testing different nucleic acid substrates. The accessibility of the catalytic pocket to single-stranded RNAs was examined using substrates with different sizes. NEF-sp was active towards RNAs of 40 nt, 21 nt and 10 nt and these were digested down to 2-3 nt. We also observed an RNA ladder pattern of degradation indicating that NEF-sp has a distributive 3' → 5' hydrolytic activity. The distributive mechanism implies that NEF-sp repeatedly binds to the substrate, hydrolysis one of the nucleotides and releases again the substrate. The cycle continues until the RNA

reaches a minimal length or it is completely degraded. This mechanism of RNA degradation is similar to the one described for Rrp6, a subunit of the eukaryotic RNA exosome (Wasmuth *et al.* 2014). We have also shown that NEF-sp cannot accommodate RNA duplexes within the catalytic pocket. DNA was also a poor substrate for the enzyme in both its single-stranded and double-stranded forms. Finally, 2'-*O*-methyl modification on the 3' terminal nucleotide of RNA, which acts as a stabilizing mark by protecting them from exonucleases (Ji *et al.* 2012), hinders the activity of the enzyme.

The importance of the two RNA-recognition motifs (RRMs) for hNEF-sp nuclease activity was not examined in this study. These may be responsible for the selection of the RNA targets. The two tandem RRMs might fold together increasing RNA binding affinity and sequence-specificity (Deo *et al.* 1999). The selected targets might then be fed into the nuclease domain of NEF-sp. However, in our *in vitro* nuclease assays, we have used a wide variety of RNAs which were all efficiently processed. We cannot exclude that the RRMs might be involved in protein-protein interactions as was previously observed for other proteins (Kadlec *et al.* 2004).

We have demonstrated that NEF-sp localizes in the nucleus of HeLa transfected cells. Our efforts to detect the endogenous protein in testes sections failed. Since our rabbit polyclonal antibodies detect transiently expressed HA-mNEF-sp in human cell cultures, we conclude that perhaps abundance of the endogenous protein is below the detection limit of our antibodies. To support our observation of nuclear localization, we tried to predict a nuclear localization signal (NLS). We were able to identify and verify that the N-terminus of NEF-sp contains a NLS. The software used predict the NLS (Kosugi *et al.* 2009) identified in mouse NEF-sp and in human NEF-sp the following amino acid sequences REPQAKKARLSTILF and PEAKKARLSTILF respectively. NEF-sp contains a canonical monopartite NLS composed by a single cluster of three basic amino acids with the consensus sequence of K–K/R–X–K/R (X represents any amino acid) (Chelsky *et al.* 1989). Typically, proteins containing this type of NLS are imported to the nucleus by a ternary complex composed of importin β1, importin α and the cargo protein. The recognition of the NLS is mediated by importin α (Kosugi *et al.* 2008). Contrary to mouse NEF-sp, the human orthologue specifically accumulates in the nucleolus. Localization experiments using the mouse cell line NIH/3T3 revealed that only human NEF-sp has this specific nucleolar localization, pointing to species-specific differences. Nucleolar localization might be due to sequestration of the target protein by specific factors such as rDNA, rRNA or interactions with other proteins (Carmo-Fonseca *et al.* 2000). However, some studies suggested that specific nucleolar localization sequences (NoLSs) might be responsible for the accumulation in the nucleolus (Scott *et al.* 2010). Prediction of NoLSs using a suitable software (Scott *et al.* 2011), identified two motifs in human NEF-sp. Interestingly, these motifs are absent in mouse NEF-sp. Thus, the different nuclear localization observed between mouse and human homologs might be due to the presence/absence of these NoLSs in human and mouse NEF-sp respectively.

Analysis of *Nef-sp* mouse knockout mutant revealed no obvious phenotype. Our initial hypothesis was that NEF-sp could play a role in piRNA 3' end maturation. Deletion of factors implicated in the mouse piRNA pathway leads to transposons activation, arrest of gametogenesis and sterility in males mice (Siomi *et al.* 2011). However, *Nef-sp* knock-out mice of both sexes were shown to be fertile. Adult testes of *Nef-sp* knockout do not show activation of transposons. While my study was in progress, the protein responsible for piRNA 3' end shortening was identified in the BmN4 cells (Izumi *et al.* 2016). PNLDC1 is the 3'-5' exonuclease responsible for trimming pre-piRNAs to its mature size in the Bombyx system. This protein is also conserved in mice and humans indicating that this might be the Trimmer in these organisms.

Analysis of adult testis transcriptome revealed no major differences between the wild-type and the knock-out. Only *Nef-sp* transcript itself was shown to be downregulated. This decrease correlates with the presence of the premature stop codon (D220*) and rapid degradation of the mutated transcript by NMD pathway {Kervestin, 2012 #282}. Some repeat transcripts RLTR10B, RLTR10B2 and RLTR10-int transposon were also upregulated in the mutant. RLTR10Bs are the youngest retrotransposon elements in the mouse genome and they might have retained the ability to be expressed (Giordano *et al.* 2007). However, activation of these of transposons was not sufficient to have an effect on spermatogenesis and on mouse fertility. The lack of phenotypic changes upon NEF-sp deletion suggests that its function in the testis might be redundant. NEF-sp might be important for precise trimming of the unknown substrate(s) and complementation by other nucleases results in no obvious phenotype. Similar situation exists in the budding yeast *Saccharomyces cerevisiae*, where RNA exonuclease 1 (Rex1p)/Rnh70p/yGR276 bears similarity to the nuclease domain of NEF-sp (van Hoof *et al.* 2000). Rex1p is a 3'-5' exonuclease involved in processing the 3' end of 5S rRNA, $tRNA_{Arg}$ and $tRNA_{Met}$ (van Hoof *et al.* 2000, Ozanick *et al.* 2009). Unprocessed 5S rRNA or tRNA precursors tend to accumulate in the Rex1 mutant yeast (van Hoof *et al.* 2000). However, maturation mediated by Rexp1 is not essential for viability and the presence of these 3' extended 5S rRNA precursors do not affect maturation of rRNA in ribosomes (Nariai *et al.* 2005). Sufficient mature rRNA and tRNA molecules are produced due to the action of a number of redundant nucleases (van Hoof *et al.* 2000, Ozanick *et al.* 2009). NEF-sp might also be involved in 3' end processing of some specific class of RNA and its activity might be redundant.

Given the easy access to CRISPR-Cas9-mediated genome editing in mouse embryos, one experiment that should be tried in the future is to insert a 3xFLAG-HA tag at the N-terminus of the protein by targeting the endogenous gene locus. This will enable immunofluorescence analysis and also purification of endogenous protein complexes. Another, perhaps more interesting experiment would be making a catalytic-dead mutation in the enzyme, with the hope of trapping processing intermediates that now fail to be trimmed by the enzyme, in vivo. This analysis will not be hampered by redundancy with other enzymes, as the substrates will be bound by the inactive enzyme, but not processed. One may even hope to get a dominant-negative phenotype.

Overall, we have demonstrated that NEF-sp is an active exoribonuclease that acts preferentially on ssRNAs and may function in the nuclear/nucleolar compartment. Our study provides the first biochemical and genetic characterization of the mammalian NEF-sp exoribonuclease.

4. Characterization of METTL4 – a putative 6mA DNA methyltransferase

Résumé

Caractérisation de METTL4 – une hypothétique ADN 6mA méthyltransférase

La méthylation N6-méthyladénine (6mA) de l'ADN a été étudiée pendant de nombreuses années. Cependant, la fonction biologique de la modification de 6mA est restée insaisissable en raison de stratégies de détection limitées. Récemment, le développement de nouvelles techniques a permis une meilleure caractérisation de 6mA dans les eucaryotes. Il a été démontré que le marqueur 6mA est présent dans le génome de certains organismes tels que *Chlamydomonas*, *C. elegans*, *D. melanogaster* et les mammifères.

Certaines des protéines impliquées dans le dépôt et l'élimination du marqueur 6mA ont déjà été identifiées. Dans *D. melanogaster*, DMAD s'est avéré être responsable du dépôt du marqueur. Dans *C. elegans*, NMAD-1 et DAMT-1 ont montré jouer un rôle dans l'élimination et le dépôt du marqueur respectivement. Alkbh1 a également été démontré qu'il était important pour l'effacement de 6mA de l'ADN chez les mammifères.

L'ADN méthyltransférase impliqué dans le dépôt du marqueur 6mA chez les mammifères et les mouches n'était toujours pas caractérisé. On a suggéré que METTL4 était l'ADN méthyltransférase chez les mammifères, mais aucune preuve expérimentale n'a été signalée. METTL4 possède un domaine de méthyltransférase (MTase) à l'extrémité C-terminale et les résidus conservés (DPPW) importants pour l'activité. Tout au long de ce projet, nous avons essayé de comprendre si METTL4 joue un rôle dans le dépôt de 6mA chez les mammifères et les mouches. Nous avons essayé de caractériser biochimiquement et structurellement la souris METTL4. De plus, nous avons analysé son importance biologique en utilisant des souris mutantes et des mouches comme système modèle.

4.1 Introduction

N^6–methyladenine (6mA) DNA methylation has been shown to be present in the genome of prokaryotes and lower eukaryotes. In higher eukaryotes, its presence and biological function remained elusive due to limited detection strategies. Recent development of new techniques allowed a better characterization of 6mA in eukaryotes. It has been shown that 6mA mark is present in the genome of some organisms such *Chlamydomonas, C. elegans, D. melanogaster* and mice (Fu *et al.* 2015, Greer *et al.* 2015, Zhang *et al.* 2015, Koziol *et al.* 2016, Wu *et al.* 2016). Some of the proteins involved in deposition and removal of 6mA mark have already been identified. In *D. melanogaster*, DMAD has been shown to be responsible for the deposition of the mark (Zhang *et al.* 2015). In *C. elegans,* NMAD-1 and DAMT-1 were shown to play a role in the removal and deposition of 6mA respectively (Greer *et al.* 2015). However, the DNA methyltransferase involved in the deposition of 6mA mark in mammals and flies has not been characterized yet.

METTL4 is the closest predicted orthologue of DAMT-1 in mouse and humans. Therefore, METTL4 was proposed to be responsible for the deposition of the 6mA mark in mammals (Greer *et al.* 2015). However, neither its DNA methyltransferase activity nor its biological relevance has been demonstrated experimentally. Domain predictions show that METTL4 has a methyltransferase domain (MTase) at C-terminus. Four catalytic residues (DPPW) essential for the methyltransferase activity are also present. Typically, this type of enzymes uses S-adenosylmethionine as a methyl group donor (Figure 4.1A) (Iyer *et al.* 2011).

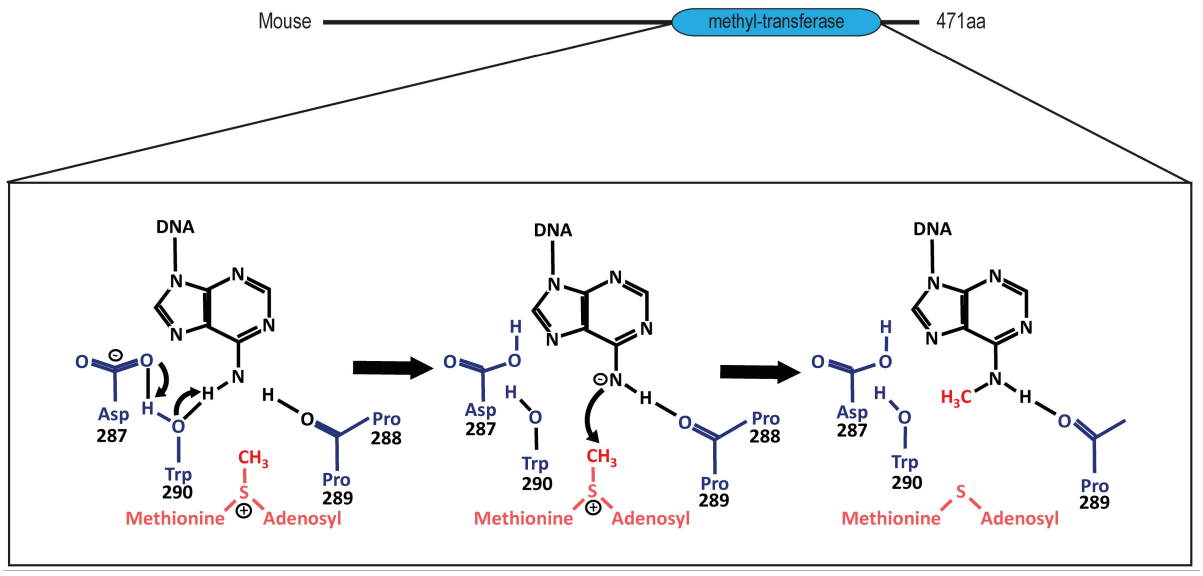


Figure 4.1 – Domain organization of mouse METTL4 and representation of methyltransferase reaction. The catalytic residues (Asp-Pro-Pro-Trp) important for DNA methyltransferase activity are represented in blue and the methyl donor (S-adenosylmethionine) in red. Adenine base is represented in black.

Throughout this project we tried to understand if METTL4 plays a role in 6mA deposition in mammals and flies. We tried to characterize biochemically and structurally mouse METTL4. In addition, we analysed its biological importance by using mutant knockout mouse and flies.

4.2 Results

4.2.1 Recombinant mouse METTL4 expression

We started by expressing recombinant mouse and fly METTL4 in bacteria (*E. coli Rosetta* strain) with an N-terminal 6xHis-SUMO fusion tag. The constructs from both species were however insoluble: DmMETTL4 clearly accumulated in the pellet fraction while mMETTL4 seemed to be poorly expressed (Figure 4.1A). Therefore, we decided to change to the eukaryotic expression system. Insect cells were used to express METTL4 with and N-terminal 6xHis-Strep-SUMO fusion tag. Constructs from two different species were tested: mouse METTL4 and *Bombyx mori* METTL4. These were not expressed and even after Ni-NTA affinity purification we could not detect any enrichment (Figure 4.1B). To confirm that METTL4 was poorly expressed, we tried to detect it by western blot using α-His antibody (Figure 4.1C). Multiple bands were detected and the higher size bands most likely correspond to full length protein. METTL4 expression levels were therefore extremely low and unsuitable for further purification.

We changed our strategy by tagging METTL4 at C-terminal instead of the commonly used N-terminal tag. We presumed that the N-terminal region of the protein might be essential for its folding and tagging METTL4 in this region might be preventing a proper folding. We tested our hypothesis by expressing mouse METTL4 fused to C-terminal SUMO-6xHis tag in bacteria (*Rosetta* strain). Bacteria were grown overnight at two different temperatures 17 °C and 37 °C. We observed that METTL4 accumulated in the pellet when expressed at 37 °C (Figure 4.1D). This was not so clear for the expression trial done at 17 °C. Therefore, a Ni-NTA affinity purification was done to enrich for mMETTL4 (Figure 4.1E). We obtained three different bands around the predicted size of METTL4. Mass spectrometry (MS) analysis revealed that highest size band corresponds to mMETTL4 (Figure 4.1E). Some bacterial proteins were also identified by MS. We decided to not further purify this construct due to METTL4 low expression levels and being contaminated with bacterial proteins.

A final attempt to have soluble mouse METTL4 was done in insect cells using the SUMO-6xHis C-terminal tag construct. METTL4 was soluble and enriched by Ni-NTA purification (Figure 4.1F). We further purified mMETTL4 by Q-column and the purest fraction were pulled together and loaded in a Heparin column (Figure 4.1G) followed by size exclusion chromatography (Figure 4.1H). The pure fractions corresponding to full-length mouse METTL4 were concentrated and used in downstream applications. We confirmed by mass spectrometry the identity of the purified protein.

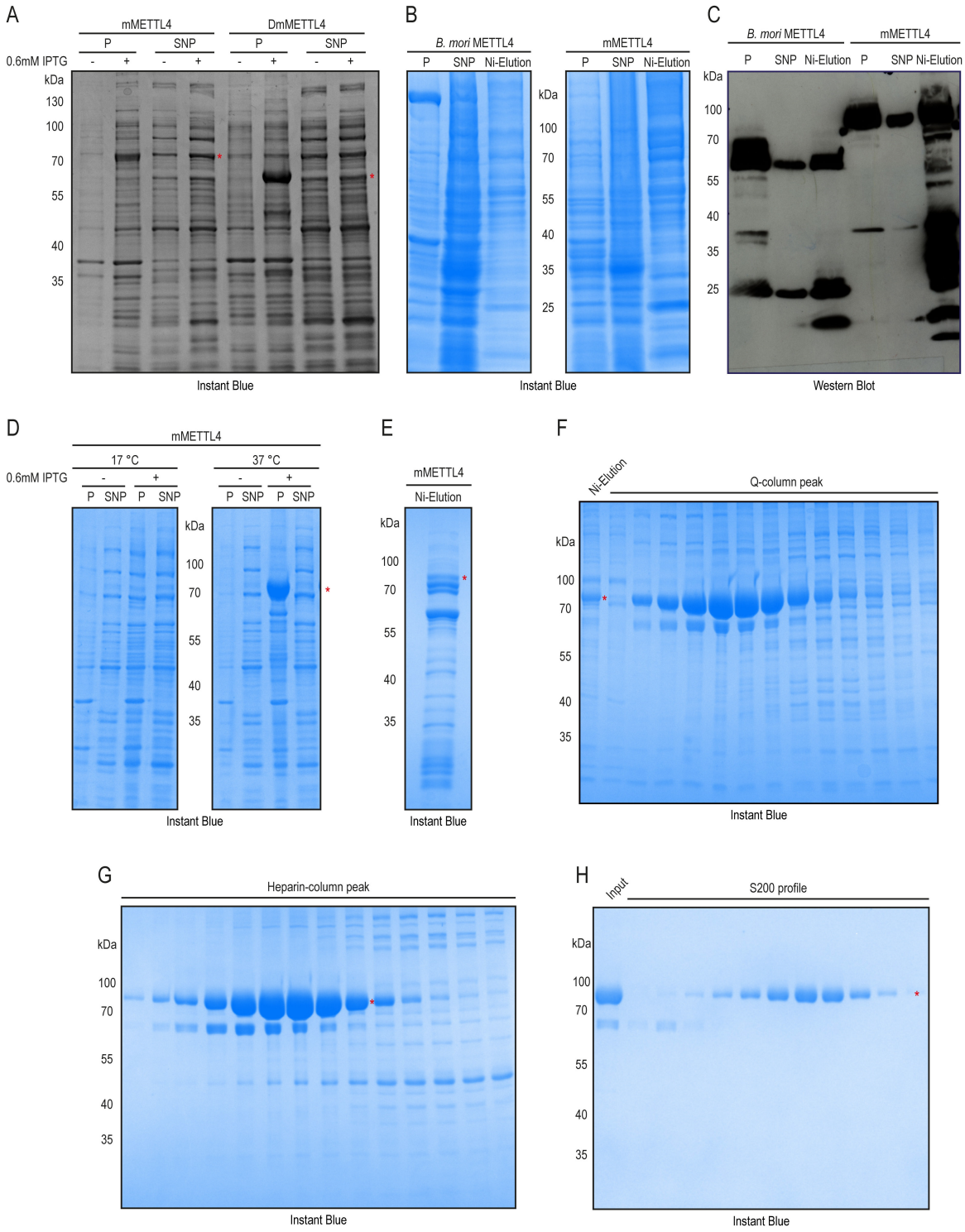


Figure 4.1 – Expression and purification of METTL4. (A) Expression trials in *E. coli* of mouse and *D. melanogaster* METTL4 tagged at N-terminal with a 6xHis-SUMO. (B) Expression of mouse and *B. mori* METTL4 using insect cells. (C) Western blot for detection of 6xHis-Strep-SUMO-METTL4 using α-His antibody and α-rabbit IgG HRP-linked as a secondary antibody. (D) Expression in bacteria of mouse METTL4 tagged at C-terminal. (E) Ni-NTA affinity purification of mouse METTL4 expressed in *E. coli* at 17 °C. (F) Purification by Ni-NTA affinity purification and Q-column of C-terminal tagged METTL4 expressed in insect cells. (G) Heparin purification of mouse METTL4 followed by (H) size exclusion chromatography. * Red asterisk indicates the band corresponding to the different METTL4 constructs tested in the purification trials.

4.2.2 Biochemical characterization of putative methyltransferase activity of mouse METTL4

METTL4 is predicted to be responsible for the deposition of the N6-Methyladenosine (6mA) mark. The protein has the predicted catalytic residues (DPPW) necessary for activity. We tested recombinant mouse METTL4 methyltransferase activity *in vitro*. Methylation assays were performed using wild-type METTL4 and the point mutant (DPPW – APPW) as a control (Figure 4.2A). dsDNA oligonucleotides were labelled at 5' end with [γ-^{32}P] ATP and incubated with mouse METTL4. After incubation, the dsDNA probes were digested with DpnI or DpnII. These two restriction enzymes bind GATC consensus sequence. DpnI cuts only 6mA-methylated DNA and DpnII is inhibited by the presence of the mark (Figure 4.2B). The digestion fragments obtained were run in a 15% Urea denaturing polyacrylamide gels, exposed overnight to Phosphor Storage screen and scanned using the Typhon (see section 2.2.4 of Material and Methods).

Digestion efficiency of the two restriction enzymes was tested (Figure 4.2C). We observed that DpnI (lane 1) is not able to digest the unmethylated oligonucleotide while the DNA probe was digested by DpnII (lane 2). Dam from *E. coli*, which is an active 6mA DNA methyltransferase that binds preferentially to the consensus sequence GATC, was used as a positive control. After incubating the dsDNA with Dam, we digested the DNA probe with the two restriction enzymes. DpnI was able to digest around 50% of the DNA probe (lane 3) indicating that the dsDNA was methylated by the Dam. Digestion with DpnII was inhibited due to the presence of the methylation mark (lane 4). Recombinant mMETTL4 DNA methyltransferase activity was similarly tested. Two different protein concentrations of 1 µg and 5 µg were used. After incubation, DpnI and DpnII were used to digested dsDNA probe as described above. DpnI was unable to digest the dsDNA due to the absence of the 6mA mark (lane 5 and lane 7). Similarly, the dsDNA was cleaved by DpnII (lane 6 and lane 8) suggesting that the methylation mark 6mA was absent. The same results were obtained using mMETTL4^{Mut} (lane 9 to 12) reinforcing the observation that METTL4 is unable to methylate dsDNA in these conditions.

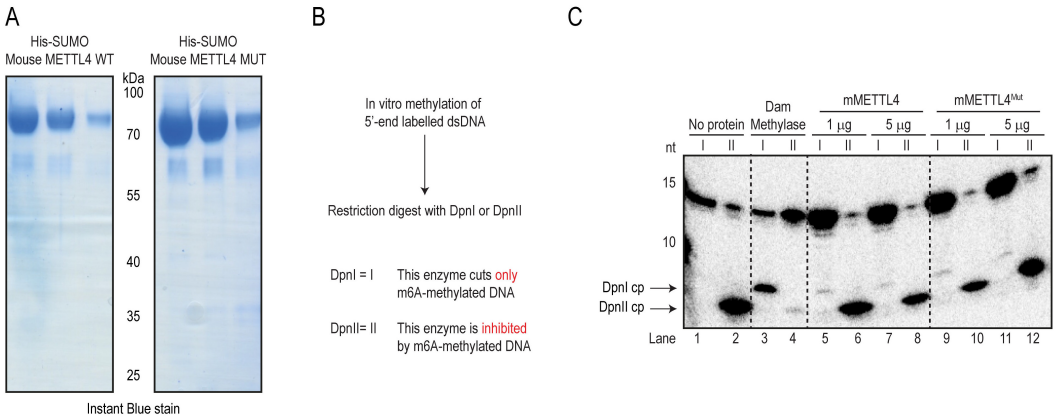


Figure 4.2 – DNA methylation assay using recombinant mouse METTL4. (A) SDS-PAGE gel after purification of mMETTL4 wild-type and catalytic mutant (APPW). (B) Schematic representation of the *in vitro* methylation assay using 6mA methylation sensitive restriction enzymes. (C) Methylation sensitive restriction digestion assay using mMETTL4, mMETTL4^{Mut} and Dam from *E. coli*.

We decided to test mMETTL4 methyltransferase activity under different conditions. Radioactively-labelled SAM-[methyl-^{14}C] was incubated with mMETTL4 and different substrates. Conserved motifs for 6mA deposition were identified in *C. elegans* (GAGG), *Chlamydomonas* (CATG and GATC) and ES cells (TTAGAA). ssDNA and dsDNA probes having these consensus sequences were used to teste mMETTL4 ability to methylate these substrates (Figure 4.3A). We did not observe any mMETTL4 methyltransferase activity using either ssDNA substrates (lane 1 – 4) or dsDNA substrates (lane 5 – 8). Dam methyltransferase from *E. coli* was used as a positive control. Both ssDNA (lane 9) and dsDNA (lane 10 -11) substrates having the required consensus sequence (GATC) were methylated.

We hypothesized that the consensus sequences used in the methylation assay might not the one recognized by mMETTL4. Therefore, we purchased a dsDNA fragment with all the NNANN combinations and we tested if METTL4 could recognize and methylate some of these sites (Figure 4.3B). The assay was again done using ssDNA and dsDNA. mMETTL4 was unable to methylate the ssDNA (lane 1) and dsDNA substrates (lane 1). Dam methyltransferase was able to recognize and methylate its consensus sequence (lane 3). Preference for hemi-methylated DNA substrates was also tested. We used naked (lane 1) and hemi-methylated (lane 2) dsDNA having a consensus sequence (GTAC) obtained from Bind-N-seq experiments (see section 4.2.3). We did not observe any activity (Figure 4.3C). Lastly, we decided to test if mMETTL4 was able to methylate RNA substrates instead of DNA. Seven different RNAs were used, however no mMETTL4 methyltransferase activity was observed (Figure 4.3D). The human complex METTL3-METTL14, available in the lab and previously described as a 6mA RNA methyltransferase (Liu *et al.* 2014), was used as a positive control.

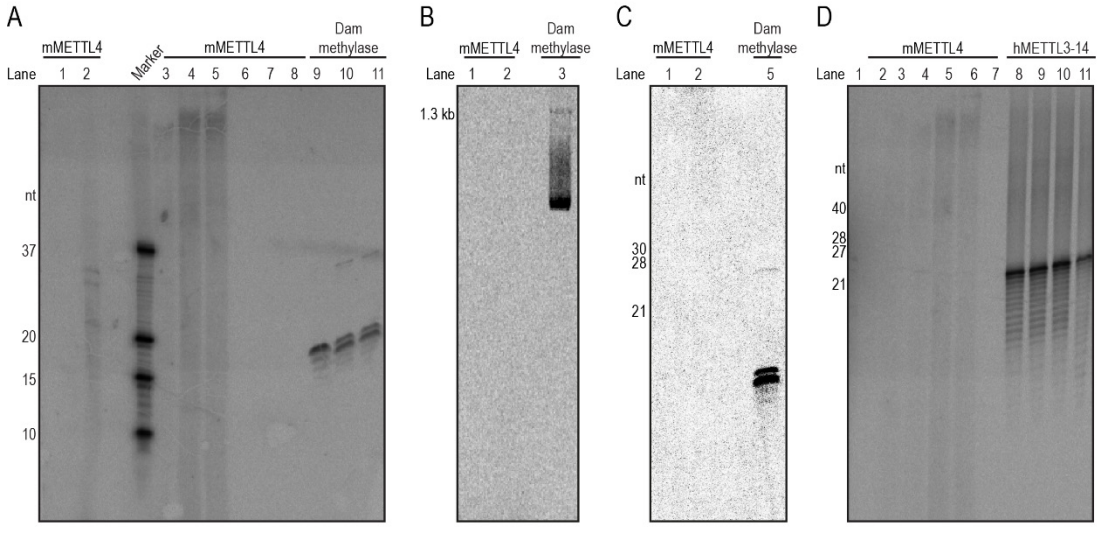


Figure 4.3 – METTL4 DNA methyltransferase activity using different substrates: (A) ssDNA and dsDNA having the consensus sequences of 6mA deposition identified previously in other species. (B) Reporter with all the NNANN possible combinations. (C) dsDNA (naked and hemi-methylated) having the consensus sequence identified by Bind-N-seq experiments. (D) different ssRNA substrates. Dam methyltransferase was used as a positive control in the experiments using DNA and the complex METTL3-METTL14 in the experiments using RNA.

4.2.3 Bind-N-seq and 6mA-IP-seq

Simultaneously with characterization of mMETTL4 methyltransferase activity (described in section 4.2.2), we also tried to determine mMETTL4 DNA-binding consensus sequence. To evaluate METTL4 binding specificity, we used a library of double-stranded DNA oligonucleotides having a 21 nt randomized sequence. The DNA probes were incubated either with purified mMETTL4 or mMETTL4^{Mut}. Samples were treated in two different ways. In one experimental set up, after extensive washes, mMETTL4 bound DNA was recovered and deep sequenced (see section 2.2.5 of Material and Methods). In the second experimental set up, after incubation with mMETTL4 and DNA recovery, the α-6mA antibody was used to pull down the putatively methylated probes (Figure 4.4A).

mMETTL4 and mMETTL4^{Mut} bound DNAs were compared with the control where no protein was used. We obtained five different consensus sequences that METTL4 could potentially bind. However, these results were not statistically significant and therefore mMETTL4 preference for these sequences is unlikely. In the experiments in which 6mA antibody was used to pull-down the methylated DNA, we compared the datasets obtained from mMETTL4 and mMETTL4^{Mut}. Three possible consensus sequences were obtained, however GTAC was the only that had statically significance (Figure 4.4B). This consensus sequence was used in the methylation assays described in section 4.2.2. The results using DNA substrate having this motif were also negative. Dam methyltransferase was used as a technical control. The most significant consensus sequence obtained for Dam methyltransferase was GATC (Figure 4.4C), which is in accordance with previously described data.

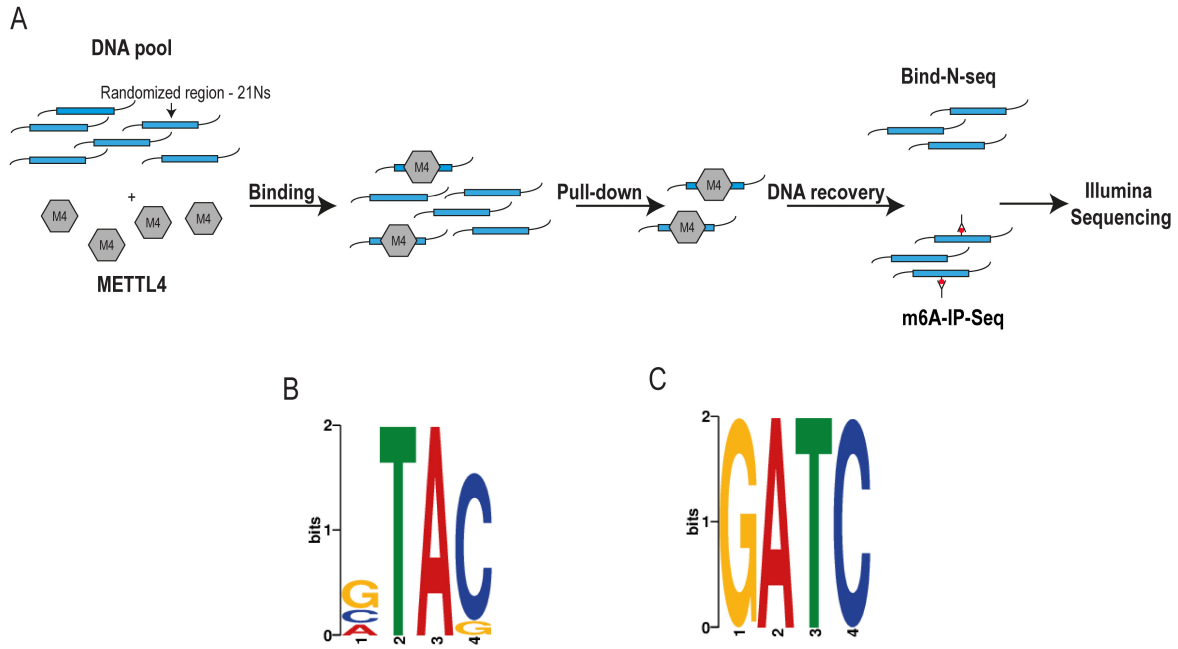


Figure 4.4 – Analysis of DNA-binding consensus sequence of mouse METTL4. (A) Schematic representation of the Bind-N-Seq high-throughput analysis of *in vitro* protein-DNA interactions. (B) Consensus DNA sequence preferentially methylated by mMETTL4. (C) Dam methyltransferase DNA-binding consensus sequence.

4.2.4 Characterization of METTL4 DNA binding activity

METTL4 ability to bind DNA was also tested. We determined protein : DNA interactions by electrophoretic mobility shift assay (EMSA). Purified mMETTL4 was incubated with dsDNA probes of different sizes: 32 bp, 23 bp and 13 bp. Samples were resolved in native-PAGE gels followed by staining with instant blue and ethidium bromide (see section 2.2.6 of Material and Methods). Mouse METTL4 did not bind to the dsDNA substrates tested (Figure 4.5A). dsDNA probes incubated with METTL4 ran similarly to the dsDNA probes without protein suggesting that METTL4 is not able to bind to dsDNA in the tested conditions. Capacity to bind ssDNA having the conserved motifs identified in *C. elegans* (GAGG), in *Chlamydomonas* (CATG and GATC) and in ES cells (TTAGAA) was also tested. METTL4 did not bind any of these DNA substrates.

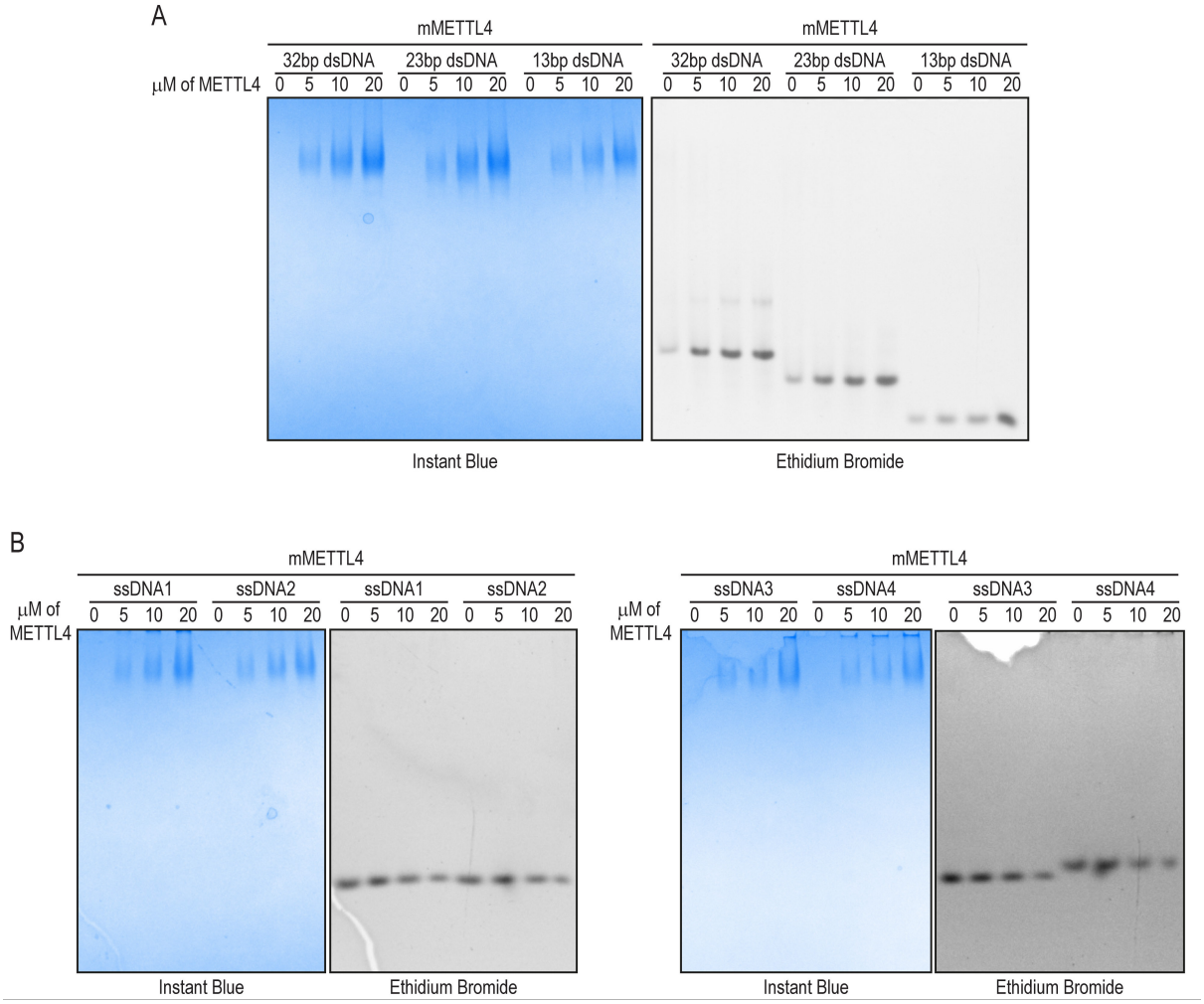


Figure 4.5 – Electrophoretic mobility shift assay using recombinant mMETTL4. (A) Mouse METTL4 was incubated with dsDNA substrates of different sizes. Samples were resolved in native-PAGE gel stained with instant blue and ethidium bromide. (B) ssDNA probes having consensus sequences of 6mA methylation sites identified in other organisms were also tested.

4.2.5 Crystallization trials of mouse METTL4

We tried to determine the crystal structure of mouse METTL4. Protein tagged at C-terminal with SUMO-6xHis and expressed in insect cells (see section 2.2.3 of Material and Methods). After Ni-NTA affinity purification, the tag was removed by TEV cleavage overnight. Afterwards protein was re-incubated with Ni-NTA beads and the collected flow-through was loaded into a Heparin column (Figure 4.6A). The most pure samples were pulled together and mono-dispersed by size exclusion chromatography (Figure 4.6B). Protein was concentrated to 10 $mg.mL^{-1}$ and used in 6 standard crystallization screens provided by the HTX lab at EMBL Grenoble. No hits were obtained in the tested conditions.

Limited proteolysis to identify stable domains was performed using trypsin and chymotrypsin. Digestion was done at 4 °C using different incubation times (Figure 4.6C). We obtained a wide range of size fragments: from 15 kDa to around 70 kDa. Some of digested fragments were analysed by LC-TOF mass spectrometry (EMBL Heidelberg) and the identified boundaries are schematically represented (Figure 4.6D). The new constructs were overexpressed in *E. coli* with a C-terminal SUMO-6xHis tag. Solubility was tested and we observed that mMETTL4-A and mMETTL4-B were insoluble and accumulate in the pellet. mMETTL4-C was present in both pellet and supernatant fractions. mMETTL4-D and mMETTL4-E solubility was inconclusive (Figure 4.6E). Ni-NTA affinity purification was performed in order to check if we could purify the different constructs from the soluble fraction (Figure 4.6F). mMETTL4-C was the most promising construct. Nevertheless, we cleaved the 6xHis-SUMO tag from the five constructs by TEV digestion overnight and re-incubated with Ni-NTA beads. The TEV cleaved fraction and the flow-through from the Ni-NTA were analysed by SDS-PAGE (Figure 4.6G). mMETTL4-C was the only construct that was highly expressed, however we could not remove the tag by TEV cleavage. None of these constructs were not suitable for crystallization trials.

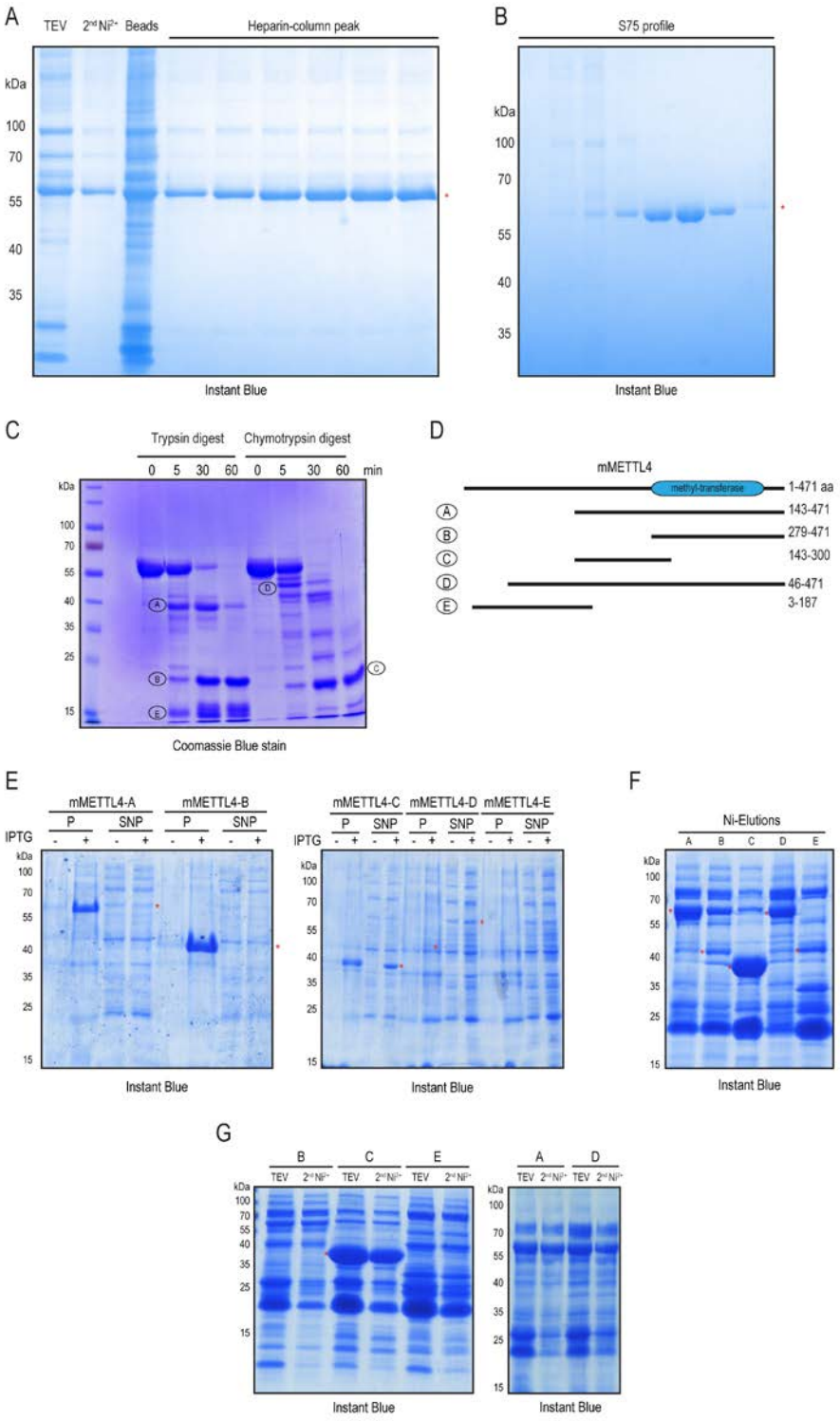


Figure 4.6 – mouse METTL4 crystallization trials. (A) Purification of mouse METTL4 expressed in insect cells. C-terminal tag was removed by TEV cleavage. Untagged mMETTL4 was further purified by Heparin followed by (B) size exclusion chromatography. (C) mMETTL4 limited proteolysis using trypsin and chymotrypsin. (D) Boundaries of the different fragments identified by mass spectrometry are represented. (E) The five constructs identified by limited proteolysis were expressed in bacterial and purified by (F) Nickel-NTA affinity purification. (G) C-terminal SUMO-6xHis tag was removed by TEV cleavage overnight. After TEV digestion, protein preparation was re-incubated with Ni-NTA beads and the flow-through was collected and analysed by SDS-PAGE. *red asterisk indicates the band corresponding to the different mouse METTL4 constructs.

4.2.6 α-mMETTL4 antibody production and validation

Polyclonal antibodies against mouse METTL4 were also raised. We expressed a small peptide (1-145 aa) from the N-terminal region of METTL4 (Figure 4.7A) in *E. coli* with a 6xHis fusion tag. The antigen was insoluble and therefore we purified it from inclusion bodies followed by Ni-NTA affinity purification under denaturing conditions (Figure 4.7B). We obtained highly pure antigen which was injected into one rabbit. Serum from the immunized rabbit was collected at several time points. The final bleed was used for the downstream experiments. Lysates from HEK293T cells transfected with human and mouse METTL4-N-HA were used to test α-mMETTL4 antibody. The homemade α-mMETTL4 was able to detect mouse METTL4 in HEK293T transfected cells. However, α-mMETTL4 was not able to identify human METTL4 (Figure 4.7C, second panel). Proteins were also detected using α-HA antibody (Figure 4.7C, first panel). We tried to detect endogenous METTL4 in several mouse tissues whole cell lysate (Figure 4.7D). A faint signal around 55 kDa was detected in heart, kidney and brain. However, we could not determine if the detected bands correspond to endogenous METTL4. α-mMETTL4 was also used for cell imaging in HeLa cells transfected with mMETTL4-N-HA. We observed that the raised antibody can recognize mMETTL4-N-HA (green signal) and the signal co-localizes with the staining obtained from the α-HA antibody (red signal) (Figure 4.7D).

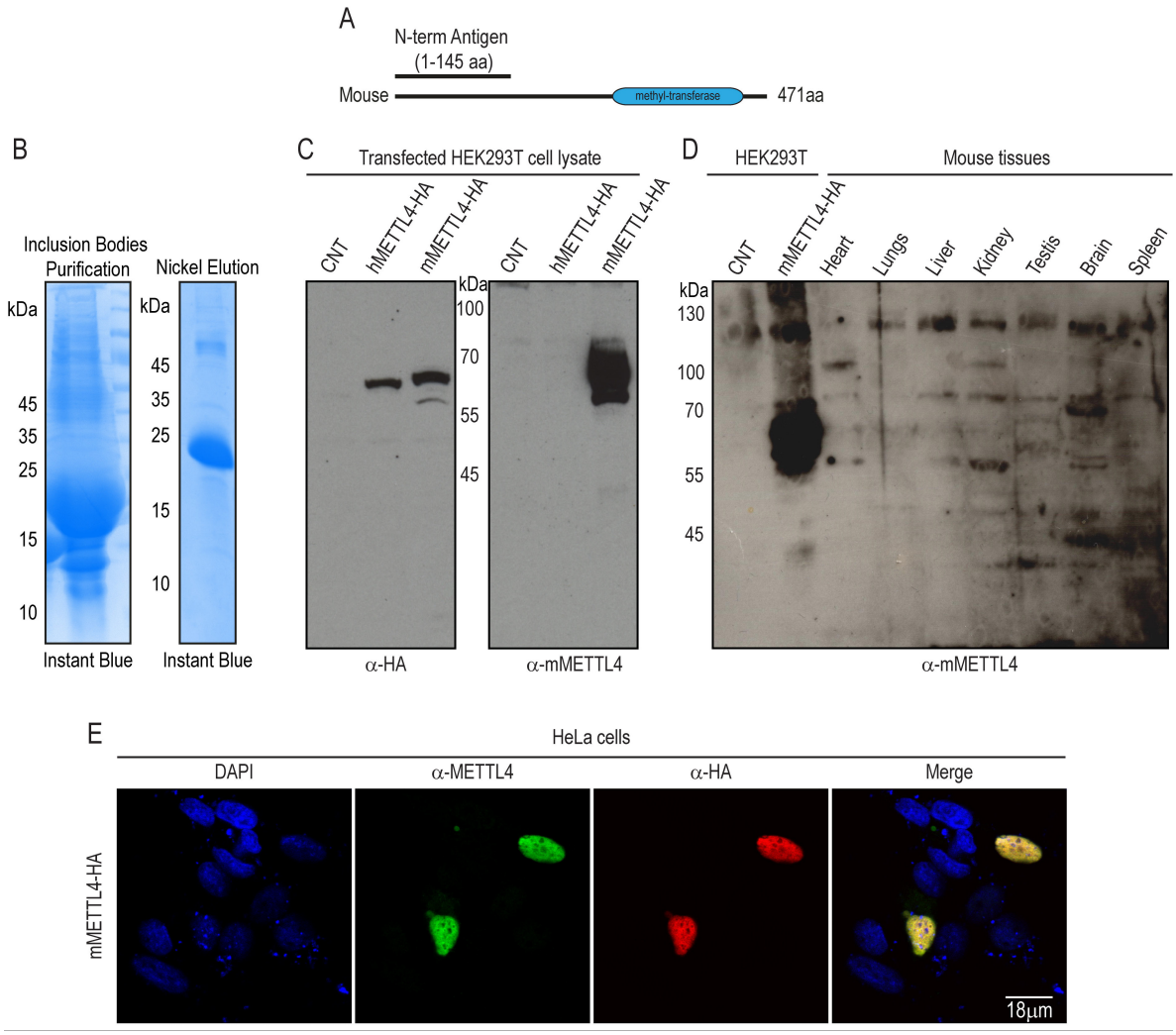


Figure 4.7 – Purification of the antigen used for the production of α-mMETTL4 and validation of the polyclonal antibodies obtained. (A) Schematic representation of N-terminal antigen used to produce α-mMETTL4 antibody. (B) Purification of the antigen from the inclusion bodies followed by nickel purification under denaturing conditions. *red asterisk indicates the band corresponding to the N-terminal antigen. (C) Western blot using lysate of HEK293T cells transfected with hMETTL4-N-HA and mMETTL4-N-HA and probed with α-HA antibody (first panel) and with α-mMETTL4 antibody (second panel). (D) Multiple tissues western blot using α-mMETTL4 antibody. Lysate of HEK293T cells transfected was used as control. (E) Detection of mMETTL4-HA in HeLa transfected cells by immunofluorescence using α-mMETTL4 antibody (green) and α-HA antibody (red). Merged images of the three signals are also represented. Scale bar is shown in µm.

4.2.7 Generation of *Mettl4* knock-out fly

To understand the role of METTL4 in flies, we disrupted the coding sequence of the *Mettl4* gene locus in the fly genome using the CRISPR-Cas9 system. Two gRNAs targeting the sole exon of *Mettl4* were used. These would generate a deletion of around 50 bp (Figure 4.8A). Several mutant flies were obtained and four independent lines were kept for future experiments (Figure 4.8B and Figure 4.8C). All the deletions obtained generates a premature stop codon and therefore METTL4 is not be expressed (Figure 4.8B).

Analysis of homozygous *Mettl4* knock-out lines #7, 18 and 6 revealed that fertility and development are not affected. These fly stocks are maintained in the lab as homozygous lines, ruling out any maternal effect of this mutation. However, homozygous *Mettl4* knock-out line #27, which has the largest deletion (74 bp) displays sub-viability (as determined by ratios of surviving heterozygotes vs homozygotes) and are sterile. They also display a curved wing phenotype that is not seen in the other lines. Therefore, this particular line might carry another mutation elsewhere in the Chr3 and the phenotype observed is not caused by loss of *Mettl4*. Only fertility and viability were analysed in the knock-out flies.

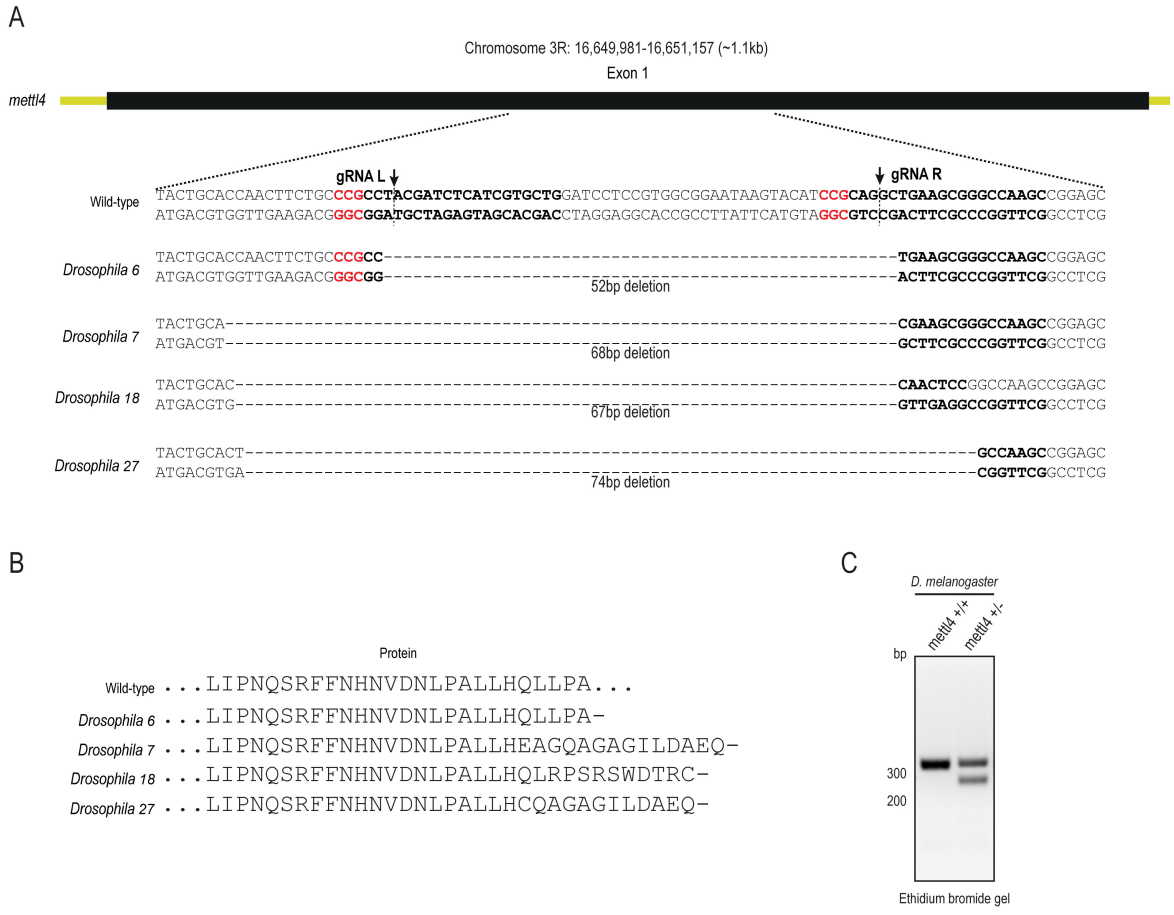


Figure 4.8 – Generation of *Mettl4* knock-out flies using the CRISPR-Cas9 system. (A) Gene organization of *Mettl4* locus and schematic representation of the gRNAs used to obtain the knock-out flies. The four independent lines analysed are also represented (B) The different deletions resulted in the generation of a premature stop codon. (C) Genotype of knock-out flies was confirmed by PCR.

4.2.8 Generation of *Mettl4*$^{-/-}$ and *Mettl4*$^{KI/KI}$ mice and phenotype characterization

We also analysed the *in vivo* role of METTL4 in mice. CRISPR-Cas9 system was used to target *mettl4* locus (see section 2.2.11 of Material and Methods). We used a single guide RNA targeting exon 4. This exon codifies for the catalytic residues (DPPW) essential for the putative methyltransferase activity of METTL4 (Figure 4.9A). We also provided an ssDNA donor template in the injection mixture. If the DNA donor template is used for repair of the double strand break, we would mutate the DPPW catalytic tetrad to APPW. We obtained one founder with a 2 bp deletion in exon 4. This generates a premature stop codon in the coding sequence of *mettl4* and therefore a knock-out allele (-). A second founder was also obtained, in which the double strand break created by Cas9 was repaired using the ssDNA donor template provided. A point-mutant (D287A) knock-in (KI) mouse was obtained (Figure 4.9B). We report that homozygous *Mettl4*$^{-/-}$ and *Mettl4*$^{KI/KI}$ animals of both sexes are viability and display normal fertility. To the date, no further experiments were done to better characterize the knock-out and knock-in mice.

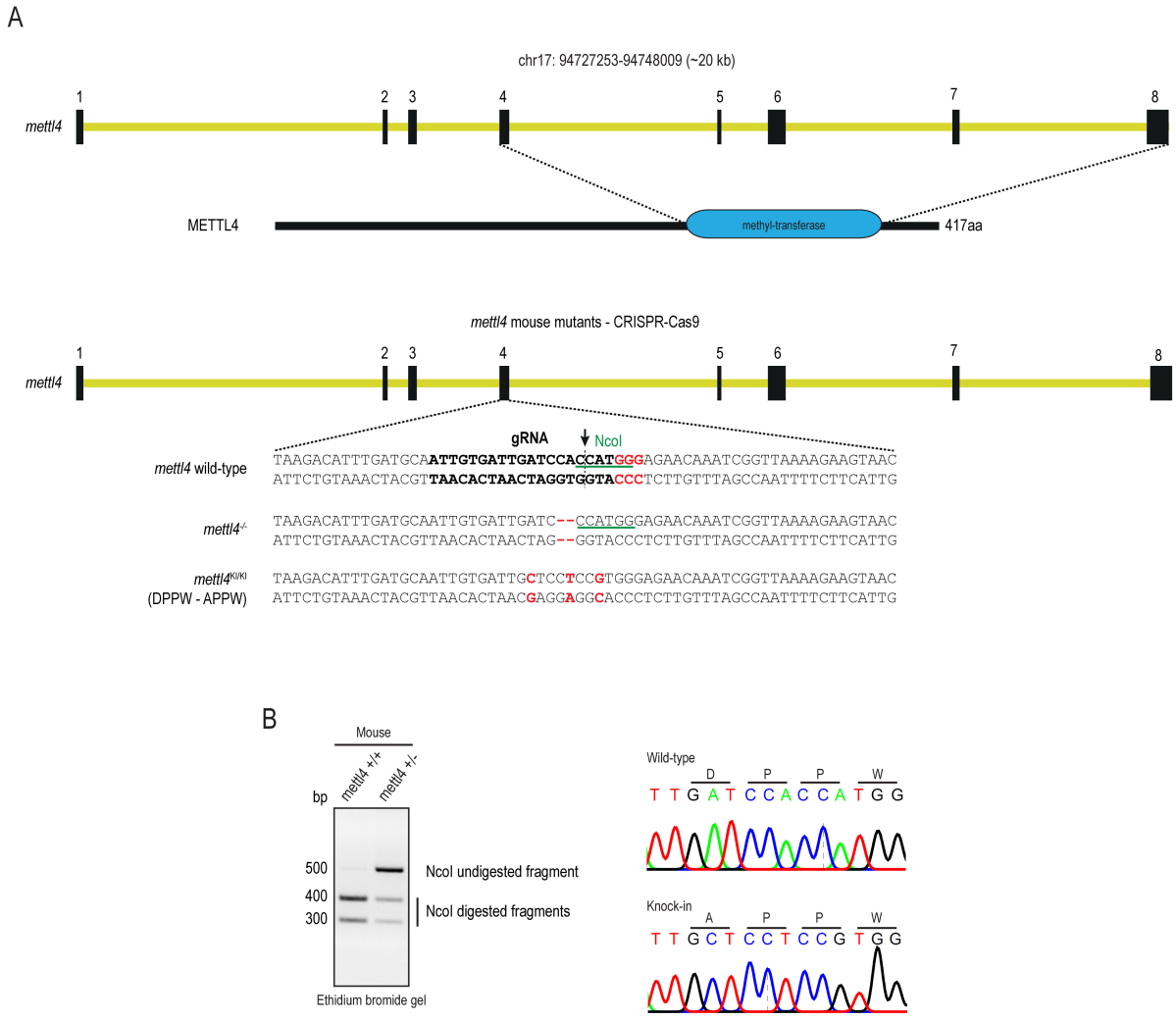


Figure 4.9 – Generation of *Mettl4* $^{-/-}$ and *Mettl4* $^{KI/KI}$ mice using the CRISPR-Cas9 system. (A) Gene organization of mouse *Mettl4* and protein domain architecture. Representation of the guide RNA used for generation of the mutant mice and the two alleles obtained: *Mettl4* $^{-/-}$ and *Mettl4* $^{KI/KI}$. (B) Genotyping and sequencing showing APPW mutation introduced in the *Mettl4* genomic loci.

4.3 Discussion and perspectives

Throughout this study we tried to biochemically characterize METTL4 methyltransferase activity *in vitro*. This protein is the closest predicted orthologue of DAMT-1, a 6mA DNA methyltransferase present in *C. elegans* (Greer *et al.* 2015). Therefore, it has been suggested that METTL4 might be the responsible for the deposition of the 6mA mark in flies and mouse.

We used purified recombinant mouse METTL4 to perform 6mA methylation assays. However, we could not detect any methyltransferase activity in the tested conditions. METTL4 did not methylate ssDNA or dsDNA substrates *in vitro*. We also tested METTL4 activity on hemi-methylated DNA. In mammals, DNMT1 is a 5mC DNA methyltransferase that acts preferentially on hemi-methylated DNA (Pradhan *et al.* 1999, Fatemi *et al.* 2001). However, METTL4 did not show any activity on this type of DNA.

As referred previously, METTL4 is an orthologue of the nematode DAMT-1 (Greer *et al.* 2015). In *C. elegans*, METTL4 is the only representative of Ime4-like/MT-A70 clade in the genome (Iyer *et al.* 2016a). Therefore, it might compensate for the absence of METTL3-METTL14 complex (Liu *et al.* 2014) and operate on ssRNA instead of DNA (Iyer *et al.* 2016a). Our trials have shown that mouse METTL4 has no methyltransferase activity towards ssRNAs.

6mA DNA consensus sequences were found in some organisms (Fu *et al.* 2015, Greer *et al.* 2015, Wu *et al.* 2016). A specific consensus sequence might also be required for mouse METTL4 to efficiently bind and methylate DNA targets. However, METTL4 did not show any methyltransferase activity when a dsDNA fragment with all the NNANN combinations was used. Similarly, DNA methylation assays using the sequence identified by Bind-N-Seq experiments were negative. Overall, these results suggest that METTL4 might not be involved in 6mA DNA methylation. However, we cannot exclude that the conditions tested might not be the most suitable. Additionally, METTL4 might require an interaction partner to be an active DNA methyltransferase. METTL3 and METTL14 are two RNA methyltransferases that exhibit weak methyltransferase activity individually. These need to form a complex to be able to efficiently methylate mammalian RNA (Liu *et al.* 2014). A third protein called WTAP was also shown to be a crucial component of the mammalian 6mA RNA methyltransferase complex (Liu *et al.* 2014, Ping *et al.* 2014a). Similarly, METTL4 might need a second DNA methyltransferase or a regulatory factor to become active. The requirement of a specific target or a specific genomic environment cannot be excluded. For instance, the 6mA methyltransferase METTL16 is responsible for the methylation of a small hairpin and it is only active on this specific RNA (Pendleton *et al.* 2017).

Analysis of *Mettl4* knock-out flies revealed that these were viable and fertile. In *D. melanogaster*, an increase of 6mA levels correlates with an overexpression of the methylated transposons (Zhang *et al.* 2015). If we assume that METTL4 is the 6mA DNA methyltransferase in flies, its deletion probably would not result in any development defect or infertility. METTL4 would be responsible for deposition of the 6mA mark and therefore for transposon activation. However, our attempts to detect the 6mA mark in flies using 6mA-specific antibodies (data not shown) were unsuccessful. Therefore, we cannot demonstrate that METTL4 is involved in 6mA DNA methylation in flies.

Analysis of *Mettl4* knock-out mouse also revealed that this protein is not essential for viability and fertility of homozygous animals. Analysis of the knock-in catalytic mutant mice confirmed that METTL4 putative methyltransferase activity is also not important for viability and fertility.

Overall, in our *in vitro* tested conditions METTL4 is not an active DNA methyltransferase. The knock-out mutant animals are viable and fertile and we were not successful in detecting the presence of 6mA mark in mouse and fly genomes. Therefore, METTL4 function is still enigmatic as well as the biological relevance of 6mA modification.

5. Concluding Remarks

Concluding Remarks

piRNAs are a class of small non-coding RNAs that play an important role on transposon silencing. These are particularly important in animal gonads where they protect the genome integrity of the germ line. It has been more than a decade since piRNAs were first described, and biochemical and genetic studies have greatly contributed for an extensive characterization of piRNA pathway. However, when this study started (back in 2014) one fundamental question regarding piRNA maturation was still unsolved: which was the protein responsible for shortening pre-piRNAs 3' ends. Our group has been trying to identify the nuclease involved in piRNA maturation and searching for potential nucleases led to the identification of NEF-sp in testis transcriptome.

Our study provides a biochemical and genetic characterization of the mammalian NEF-sp exoribonuclease. After extensive characterization, which has been described throughout this thesis, we concluded that NEF-sp is not involved in piRNA pathway as initially thought. NEF-sp is indeed a testis specific nuclease and degrades preferentially single stranded RNAs from the 3' to the 5' end in an Mg^{2+} dependent manner. NEF-sp localizes in the nucleus in HeLa transfected cells and therefore we believe that this protein works in this cellular compartment. Analysis of the *Nef-sp* mouse knock-out mutant revealed no obvious phenotype. This result made us to conclude that NEF-sp might be involved in the precise trimming of the unknown substrate(s) and its activity might be compensated by complementation by other nucleases.

In the second part of this thesis we explored the role of 6mA DNA methylation in mammals and flies. This type of DNA methylation has been recently studied in eukaryotes. However, its biological relevance is still not fully understood, particularly in mammals. We tried to characterize METTL4, a DNA methyltransferase which is predicted to be responsible for 6mA deposition in flies and mice. METTL4 was not an active DNA/RNA methyltransferase in the conditions tested. The mouse and mutant flies lack any obvious phenotype. These results made us believe that METTL4 might not be responsible for the deposition of m6A modification. However, we cannot exclude that m6A DNA modification might not have any biological relevance.

Overall, we combined biochemical and genetic approaches to understand the function of these two proteins. Our data contributes for a better understanding of the role of NEF-sp and METTL4 in mammals.

References

Aapola, U., K. Kawasaki, H. S. Scott, J. Ollila, M. Vihinen, M. Heino, A. Shintani, K. Kawasaki, S. Minoshima, K. Krohn, S. E. Antonarakis, N. Shimizu, J. Kudoh and P. Peterson (2000). "Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family." Genomics **65**(3): 293-298.

Aravin, A., D. Gaidatzis, S. Pfeffer, M. Lagos-Quintana, P. Landgraf, N. Iovino, P. Morris, M. J. Brownstein, S. Kuramochi-Miyagawa, T. Nakano, M. Chien, J. J. Russo, J. Ju, R. Sheridan, C. Sander, M. Zavolan and T. Tuschl (2006). "A novel class of small RNAs bind to MILI protein in mouse testes." Nature **442**(7099): 203-207.

Aravin, A. A., G. J. Hannon and J. Brennecke (2007a). "The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race." Science **318**(5851): 761-764.

Aravin, A. A., N. M. Naumova, A. V. Tulin, V. V. Vagin, Y. M. Rozovsky and V. A. Gvozdev (2001). "Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D-melanogaster germline." Current Biology **11**(13): 1017-1027.

Aravin, A. A., R. Sachidanandam, D. Bourc'his, C. Schaefer, D. Pezic, K. F. Toth, T. Bestor and G. J. Hannon (2008). "A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice." Mol Cell **31**(6): 785-799.

Aravin, A. A., R. Sachidanandam, A. Girard, K. Fejes-Toth and G. J. Hannon (2007b). "Developmentally regulated piRNA clusters implicate MILI in transposon control." Science **316**(5825): 744-747.

Aravin, A. A., G. W. Van Der Heijden, J. Castaneda, V. V. Vagin, G. J. Hannon and A. Bortvin (2009). "Cytoplasmic Compartmentalization of the Fetal piRNA Pathway in Mice." Plos Genetics **5**(12).

Barau, J., A. Teissandier, N. Zamudio, S. Roy, V. Nalesso, Y. Herault, F. Guillou and D. Bourc'his (2016). "The DNA methyltransferase DNMT3C protects male germ cells from transposon activity." Science **354**(6314): 909-912.

Bellve, A. R., J. C. Cavicchia, C. F. Millette, D. A. O'brien, Y. M. Bhatnagar and M. Dym (1977). "Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization." J Cell Biol **74**(1): 68-85.

Bestor, T., A. Laudano, R. Mattaliano and V. Ingram (1988). "Cloning and Sequencing of a Cdna-Encoding DNA Methyltransferase of Mouse Cells - the Carboxyl-Terminal Domain of the Mammalian Enzymes Is Related to Bacterial Restriction Methyltransferases." Journal of Molecular Biology **203**(4): 971-983.

Bestor, T. H. (1992). "Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain." EMBO J **11**(7): 2611-2617.

Bieniossek, C., T. Imasaki, Y. Takagi and I. Berger (2012). "MultiBac: expanding the research toolbox for multiprotein complexes." Trends Biochem Sci **37**(2): 49-57.

Bourc'his, D. and T. H. Bestor (2004). "Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L." Nature **431**(7004): 96-99.

Bourc'his, D., G. L. Xu, C. S. Lin, B. Bollman and T. H. Bestor (2001). "Dnmt3L and the establishment of maternal genomic imprints." Science **294**(5551): 2536-2539.

Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam and G. J. Hannon (2007). "Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila." Cell **128**(6): 1089-1103.

Carmell, M. A., A. Girard, H. J. G. Van De Kant, D. Bourc'his, T. H. Bestor, D. G. De Rooij and G. J. Hannon (2007). "MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline." Dev Cell **12**(4): 503-514.

Carmo-Fonseca, M., L. Mendes-Soares and I. Campos (2000). "To be or not to be in the nucleolus." Nat Cell Biol **2**(6): E107-112.

Chedin, F., M. R. Lieber and C. L. Hsieh (2002). "The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a." Proc Natl Acad Sci U S A **99**(26): 16916-16921.

Chelsky, D., R. Ralph and G. Jonak (1989). "Sequence Requirements for Synthetic Peptide-Mediated Translocation to the Nucleus." Molecular and Cellular Biology **9**(6): 2487-2492.

Cool, J., T. Defalco and B. Capel (2012). "Testis formation in the fetal mouse: dynamic and complex de novo tubulogenesis." Wiley Interdisciplinary Reviews-Developmental Biology **1**(6): 847-859.

Cora, E., R. R. Pandey, J. Xiol, J. Taylor, R. Sachidanandam, A. A. Mccarthy and R. S. Pillai (2014). "The MID-PIWI module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs." RNA **20**(6): 773-781.

Cox, D. N., A. Chao, J. Baker, L. Chang, D. Qiao and H. F. Lin (1998). "A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal." Genes Dev **12**(23): 3715-3727.

Daugeron, M. C., F. Mauxion and B. Seraphin (2001). "The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation." Nucleic Acids Res **29**(12): 2448-2455.

De Fazio, S., N. Bartonicek, M. Di Giacomo, C. Abreu-Goodger, A. Sankar, C. Funaya, C. Antony, P. N. Moreira, A. J. Enright and D. O'carroll (2011). "The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements." Nature **480**(7376): 259-U148.

Deng, W. and H. F. Lin (2002). "miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis." Dev Cell **2**(6): 819-830.

Deo, R. C., J. B. Bonanno, N. Sonenberg and S. K. Burley (1999). "Recognition of polyadenylate RNA by the poly(A)-binding protein." Cell **98**(6): 835-845.

Di Giacomo, M., S. Comazzetto, H. Saini, S. De Fazio, C. Carrieri, M. Morgan, L. Vasiliauskaite, V. Benes, A. J. Enright and D. O'carroll (2013). "Multiple Epigenetic Mechanisms and the piRNA Pathway Enforce LINE1 Silencing during Adult Spermatogenesis." Molecular Cell **50**(4): 601-608.

Dominissini, D., S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek and G. Rechavi (2012). "Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq." Nature **485**(7397): 201-206.

Dupeux, F., M. Rower, G. Seroul, D. Blot and J. A. Marquez (2011). "A thermal stability assay can help to estimate the crystallization likelihood of biological samples." Acta Crystallogr D Biol Crystallogr **67**(Pt 11): 915-919.

Ehrlich, M., M. A. Gama-Sosa, L. H. Huang, R. M. Midgett, K. C. Kuo, R. A. Mccune and C. Gehrke (1982). "Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells." Nucleic Acids Res **10**(8): 2709-2721.

Fatemi, M., A. Hermann, S. Pradhan and A. Jeltsch (2001). "The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA." Journal of Molecular Biology **309**(5): 1189-1199.

Fu, Y., G. Z. Luo, K. Chen, X. Deng, M. Yu, D. Han, Z. Hao, J. Liu, X. Lu, L. C. Dore, X. Weng, Q. Ji, L. Mets and C. He (2015). "N6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas." Cell **161**(4): 879-892.

Ginsburg, M., M. H. L. Snow and A. Mclaren (1990). "Primordial Germ-Cells in the Mouse Embryo during Gastrulation." Development **110**(2): 521-&.

Giordano, J., Y. C. Ge, Y. Gelfand, G. Abrusan, G. Benson and P. E. Warburton (2007). "Evolutionary history of mammalian transposons determined by genome-wide defragmentation." Plos Computational Biology **3**(7): 1321-1334.

Girard, A., R. Sachidanandam, G. J. Hannon and M. A. Carmell (2006). "A germline-specific class of small RNAs binds mammalian Piwi proteins." Nature **442**(7099): 199-202.

Goll, M. G., F. Kirpekar, K. A. Maggert, J. A. Yoder, C. L. Hsieh, X. Y. Zhang, K. G. Golic, S. E. Jacobsen and T. H. Bestor (2006). "Methylation of tRNA(AsP) by the DNA methyltransferase homolog Dnmt2." Science **311**(5759): 395-398.

Greer, E. L., M. A. Blanco, L. Gu, E. Sendinc, J. Liu, D. Aristizabal-Corrales, C. H. Hsu, L. Aravind, C. He and Y. Shi (2015). "DNA Methylation on N6-Adenine in C. elegans." Cell **161**(4): 868-878.

Grivna, S. T., E. Beyret, Z. Wang and H. Lin (2006). "A novel class of small RNAs in mouse spermatogenic cells." Genes Dev **20**(13): 1709-1714.

Gunawardane, L. S., K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi and M. C. Siomi (2007). "A slicer-mediated mechanism for repeat-associated siRNA 5 ' end formation in Drosophila." Science **315**(5818): 1587-1590.

Guo, F., X. L. Li, D. Liang, T. Li, P. Zhu, H. S. Guo, X. L. Wu, L. Wen, T. P. Gu, B. Q. Hu, C. P. Walsh, J. S. Li, F. C. Tang and G. L. Xu (2014). "Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote." Cell Stem Cell **15**(4): 447-458.

Hackett, J. A., R. Sengupta, J. J. Zylicz, K. Murakami, C. Lee, T. A. Down and M. A. Surani (2013). "Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine." Science **339**(6118): 448-452.

Hajkova, P., S. Erhardt, N. Lane, T. Haaf, O. El-Maarri, W. Reik, J. Walter and M. A. Surani (2002). "Epigenetic reprogramming in mouse primordial germ cells." Mech Dev **117**(1-2): 15-23.

Han, B. W., W. Wang, C. J. Li, Z. P. Weng and P. D. Zamore (2015). "piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production." Science **348**(6236): 817-821.

Handler, D., K. Meixner, M. Pizka, K. Lauss, C. Schmied, F. S. Gruber and J. Brennecke (2013). "The genetic makeup of the Drosophila piRNA pathway." Mol Cell **50**(5): 762-777.

Hata, K., M. Kusumi, T. Yokomine, E. Li and H. Sasaki (2006). "Meiotic and epigenetic aberrations in Dnmt3L-deficient male germ cells." Molecular Reproduction and Development **73**(1): 116-122.

Hata, K., M. Okano, H. Lei and E. Li (2002). "Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice." Development **129**(8): 1983-1993.

Hattman, S. (2005). "DNA-[adenine] methylation in lower eukaryotes." Biochemistry (Mosc) **70**(5): 550-558.

Hattman, S., C. Kenny, L. Berger and K. Pratt (1978). "Comparative study of DNA methylation in three unicellular eucaryotes." J Bacteriol **135**(3): 1156-1157.

Heyn, H. and M. Esteller (2015). "An Adenine Code for DNA: A Second Life for N6-Methyladenine." Cell **161**(4): 710-713.

Homolka, D., R. R. Pandey, C. Goriaux, E. Brasset, C. Vaury, R. Sachidanandam, M. O. Fauvarque and R. S. Pillai (2015). "PIWI Slicing and RNA Elements in Precursors Instruct Directional Primary piRNA Biogenesis." Cell Rep **12**(3): 418-428.

Honda, S., Y. Kirino, M. Maragkakis, P. Alexiou, A. Ohtaki, R. Murali, Z. Mourelatos and Y. Kirino (2013). "Mitochondrial protein BmPAPI modulates the length of mature piRNAs." Rna-a Publication of the Rna Society **19**(10): 1405-1418.

Horwich, M. D., C. J. Li, C. Matranga, V. Vagin, G. Farley, P. Wang and P. D. Zamore (2007). "The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC." Current Biology **17**(14): 1265-1272.

Howell, C. Y., T. H. Bestor, F. Ding, K. E. Latham, C. Mertineit, J. M. Trasler and J. R. Chaillet (2001). "Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene." Cell **104**(6): 829-838.

Huang, Y., L. Ji, Q. Huang, D. G. Vassylyev, X. Chen and J. B. Ma (2009). "Structural insights into mechanisms of the small RNA methyltransferase HEN1." Nature **461**(7265): 823-827.

Huckins, C. (1971). "The spermatogonial stem cell population in adult rats. I. Their morphology, proliferation and maturation." Anat Rec **169**(3): 533-557.

Ipsaro, J. J., A. D. Haase, S. R. Knott, L. Joshua-Tor and G. J. Hannon (2012). "The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis." Nature **491**(7423): 279-U151.

Iyer, L. M., S. Abhiman and L. Aravind (2011). "Natural history of eukaryotic DNA methylation systems." Prog Mol Biol Transl Sci **101**: 25-104.

Iyer, L. M., D. Zhang and L. Aravind (2016a). "Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification." Bioessays **38**(1): 27-40.

Iyer, L. M., D. P. Zhang and L. Aravind (2016b). "Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification." Bioessays **38**(1): 27-40.

Izumi, N., K. Shoji, Y. Sakaguchi, S. Honda, Y. Kirino, T. Suzuki, S. Katsuma and Y. Tomari (2016). "Identification and Functional Analysis of the Pre-piRNA 3 ' Trimmer in Silkworms." Cell **164**(5): 962-973.

Ji, L. and X. Chen (2012). "Regulation of small RNA stability: methylation and beyond." Cell Res **22**(4): 624-636.

Jia, G. F., Y. Fu, X. Zhao, Q. Dai, G. Q. Zheng, Y. Yang, C. Q. Yi, T. Lindahl, T. Pan, Y. G. Yang and C. He (2011). "N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO." Nature Chemical Biology **7**(12): 885-887.

Joyce, C. M. and T. A. Steitz (1994). "Function and Structure Relationships in DNA-Polymerases." Annual Review of Biochemistry **63**: 777-822.

Kadlec, J., E. Izaurralde and S. Cusack (2004). "The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3." Nat Struct Mol Biol **11**(4): 330-337.

Kalmykova, A. I., M. S. Klenov and V. A. Gvozdev (2005). "Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline." Nucleic Acids Res **33**(6): 2052-2059.

Kaneda, M., M. Okano, K. Hata, T. Sado, N. Tsujimoto, E. Li and H. Sasaki (2004). "Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting." Nature **429**(6994): 900-903.

Kato, Y., M. Kaneda, K. Hata, K. Kumaki, M. Hisano, Y. Kohara, M. Okano, E. Li, M. Nozaki and H. Sasaki (2007). "Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse." Human Molecular Genetics **16**(19): 2272-2280.

Katz, D. J., T. M. Edwards, V. Reinke and W. G. Kelly (2009). "A C. elegans LSD1 Demethylase Contributes to Germline Immortality by Reprogramming Epigenetic Memory." Cell **137**(2): 308-320.

Kawaoka, S., N. Hayashi, Y. Suzuki, H. Abe, S. Sugano, Y. Tomari, T. Shimada and S. Katsuma (2009). "The Bombyx ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes." Rna-a Publication of the Rna Society **15**(7): 1258-1264.

Kawaoka, S., N. Izumi, S. Katsuma and Y. Tomari (2011). "3 ' End Formation of PIWI-Interacting RNAs In Vitro." Mol Cell **43**(6): 1015-1022.

Kenan, D. J., C. C. Query and J. D. Keene (1991). "Rna Recognition - Towards Identifying Determinants of Specificity." Trends in Biochemical Sciences **16**(6): 214-220.

Kirino, Y. and Z. Mourelatos (2007). "The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs." Rna-a Publication of the Rna Society **13**(9): 1397-1401.

Klass, M., P. N. Nguyen and A. Dechavigny (1983). "Age-Correlated Changes in the DNA-Template in the Nematode Caenorhabditis-Elegans." Mechanisms of Ageing and Development **22**(3-4): 253-263.

Kosugi, S., M. Hasebe, T. Entani, S. Takayama, M. Tomita and H. Yanagawa (2008). "Design of peptide inhibitors for the importin alpha/beta nuclear import pathway by activity-based profiling." Chem Biol **15**(9): 940-949.

Kosugi, S., M. Hasebe, M. Tomita and H. Yanagawa (2009). "Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs." Proc Natl Acad Sci U S A **106**(25): 10171-10176.

Kotaja, N., H. F. Lin, M. Parvinen and P. Sassone-Corsi (2006). "Interplay of PIWI/Argonaute protein MIWI and kinesin KIF17b in chromatoid bodies of male germ cells." Journal of Cell Science **119**(13): 2819-2825.

Koziol, M. J., C. R. Bradshaw, G. E. Allen, A. S. H. Costa, C. Frezza and J. B. Gurdon (2016). "Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications." Nat Struct Mol Biol **23**(1): 24-30.

Kuramochi-Miyagawa, S., T. Kimura, T. W. Ijiri, T. Isobe, N. Asada, Y. Fujita, M. Ikawa, N. Iwai, M. Okabe, W. Deng, H. F. Lin, Y. Matsuda and T. Nakano (2004). "Mili, a mammalian member of piwi family gene, is essential for spermatogenesis." Development **131**(4): 839-849.

Kuramochi-Miyagawa, S., T. Kimura, K. Yomogida, A. Kuroiwa, Y. Tadokoro, Y. Fujita, M. Sato, Y. Matsuda and T. Nakano (2001). "Two mouse piwi-related genes: miwi and mili." Mech Dev **108**(1-2): 121-133.

Kuramochi-Miyagawa, S., T. Watanabe, K. Gotoh, Y. Totoki, A. Toyoda, M. Ikawa, N. Asada, K. Kojima, Y. Yamaguchi, T. W. Ijiri, K. Hata, E. Li, Y. Matsuda, T. Kimura, M. Okabe, Y. Sakaki, H. Sasaki and T. Nakano (2008). "DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes." Genes Dev **22**(7): 908-917.

Lane, N., W. Dean, S. Erhardt, P. Hajkova, A. Surani, J. Walter and W. Reik (2003). "Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse." Genesis **35**(2): 88-93.

Lauster, R., T. A. Trautner and M. Noyerweidner (1989). "Cytosine-Specific Type-Ii DNA Methyltransferases - a Conserved Enzyme Core with Variable Target-Recognizing Domains." Journal of Molecular Biology **206**(2): 305-312.

Lawson, K. A., N. R. Dunn, B. a. J. Roelen, L. M. Zeinstra, A. M. Davis, C. V. E. Wright, J. P. W. F. M. Korving and B. L. M. Hogan (1999). "Bmp4 is required for the generation of primordial germ cells in the mouse embryo." Genes Dev **13**(4): 424-436.

Lei, H., S. P. Oh, M. Okano, R. Juttermann, K. A. Goss, R. Jaenisch and E. Li (1996). "De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells." Development **122**(10): 3195-3205.

Leonhardt, H., A. W. Page, H. U. Weier and T. H. Bestor (1992). "A Targeting Sequence Directs DNA Methyltransferase to Sites of DNA-Replication in Mammalian Nuclei." Cell **71**(5): 865-873.

Li, C. J., V. V. Vagin, S. H. Lee, J. Xu, S. M. Ma, H. L. Xi, H. Seitz, M. D. Horwich, M. Syrzycka, B. M. Honda, E. L. W. Kittler, M. L. Zapp, C. Klattenhoff, N. Schulz, W. E. Theurkauf, Z. P. Weng and P. D. Zamore (2009). "Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies." Cell **137**(3): 509-521.

Li, E., T. H. Bestor and R. Jaenisch (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." Cell **69**(6): 915-926.

Li, J. Y., D. J. Lees-Murdock, G. L. Xu and C. P. Walsh (2004). "Timing of establishment of paternal methylation imprints in the mouse." Genomics **84**(6): 952-960.

Liu, F. G., W. Clark, G. Z. Luo, X. Y. Wang, Y. Fu, J. B. Wei, X. Wang, Z. Y. Hao, Q. Dai, G. Q. Zheng, H. H. Ma, D. L. Han, M. Evans, A. Klungland, T. Pan and C. He (2016). "ALKBH1-Mediated tRNA Demethylation Regulates Translation." Cell **167**(3): 816-+.

Liu, J., Y. Yue, D. Han, X. Wang, Y. Fu, L. Zhang, G. Jia, M. Yu, Z. Lu, X. Deng, Q. Dai, W. Chen and C. He (2014). "A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation." Nat Chem Biol **10**(2): 93-95.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biol **15**(12): 550.

Low, D. A., N. J. Weyand and M. J. Mahan (2001). "Roles of DNA adenine methylation in regulating bacterial gene expression and virulence." Infection and Immunity **69**(12): 7197-7204.

Lu, M., J. L. Campbell, E. Boye and N. Kleckner (1994). "Seqa - a Negative Modulator of Replication Initiation in Escherichia-Coli." Cell **77**(3): 413-426.

Lyko, F. and R. Maleszka (2011). "Insects as innovative models for functional studies of DNA methylation." Trends in Genetics **27**(4): 127-131.

Mclaren, A. (2003). "Primordial germ cells in the mouse." Developmental Biology **262**(1): 1-15.

Mertineit, C., J. A. Yoder, T. Taketo, D. W. Laird, J. M. Trasler and T. H. Bestor (1998). "Sex-specific exons control DNA methyltransferase in mammalian germ cells." Development **125**(5): 889-897.

Meyer, K. D., Y. Saletore, P. Zumbo, O. Elemento, C. E. Mason and S. R. Jaffrey (2012). "Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons." Cell **149**(7): 1635-1646.

Mohn, F., D. Handler and J. Brennecke (2015). "piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis." Science **348**(6236): 812-817.

Monk, M., M. Boubelik and S. Lehnert (1987). "Temporal and Regional Changes in DNA Methylation in the Embryonic, Extraembryonic and Germ-Cell Lineages during Mouse Embryo Development." Development **99**(3): 371-382.

Nagai, K., C. Oubridge, T. H. Jessen, J. Li and P. R. Evans (1990). "Crystal-Structure of the Rna-Binding Domain of the U1 Small Nuclear Ribonucleoprotein-A." Nature **348**(6301): 515-520.

Nariai, M., T. Tanaka, T. Okada, C. Shirai, C. Horigome and K. Mizuta (2005). "Synergistic defect in 60S ribosomal subunit assembly caused by a mutation of Rrs1p, a ribosomal protein L11-binding protein, and 3'-extension of 5S rRNA in Saccharomyces cerevisiae." Nucleic Acids Res **33**(14): 4553-4562.

Nishimasu, H., H. Ishizu, K. Saito, S. Fukuhara, M. K. Kamatani, L. Bonnefond, N. Matsumoto, T. Nishizawa, K. Nakanaga, J. Aoki, R. Ishitani, H. Siomi, M. C. Siomi and O. Nureki (2012). "Structure and function of Zucchini endoribonuclease in piRNA biogenesis." Nature **491**(7423): 284-U157.

Oakberg, E. F. (1971). "Spermatogonial stem-cell renewal in the mouse." Anat Rec **169**(3): 515-531.

Ohinata, Y., B. Payer, D. O'carroll, K. Ancelin, Y. Ono, M. Sano, S. C. Barton, T. Obukhanych, M. Nussenzweig, A. Tarakhovsky, M. Saitou and M. A. Surani (2005). "Blimp1 is a critical determinant of the germ cell lineage in mice." Nature **436**(7048): 207-213.

Okano, M., D. W. Bell, D. A. Haber and E. Li (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-257.

Okano, M., S. P. Xie and E. Li (1998). "Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases." Nature Genetics **19**(3): 219-220.

Olek, A. and J. Walter (1997). "The pre-implantation ontogeny of the H19 methylation imprint." Nature Genetics **17**(3): 275-276.

Olson, A. J., J. Brennecke, A. A. Aravin, G. J. Hannon and R. Sachidanandam (2008). "Analysis of large-scale sequencing of small RNAs." Pac Symp Biocomput: 126-136.

Oswald, J., S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik and J. Walter (2000). "Active demethylation of the paternal genome in the mouse zygote." Current Biology **10**(8): 475-478.

Ozanick, S. G., X. Wang, M. Costanzo, R. L. Brost, C. Boone and J. T. Anderson (2009). "Rex1p deficiency leads to accumulation of precursor initiator tRNAMet and polyadenylation of substrate RNAs in Saccharomyces cerevisiae." Nucleic Acids Res **37**(1): 298-308.

Pandey, R. R., Y. Tokuzawa, Z. Yang, E. Hayashi, T. Ichisaka, S. Kajita, Y. Asano, T. Kunieda, R. Sachidanandam, S. Chuma, S. Yamanaka and R. S. Pillai (2013). "Tudor domain containing 12 (TDRD12) is essential for secondary PIWI interacting RNA biogenesis in mice." Proc Natl Acad Sci U S A **110**(41): 16492-16497.

Pane, A., K. Wehr and T. Schupbach (2007). "zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline." Dev Cell **12**(6): 851-862.

Patil, D. P., C. K. Chen, B. F. Pickering, A. Chow, C. Jackson, M. Guttman and S. R. Jaffrey (2016). "m(6)A RNA methylation promotes XIST-mediated transcriptional repression." Nature **537**(7620): 369-+.

Pendleton, K. E., B. Chen, K. Liu, O. V. Hunter, Y. Xie, B. P. Tu and N. K. Conrad (2017). "The U6 snRNA m6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention." Cell **169**(5): 824-835 e814.

Phalke, S., O. Nickel, D. Walluscheck, F. Hortig, M. C. Onorati and G. Reuter (2009). "Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2." Nature Genetics **41**(6): 696-702.

Pillai, R. S., S. N. Bhattacharyya, C. G. Artus, T. Zoller, N. Cougot, E. Basyuk, E. Bertrand and W. Filipowicz (2005). "Inhibition of translational initiation by Let-7 microRNA in human cells." Science **309**(5740): 1573-1576.

Ping, X. L., B. F. Sun, L. Wang, W. Xiao, X. Yang, W. J. Wang, S. Adhikari, Y. Shi, Y. Lv, Y. S. Chen, X. Zhao, A. Li, Y. Yang, U. Dahal, X. M. Lou, X. Liu, J. Huang, W. P. Yuan, X. F. Zhu, T. Cheng, Y. L. Zhao, X. Wang, J. M. Rendtlew Danielsen, F. Liu and Y. G. Yang (2014a). "Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase." Cell Res **24**(2): 177-189.

Ping, X. L., B. F. Sun, L. Wang, W. Xiao, X. Yang, W. J. Wang, S. Adhikari, Y. Shi, Y. Lv, Y. S. Chen, X. Zhao, A. Li, Y. Yang, U. Dahal, X. M. Lou, X. Liu, J. Huang, W. P. Yuan, X. F. Zhu, T. Cheng, Y. L. Zhao, X. Q. Wang, J. M. R. Danielsen, F. Liu and Y. G. Yang (2014b). "Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase." Cell Res **24**(2): 177-189.

Posfai, J., A. S. Bhagwat, G. Posfai and R. J. Roberts (1989). "Predictive Motifs Derived from Cytosine Methyltransferases." Nucleic Acids Research **17**(7): 2421-2435.

Pradhan, S., A. Bacolla, R. D. Wells and R. J. Roberts (1999). "Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation." J Biol Chem **274**(46): 33002-33010.

Prilusky, J., C. E. Felder, T. Zeev-Ben-Mordehai, E. H. Rydberg, O. Man, J. S. Beckmann, I. Silman and J. L. Sussman (2005). "FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded." Bioinformatics **21**(16): 3435-3438.

Raddatz, G., P. M. Guzzardo, N. Olova, M. R. Fantappie, M. Rampp, M. Schaefer, W. Reik, G. J. Hannon and F. Lyko (2013). "Dnmt2-dependent methylomes lack defined DNA methylation patterns." Proc Natl Acad Sci U S A **110**(21): 8627-8631.

Ramachandran, V. and X. Chen (2008). "Degradation of microRNAs by a family of exoribonucleases in Arabidopsis." Science **321**(5895): 1490-1492.

Reuter, M., P. Berninger, S. Chuma, H. Shah, M. Hosokawa, C. Funaya, C. Antony, R. Sachidanandam and R. S. Pillai (2011). "Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing." Nature **480**(7376): 264-U154.

Reuter, M., S. Chuma, T. Tanaka, T. Franz, A. Stark and R. S. Pillai (2009). "Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile." Nat Struct Mol Biol **16**(6): 639-646.

Rogers, S. D., M. E. Rogers, G. Saunders and G. Holt (1986). "Isolation of mutants sensitive to 2-aminopurine and alkylating agents and evidence for the role of DNA methylation in Penicillium chrysogenum." Curr Genet **10**(7): 557-560.

Saito, K., K. M. Nishida, T. Mori, Y. Kawamura, K. Miyoshi, T. Nagami, H. Siomi and M. C. Siomi (2006). "Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome." Genes Dev **20**(16): 2214-2222.

Saito, K., Y. Sakaguchi, T. Suzuki, T. Suzuki, H. Siomi and M. C. Siomi (2007). "Pimet, the Drosophila homolog of HEN1, mediates 2 '-O-methylation of PIWI-interacting RNAs at their 3 ' ends." Genes Dev **21**(13): 1603-1608.

Santos, F., B. Hendrich, W. Reik and W. Dean (2002). "Dynamic reprogramming of DNA methylation in the early mouse embryo." Dev Biol **241**(1): 172-182.

Savitsky, M., D. Kwon, P. Georgiev, A. Kalmykova and V. Gvozdev (2006). "Telomere elongation is under the control of the RNAi-based mechanism in the Drosophila germline." Genes Dev **20**(3): 345-354.

Saxe, J. P., M. J. Chen, H. Y. Zhao and H. F. Lin (2013). "Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline." Embo Journal **32**(13): 1869-1885.

Saxonov, S., P. Berg and D. L. Brutlag (2006). "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters." Proc Natl Acad Sci U S A **103**(5): 1412-1417.

Schmidt, A., G. Palumbo, M. P. Bozzetti, P. Tritto, S. Pimpinelli and U. Schafer (1999). "Genetic and molecular characterization of sting, a gene involved in crystal formation and meiotic drive in the male germ line of Drosophila melanogaster." Genetics **151**(2): 749-760.

Scott, M. S., F. M. Boisvert, M. D. Mcdowall, A. I. Lamond and G. J. Barton (2010). "Characterization and prediction of protein nucleolar localization sequences." Nucleic Acids Research **38**(21): 7388-7399.

Scott, M. S., P. V. Troshin and G. J. Barton (2011). "NoD: a Nucleolar localization sequence detector for eukaryotic and viral proteins." Bmc Bioinformatics **12**.

Shima, F., H. Narita, A. Hiura, H. Shimoda and M. Akashi (2017). "Construction and histological analysis of a 3D human arterial wall model containing vasa vasorum using a layer-by-layer technique." J Biomed Mater Res A **105**(3): 814-823.

Shimizu-Yoshida, Y., M. Sasamoto, A. Yoshida, T. Yoshioka, A. Matsumoto and A. Sakai (1999). "Mouse CAF1, a mouse homologue of the yeast POP2 gene, complements the yeast pop2 null mutation." Yeast **15**(13): 1357-1364.

Siomi, M. C., K. Sato, D. Pezic and A. A. Aravin (2011). "PIWI-interacting small RNAs: the vanguard of genome defence." Nat Rev Mol Cell Biol **12**(4): 246-258.

Steitz, T. A. and J. A. Steitz (1993). "A general two-metal-ion mechanism for catalytic RNA." Proc Natl Acad Sci U S A **90**(14): 6498-6502.

Suetake, I., F. Shinozaki, J. Miyagawa, H. Takeshima and S. Tajima (2004). "DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction." Journal of Biological Chemistry **279**(26): 27816-27823.

Tanaka, S. S., Y. Toyooka, R. Akasu, Y. Katoh-Fukui, Y. Nakahara, R. Suzuki, M. Yokoyama and T. Noce (2000). "The mouse homolog of Drosophila Vasa is required for the development of male germ cells." Genes & Development **14**(7): 841-853.

Thore, S., F. Mauxion, B. Seraphin and D. Suck (2003). "X-ray structure and activity of the yeast Pop2 protein: a nuclease subunit of the mRNA deadenylase complex." Embo Reports **4**(12): 1150-1155.

Vagin, V. V., M. S. Klenov, A. I. Kalmykova, A. D. Stolyarenko, R. N. Kotelnikov and V. A. Gvozdev (2004). "The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of Drosophila melanogaster." RNA Biol **1**(1): 54-58.

Vagin, V. V., A. Sigova, C. J. Li, H. Seitz, V. Gvozdev and P. D. Zamore (2006). "A distinct small RNA pathway silences selfish genetic elements in the germline." Science **313**(5785): 320-324.

Van Hoof, A., P. Lennertz and R. Parker (2000). "Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast." EMBO J **19**(6): 1357-1365.

Viswanathan, P., T. Ohn, Y. C. Chiang, J. Chen and C. L. Denis (2004). "Mouse CAF1 can function as a processive deadenylase/3'-5'-exonuclease in vitro but in yeast the deadenylase function of CAF1 is not required for mRNA poly(A) removal." J Biol Chem **279**(23): 23988-23995.

Wang, X., Z. Lu, A. Gomez, G. C. Hon, Y. Yue, D. Han, Y. Fu, M. Parisien, Q. Dai, G. Jia, B. Ren, T. Pan and C. He (2014). "N6-methyladenosine-dependent regulation of messenger RNA stability." Nature **505**(7481): 117-120.

Wang, X., B. S. Zhao, I. A. Roundtree, Z. Lu, D. Han, H. Ma, X. Weng, K. Chen, H. Shi and C. He (2015). "N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency." Cell **161**(6): 1388-1399.

Wasmuth, E. V., K. Januszyk and C. D. Lima (2014). "Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA." Nature **511**(7510): 435-439.

Watanabe, T., S. Chuma, Y. Yamamoto, S. Kuramochi-Miyagawa, Y. Totoki, A. Toyoda, Y. Hoki, A. Fujiyama, T. Shibata, T. Sado, T. Noce, T. Nakano, N. Nakatsuji, H. F. Lin and H. Sasaki (2011). "MITOPLD Is a Mitochondrial Protein Essential for Nuage Formation and piRNA Biogenesis in the Mouse Germline." Dev Cell **20**(3): 364-375.

Watanabe, T., Y. Totoki, A. Toyoda, M. Kaneda, S. Kuramochi-Miyagawa, Y. Obata, H. Chiba, Y. Kohara, T. Kono, T. Nakano, M. A. Surani, Y. Sakaki and H. Sasaki (2008). "Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes." Nature **453**(7194): 539-U539.

Waterston, R. H., K. Lindblad-Toh, E. Birney and M. G. S. Consor (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-562.

Weber, M., I. Hellmann, M. B. Stadler, L. Ramos, S. Paabo, M. Rebhan and D. Schubeler (2007). "Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome." Nature Genetics **39**(4): 457-466.

Wenda, J. M., D. Homolka, Z. Yang, P. Spinelli, R. Sachidanandam, R. R. Pandey and R. S. Pillai (2017). "Distinct Roles of RNA Helicases MVH and TDRD9 in PIWI Slicing-Triggered Mammalian piRNA Biogenesis and Function." Dev Cell **41**(6): 623-637 e629.

Wu, S. C. and Y. Zhang (2010). "Active DNA demethylation: many roads lead to Rome." Nature Reviews Molecular Cell Biology **11**(9): 607-620.

Wu, T. P., T. Wang, M. G. Seetin, Y. Q. Lai, S. J. Zhu, K. X. Lin, Y. F. Liu, S. D. Byrum, S. G. Mackintosh, M. Zhong, A. Tackett, G. L. Wang, L. S. Hon, G. Fang, J. A. Swenberg and A. Z. Xiao (2016). "DNA methylation on N-6-adenine in mammalian embryonic stem cellse." Nature **532**(7599): 329-+.

Xiao, W., S. Adhikari, U. Dahal, Y. S. Chen, Y. J. Hao, B. F. Sun, H. Y. Sun, A. Li, X. L. Ping, W. Y. Lai, X. Wang, H. L. Ma, C. M. Huang, Y. Yang, N. Huang, G. B. Jiang, H. L. Wang, Q. Zhou, X. J. Wang, Y. L. Zhao and Y. G. Yang (2016). "Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing." Mol Cell **61**(4): 507-519.

Yamaji, M., Y. Seki, K. Kurimoto, Y. Yabuta, M. Yuasa, M. Shigeta, K. Yamanaka, Y. Ohinata and M. Saitou (2008). "Critical function of Prdm14 for the establishment of the germ cell lineage in mice." Nature Genetics **40**(8): 1016-1022.

Yang, W., J. Y. Lee and M. Nowotny (2006). "Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity." Mol Cell **22**(1): 5-13.

Yang, Z. L., K. M. Chen, R. R. Pandey, D. Homolka, M. Reuter, B. K. R. Janeiro, R. Sachidanandam, M. O. Fauvarque, A. A. Mccarthy and R. S. Pillai (2016). "PIWI Slicing and EXD1 Drive Biogenesis of Nuclear piRNAs from Cytosolic Targets of the Mouse piRNA Pathway." Mol Cell **61**(1): 138-152.

Ying, Y., X. M. Liu, A. Marble, K. A. Lawson and G. Q. Zhao (2000). "Requirement of Bmp8b for the generation of primordial germ cells in the mouse." Molecular Endocrinology **14**(7): 1053-1063.

Yoshida, S. (2010). "Stem cells in mammalian spermatogenesis." Development Growth & Differentiation **52**(3): 311-317.

Zhang, G., H. Huang, D. Liu, Y. Cheng, X. Liu, W. Zhang, R. Yin, D. Zhang, P. Zhang, J. Liu, C. Li, B. Liu, Y. Luo, Y. Zhu, N. Zhang, S. He, C. He, H. Wang and D. Chen (2015). "N6-methyladenine DNA modification in Drosophila." Cell **161**(4): 893-906.

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

Zuo, Y. and M. P. Deutscher (2001). "Exoribonuclease superfamilies: structural analysis and phylogenetic distribution." Nucleic Acids Res **29**(5): 1017-1026.

Zuo, Y. H. and M. P. Deutscher (2001). "Exoribonuclease superfamilies: structural analysis and phylogenetic distribution." Nucleic Acids Research **29**(5): 1017-1026.

Acknowledgments

I would like to thank Ramesh for giving me the opportunity to work in his lab. His passion and enthusiasm for science were always an inspiration and motivation. It was great to be in a lab with such a nice environment and great colleagues.

I would like to thank my PhD jury members, Prof. Winfried Weissenhorn, Prof. Dónal O'Carroll, Dr. Alena Shkumatava and Dr. André Verdel for having the time to read this thesis. Also my TAC committee: Dr. Stephen Cusack, Dr. André Verdel and Dr. François Spitz for their guidance and helpful discussions throughout my PhD. I would also like to extend these acknowledgements to EMBL PhD program and the Republic and Canton of Geneva for the financial support. EMBL core facilities were also very important for the accomplishment of this work. Thank you to all the people at EMBL Grenoble outstation, it was wonderful to have such a pleasant work environment.

Special thanks to my lab mates Pietro, Magda, Michael, Zhaolin, Raman, David, Kuan-Ming, Joanna, Mateuzs, Pascal and Fabienne for all the help during my PhD. We spend a lot of time together (in and out of the lab) and your support was very important, I learned a lot with you guys and I also complained a lot… thank you for having the patience to hear me!

Thanks a lot to my favorite flat mates Marjory and Jorge, who also listened very patiently to my complaints and cheered me up in the most difficult times… And all the people that I met along the way, Etienne, Miriam, Paul, Pietro, Lahari, Taianá, Magda, Duygu and many more… we have shared very nice moments together and you are the reason why this experience, which is not easy, is so worth it!

Um agradecimento especial à minha família que sempre me apoiou ao longo destes anos… estar longe não é fácil mas vocês sempre apoiaram as minhas escolhas e sempre me deram a força necessária para atingir os meus objectivos. Esta conquista também é vossa!

Ao João um obrigada especial por todo o apoio, pela paciência, por todo o carinho e amor… somos uma equipa e tenho a certeza que assim continuará, por um tempo muito longo e indeterminado!

Aos meus amigos… Um muito obrigado por todo o carinho ao longo destes anos, por todos os momentos partilhados. Já la vão uns aninhos e a amizade continua e espero que assim perdure!

Appendix

Characterization of the mammalian RNA exonuclease 5/NEF-sp as a testis-specific nuclear 3′ → 5′ exoribonuclease

SARA SILVA,[1,2] DAVID HOMOLKA,[1] and RAMESH S. PILLAI[1]

[1]Department of Molecular Biology, University of Geneva, CH-1211 Geneva 4, Switzerland
[2]European Molecular Biology Laboratory, Grenoble Outstation, 38042, France

ABSTRACT

Ribonucleases catalyze maturation of functional RNAs or mediate degradation of cellular transcripts, activities that are critical for gene expression control. Here we identify a previously uncharacterized mammalian nuclease family member NEF-sp (RNA exonuclease 5 [REXO5] or LOC81691) as a testis-specific factor. Recombinant human NEF-sp demonstrates a divalent metal ion-dependent 3′ → 5′ exoribonuclease activity. This activity is specific to single-stranded RNA substrates and is independent of their length. The presence of a 2′-*O*-methyl modification at the 3′ end of the RNA substrate is inhibitory. Ectopically expressed NEF-sp localizes to the nucleolar/nuclear compartment in mammalian cell cultures and this is dependent on an amino-terminal nuclear localization signal. Finally, mice lacking NEF-sp are viable and display normal fertility, likely indicating overlapping functions with other nucleases. Taken together, our study provides the first biochemical and genetic exploration of the role of the NEF-sp exoribonuclease in the mammalian genome.

Keywords: NEF-sp; LOC81691; Q96IC2; REXON; RNA exonuclease 5; REXO5; 2610020H08Rik

INTRODUCTION

Spermatogenesis is the process by which sperm cells are produced in the male germline. Early events in this process include differentiation of embryonic primordial germ cells (PGCs) to form prospermatogonia, and later spermatogonia and spermatogonial stem cells (SSCs) (Hilscher et al. 1974). These undergo multiple rounds of amplification via mitotic divisions, which end in production of spermatocytes that enter meiosis (Bellvé et al. 1977). The spermatocytes complete meiosis to generate haploid round spermatids that enter a differentiation process called spermiogenesis which results in compaction of chromatin by removal of histones and their replacement with protamines. This is accompanied by other cellular events like formation of the acrosome that contains enzymes essential for fertilization of the egg, and formation of the flagella required for motility. Given the progressive chromatin compaction and its unavailability for transcription, some RNA products are already generated early in the spermatocytes and stored in cytoplasmic granules for translation later during spermiogenesis (Eddy 1995). Thus, RNA processing events are critical for successful completion of spermatogenesis (Dadoune et al. 2004).

Ribonucleases play an essential role in control of gene expression. All messenger RNAs (mRNAs) require ribonuclease-mediated processing to create their final 3′ ends: poly(A) tails of most mRNAs or the hairpin structure of replication-dependent histone mRNAs (Colgan and Manley 1997; Schümperli and Pillai 2004). Functional noncoding RNAs like rRNAs, tRNAs, miRNAs, piRNAs, etc. are produced as precursors that need additional ribonuclease cleavage events to produce mature RNA molecules (Allmang et al. 2000; Ghildiyal and Zamore 2009). Ribonucleases also act in important quality control steps like nonsense-mediated decay (NMD) where mistakes in transcript production that lead to premature stop codons are recognized, and such transcripts are eliminated (Mühlemann and Jensen 2012). Once these RNAs have served their purpose, removal by degradation also requires a number of ribonucleases, a step that is essential in maintenance of cellular homeostasis (Makino et al. 2013).

In our search for testis-specific nucleases, we identified the previously uncharacterized mammalian nuclease RNA exonuclease 5 (REXO5 or NEF-sp) be specifically enriched in testicular transcriptomes. We demonstrate that *Nef-sp* is exclusively expressed in mouse testes. Recombinant human NEF-sp (hNEF-sp) protein is a 3′ → 5′ exoribonuclease that acts on single-stranded RNA in a distributive manner. Modification of the RNA 3′ end is inhibitory for this activity. Ectopically expressed hNEF-sp localizes to the nucleolus in

Corresponding author: ramesh.pillai@unige.ch
Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.060723.117. Freely available online through the *RNA* Open Access option.

human cell cultures and this is dependent on an amino-terminal nuclear localization signal. In contrast, mouse NEF-sp is uniformly nuclear in both mouse and human cell cultures. The *Nef-sp* knockout mice are viable and fertile, indicating potential complementation by other nuclease(s). This study sheds light on the biochemical and physiological relevance of the NEF-sp exoribonuclease.

RESULTS

NEF-sp is a testis-specific protein

A search for potential nucleases in testis transcriptomes led to the identification of human NEF-sp (accession no. NP_001185982) as an uncharacterized protein. It is variously known in the databases as LOC81691 or Q96IC2_REXON or 2610020H08Rik or RNA exonuclease 5 (REXO5), but in this report we will refer to it as NEF-sp. We first performed reverse transcription PCR (RT-PCR) analysis using total testes RNA (Fig. 1A). Similar to the testis-specific piRNA pathway factor *PIWIL2/Mili* (Kuramochi-Miyagawa et al. 2001), the *Nef-sp* transcript was detected in both human and mouse testes. Analysis of total RNA from various mouse tissues indicates exclusive expression of *Nef-sp* in the testes (Fig. 1B). We failed to detect the transcript in commonly used human cell cultures like HeLa or HEK293T (Fig. 1B). To study it further, we cloned complementary DNAs (cDNAs) for *Nef-sp* from human and mouse testicular total RNA. The cloned human cDNA (accession no. NP_001185982) encodes for a 774 amino acid (aa) human NEF-sp (hNEF-sp) protein that is composed of an amino-terminal nuclease domain followed by two carboxy-terminal tandem RNA-recognition motifs (RRMs) (Fig. 1C). We identified two isoforms of the mouse *Nef-sp*: a full-length long isoform encoding for a protein of 784 aa (accession no. KY853396), and a shorter carboxy-terminal splice variant encoding for a protein of 696 aa (accession no. KY853397). Both the mouse proteins have the same domain architecture, but the shorter isoform carries changes downstream from the RRM domains (see Materials and Methods).

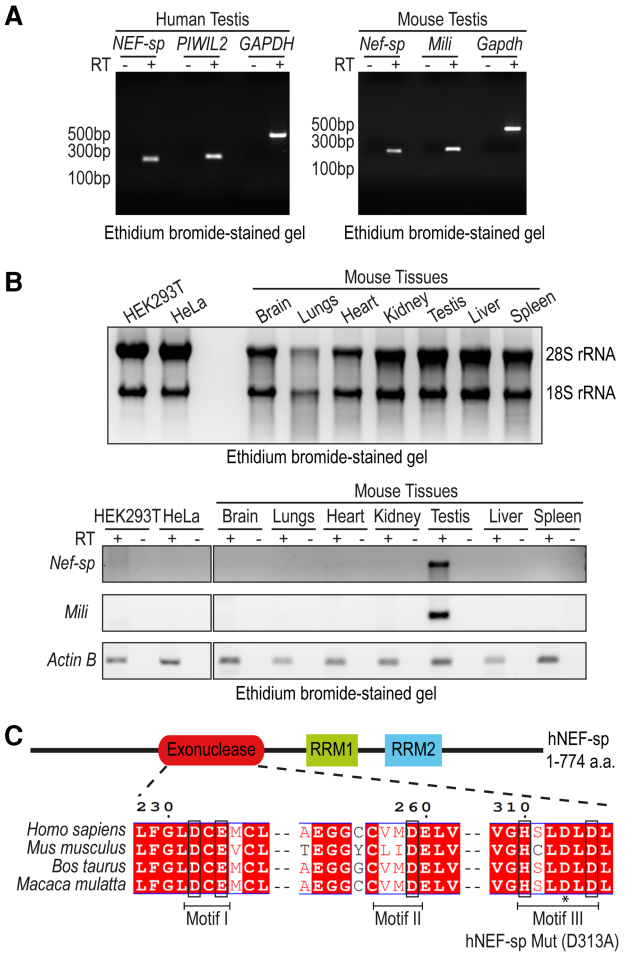


FIGURE 1. Mouse *Nef-sp* is exclusively expressed in the testes. (*A*) *Nef-sp* transcript is detected in total RNA from human and mouse testes using reverse transcription PCR (RT-PCR). Testis-specific human *PIWIL2* or mouse *Piwil2/Mili* and ubiquitously expressed *Gapdh* serve as controls. (*B*) Quality of total RNA isolated from human cell cultures (HeLa and HEK293) and different mouse tissues is indicated by ethidium bromide staining to reveal the 28S and 18S rRNAs (*top* panel). RT-PCR analysis of *Nef-sp*, *Mili*, and *Actin B* in the different RNA samples (*lower* panel). Reactions carried out without (−) and with (+) reverse transcriptase (RT) are shown. (*C*) Domain organization of human NEF-sp (hNEF-sp) showing the nuclease domain and two RNA-recognition motifs (RRMs). The conserved residues DEDDh (in boxes) distributed in three motifs, and the point mutation (D313A) introduced in motif III, are indicated.

NEF-sp is a divalent metal ion-dependent 3′ → 5′ exonuclease

To examine its biochemical properties, we produced a recombinant version of the full-length human NEF-sp with an amino-terminal 6xHis-Strep-SUMO fusion tag (Fig. 2A). This tagged protein alone was incubated with a 5′-end labeled 40-nt single-stranded RNA (ssRNA) (Table 1), and the reaction products were resolved by 15% denaturing urea polyacrylamide gel electrophoresis (PAGE). This did not reveal any activity (Fig. 2B). Supplementing the reaction with divalent metal ions (Mg^{2+} or Mn^{2+}) resulted in appearance of an RNA ladder, indicative of nuclease activity (Fig. 2B). We note that even when present at lower concentrations, the Mn^{2+} ion (2.5 mM) showed a stronger stimulation of the nuclease activity compared with the Mg^{2+} ion (25 mM). This is likely due to the relaxed coordination properties of Mn^{2+}, compared with the stringent coordination geometry requirement of Mg^{2+} (Yang et al. 2006). The Mg^{2+}-dependent exonuclease activity of human NEF-sp is completely blocked by incubation with the Mg^{2+}-chelating agent EDTA (Fig. 2B). Given that the RNA substrates are 5′-end labeled, we conclude that human NEF-sp has 3′ → 5′ exonuclease activity, with the nuclease acting on all the available ssRNA substrates and degrading it in a distributive manner, giving rise to the ladder-like pattern of degradation products.

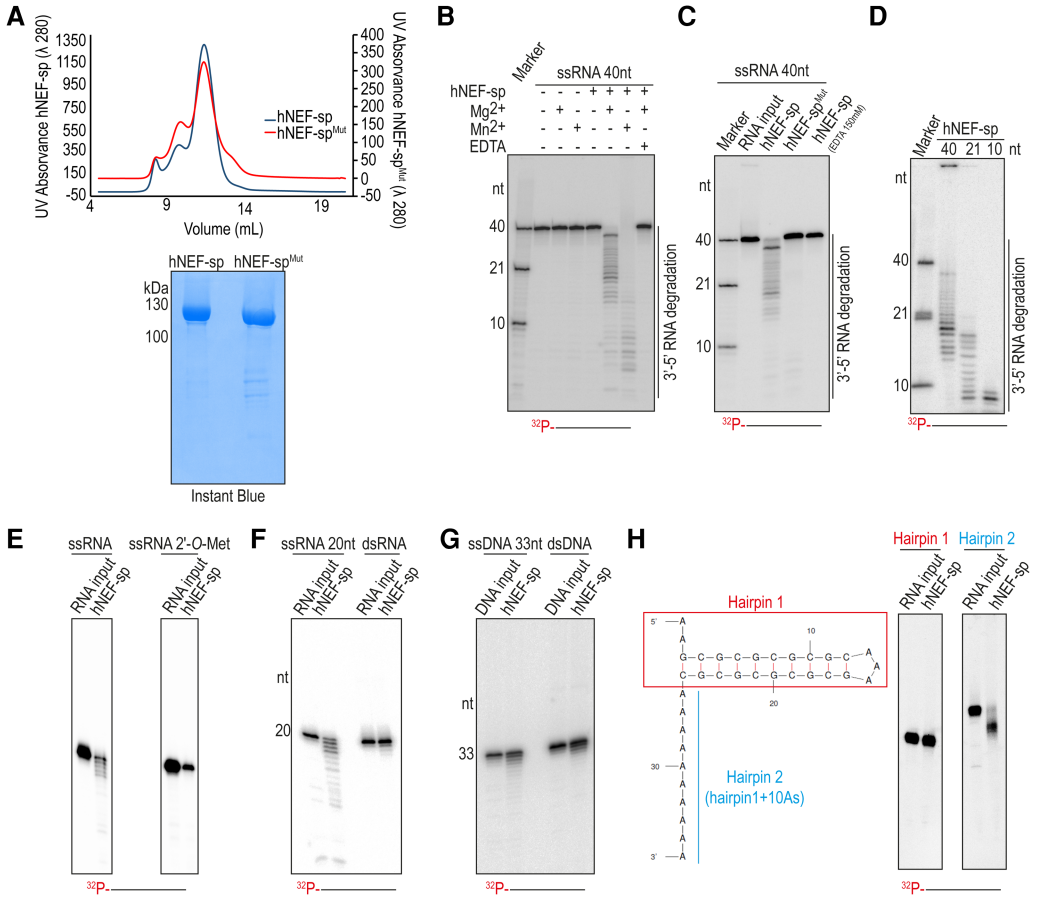


FIGURE 2. Human NEF-sp is 3′ → 5′ exoribonuclease that acts on single-stranded RNAs. (*A*) Quality of recombinant human NEF-sp wild-type and point mutant (D313A; hNEF-sp^{Mut}) proteins used for nuclease assays as visualized by Instant Blue staining. The size-exclusion chromatography profiles for the proteins are shown. The identical profiles point to an absence of dramatic structural changes in the mutant protein. (*B*) Nuclease assay with a 5′-end labeled (with [^{32}P-γ]ATP) 40-nt single-stranded RNA (ssRNA) and wild-type hNEF-sp protein. (*C*) Nuclease assays with wild-type and point-mutant (hNEF-sp^{Mut}) hNEF-sp proteins. (*D*) Nuclease assay with ssRNA substrates of different lengths (40, 21, and 10 nt). The 10-nt RNA substrate is still digested down to at least 8–7 nt fragments. (*E*) Nuclease assay with an ssRNA carrying a 3′ terminal 2′-*O*-methyl modification. The RNA substrate is not degraded by NEF-sp (*right* panel), while the unmodified ssRNA with the same sequence is digested (*left* panel). (*F*) Reactions with 20-nt ssRNA or a 20-bp double-stranded RNA (dsRNA). Only the ssRNA is digested. (*G*) Reactions with 33-nt ssDNA or a 33-bp dsDNA indicating lack of activity to DNA substrates. (*H*) Nuclease assay using hairpin RNAs without (Hairpin 1) and with (Hairpin 2) a 3′ overhang. The structure of the hairpins as predicted by the RNA folding program Mfold is shown. Note the digestion of the single-stranded 3′ extension in Hairpin 2, but not of the double-stranded region. This experiment also shows lack of endonuclease activity on the single-stranded loop of the hairpin. Products are resolved in a 15% urea polyacrylamide gel. Markers are 5′-end labeled ssRNAs and length in nucleotides (nt) are indicated.

Protein sequence analysis of NEF-sp identifies it as a member of the DEDD family of nucleases with the presence of four invariant acidic residues (D–E–D–D) distributed in three sequence motifs (Fig. 1C; Zuo and Deutscher 2001). The presence of a histidine in the motif III [H-x(4)-D] further identifies NEF-sp as belonging to the DEDDh subfamily. In other family members, these conserved residues are essential for coordination of two divalent metal ions required for catalytic activity. We prepared a point mutant (hNEF-sp^{Mut}) version of the protein carrying a single amino acid substitution (D313A) within the motif III. The purified mutant protein behaves identical to the wild-type protein during size-exclusion chromatography, indicating lack of any gross conformational changes (Fig. 2A). Importantly, this mutation inactivates the enzyme, confirming that the nuclease activity we observe is inherent to hNEF-sp (Fig. 2C). Taken together, these results show that human NEF-sp is a divalent metal ion-dependent 3′ → 5′ exonuclease.

Human NEF-sp acts on single-stranded RNAs

To examine the specificity of the enzyme, we incubated it with different nucleic acid substrates. First, we probed the accessibility of the catalytic pocket to single-stranded RNAs of different lengths. A 40-nt ssRNA was very efficiently used as a substrate, as already observed above. Similarly, a 21-nt and a 10-nt RNA were also used as substrates (the lane marked "marker" shows the actual undigested substrates used) (Fig. 2D). The 10-nt RNA was digested down to at least 7 nt, indicating a minimum length requirement for the ssRNA

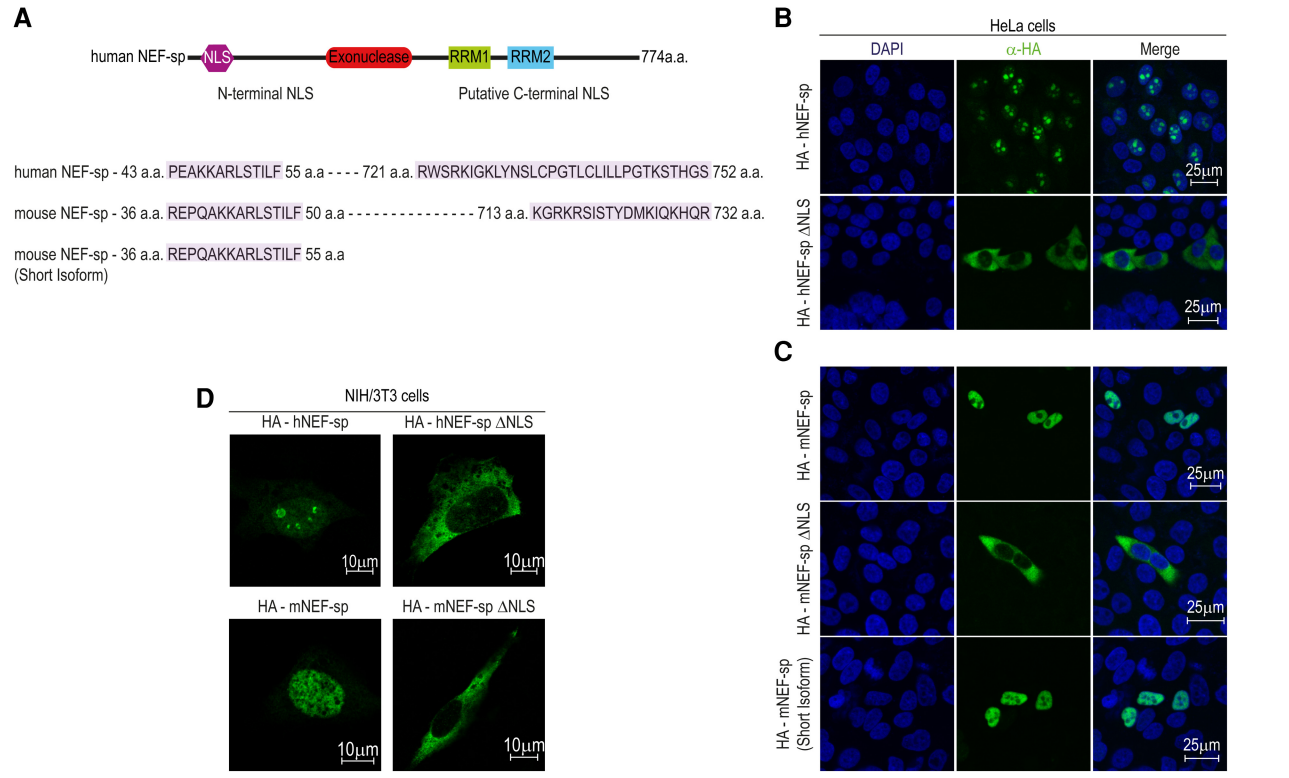


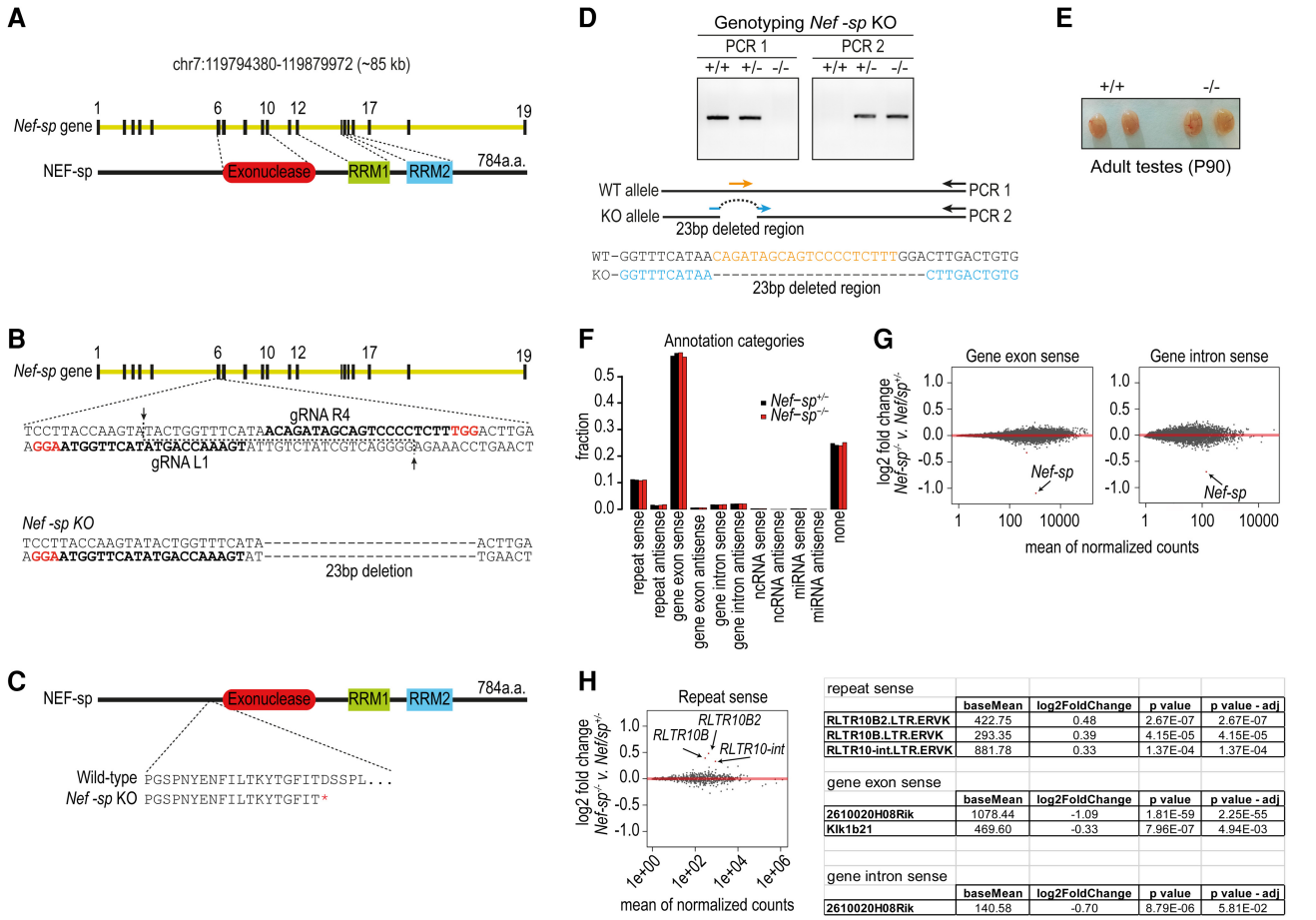
FIGURE 3. Human NEF-sp localizes to the nucleolus, while mouse NEF-sp is nuclear in human cell cultures. (*A*) Cartoon showing domain organization of hNEF-sp. The nuclear localization signal (NLS) is indicated. The predicted NLS sequences are shown *below* and amino acid (aa) boundaries are indicated. Only the amino-terminal NLS is experimentally proven to be functional. (*B*,*C*) Immunofluorescence analyses of HA-tagged proteins transiently expressed in human HeLa cell cultures. The signal from detection of the HA-tagged protein (green) and DNA with DAPI staining (blue) is shown. A merged image is provided. (*D*) Immunofluorescence detection in transfected mouse NIH3T3 cells. Scale bar in micrometers is shown.

substrate to reach into the catalytic pocket of the enzyme. The 2′-*O*-methyl modification on the 3′ terminal nucleotide of RNAs is proposed to act as a stabilizing mark by protecting them from exonucleases (Ji and Chen 2012). When hNEF-sp was incubated with ssRNA carrying such a 3′ end modification, the RNA remained undigested (Fig. 2E). This shows that 2′-*O*-methyl modification of the RNA at the 3′ end hinders the activity of hNEF-sp.

Next, we examined whether double-stranded RNA (dsRNA) could be used as a substrate. When a 20-nt ssRNA was provided as a substrate, it was cleaved by the NEF-sp enzyme (Fig. 2F). However, we did not observe any activity when this sequence was preannealed with its complementary strand to make a 20-bp dsRNA duplex prior to incubation with the enzyme (Fig. 2F). This suggests that dsRNAs are not accommodated within the catalytic pocket of the enzyme. Furthermore, DNA was also a poor substrate for the enzyme in both its single-stranded and double-stranded forms (Fig. 2G). When offered structured RNAs, hNEF-sp left a 25-nt hairpin RNA (Hairpin 1) intact (Fig. 2H). However, when incubated with the same hairpin RNA, but now carrying a 3′ overhang composed of 10As (Hairpin 2), hNEF-sp was able to degrade the single-stranded region, leaving the structured part untouched (Fig. 2H). Furthermore, the above experiment also shows that hNEF-sp is not an endonuclease, as the single-stranded loop region was not cleaved. Taken together, we show that hNEF-sp is an Mg^{2+}-dependent 3′ → 5′ exonuclease acting on ssRNAs and is inhibited by the presence of a 2′-*O*-methyl modification at the 3′ end of the substrate.

Human NEF-sp is nucleolar while mouse NEF-sp is nuclear

To study its subcellular localization, we transiently expressed HA-tagged versions of NEF-sp in human HeLa cell cultures (Fig. 3A). The full-length HA-hNEF-sp protein efficiently accumulated in the nucleolus of transfected cells (Fig. 3B). Computational search for potential nuclear localization signal (NLS) sequences identified two such elements: one at the amino terminus and another at the carboxyl terminus (Fig. 3A). Of these, the amino-terminal signal is critical for nucleolar accumulation, as its deletion restricts the hNEF-$sp^{\Delta NLS}$ version to the cytoplasm (Fig. 3B). Also, a large-scale analysis of subcellular localization of human proteins revealed that hNEF-sp correctly localizes to the nucleolar compartment in a monkey cell line (Simpson et al. 2000). Curiously, although the HA-tagged mouse NEF-sp (mNEF-



repeat sense

	baseMean	log2FoldChange	p value	p value - adj
RLTR10B2.LTR.ERVK	422.75	0.48	2.67E-07	2.67E-07
RLTR10B.LTR.ERVK	293.35	0.39	4.15E-05	4.15E-05
RLTR10-int.LTR.ERVK	881.78	0.33	1.37E-04	1.37E-04

gene exon sense

	baseMean	log2FoldChange	p value	p value - adj
2610020H08Rik	1078.44	-1.09	1.81E-59	2.25E-55
Klk1b21	469.60	-0.33	7.96E-07	4.94E-03

gene intron sense

	baseMean	log2FoldChange	p value	p value - adj
2610020H08Rik	140.58	-0.70	8.79E-06	5.81E-02

FIGURE 4. *Nef-sp* knockout mice are viable and fertile. (*A*) Gene organization of mouse *Nef-sp* and mNEF-sp protein domain architecture. (*B*) Guide RNAs (gRNAs) used for generation of a 23-bp deletion in the mouse *Nef-sp* gene locus of single-cell embryos. (*C*) Premature translation stop codon (D220*) introduced by the deletion leading to a truncation of the protein before the nuclease domain. (*D*) Genotyping of the wild-type, *Nef-sp*$^{+/-}$ and *Nef-sp*$^{-/-}$ mice. Cartoon showing the set of primers (represented by arrows) used to distinguish the wild-type allele from the knockout allele. See Materials and Methods for details. (*E*) Representative mouse testes from wild-type and mutant Nef-sp animals showing normal testes size. Mutant mice show normal fertility. (*F*) Analysis of the total testicular transcriptome from heterozygous and homozygous *Nef-sp* knockout mutant mice. Overall genome annotations of reads do not show any changes. (*G*) Comparison of read counts of sense-oriented gene exonic and intronic reads reveal down-regulation of *Nef-sp* transcript, as expected. No other significant change in transcripts is noted. (*H*) Analysis of sense-oriented repeat reads shows an up-regulation of the indicated repeat elements. Table showing fold-changes in read counts of sense-oriented repeat elements and *Nef-sp* transcript.

sp) shows nuclear localization, it was clearly excluded from the nucleoli in human HeLa cell cultures (Fig. 3C). We tested localization of mNEF-sp in mouse NIH3T3 cells and the protein was still accumulated in the nucleus while being excluded from the nucleolus, pointing to species-specific differences (Fig. 3D). Similar to the human protein, deletion of the amino-terminal NLS retains the mNEF-sp$^{\Delta NLS}$ version in the cytoplasm. The short isoform of mNEF-sp which lacks the predicted carboxy-terminal NLS (Fig. 3A) was still efficiently imported into the nucleus (Fig. 3C). Thus we conclude that the amino-terminal NLS is necessary and sufficient for nuclear/nucleolar import of NEF-sp. Our attempts to examine endogenous mNEF-sp in mouse testes with anti-mNEF-sp antibodies was unsuccessful. Taken together, we propose that NEF-sp is an exoribonuclease that may function in the nuclear/nucleolar compartment.

Analysis of *Nef-sp* knockout mouse

To study the in vivo role of NEF-sp, we introduced a deletion in the *Nef-sp* gene locus in single-cell mouse embryos using the RNA-guided Cas9 genome-editing enzyme (Fig. 4A,B). The generated mutant mice have a 23 bp deletion in the coding sequence, resulting in a premature stop codon (D220*) upstream of the nuclease domain (Fig. 4C,D). We will refer to this as the *Nef-sp* knockout allele (*Nef-sp*$^{-}$). Homozygous *Nef-sp*$^{-/-}$ males displayed normal testes sizes when compared to wild-type animals (Fig. 4E). We report that homozygous *Nef-sp*$^{-/-}$ animals of both sexes display normal fertility and viability. Deep sequencing of testicular total RNAs from wild-type and homozygous knockout mice did not reveal any dramatic changes in transcripts belonging to the different annotation categories (Fig. 4F).

TABLE 1. RNA and DNA oligonucleotides used for the characterization of human NEF-sp nuclease activity

ssRNA1	5′-GGCGAGAAAGCUAUCUGAGCACCUGUGUUCAUGUCAGCAU-3′ (40 nt)
ssRNA2	5′-UGACAUGAACACAGGUGCUCA-3′ (21 nt)
ssRNA3	5′-UGACAUGAAC-3′ (10 nt)
ssRNA4	5′-UGACAUGAACACAGGUGCUC-3′ (20 nt)
ssRNA5	5′-GAGCACCUGUGUUCAUGUCA-3′ (20 nt)
ssRNA 2-*O*-Met_1	5′-UGACAUGAACACAGGUGCUCA-3′ (21 nt) same sequence as ssRNA2
ssDNA1	5′-CAGGTTTGACCAGTTCTGCCATAACACAGCAGC-3′ (33 nt)
ssDNA2	5′-GCTGCTGTGTTATGGCAGAACTGGTCAAACCTG-3′ (33 nt)
Hairpin 1	5′-AAGCGCGCGCGCAAAGCGCGCGCGC-3′ (25 nt)
Hairpin 2	5′-AAGCGCGCGCGCAAAGCGCGCGCGCAAAAAAAAAA-3′ (35 nt)

Closer examination of genic cellular mRNAs showed unchanged levels, except for a dramatic reduction in *Nef-sp* transcript levels in the knockout mutant as expected from the deletion (Fig. 4G). Examination of repeat-associated transcripts reveals a modest increase in RLTR10B, RLTR10B2, and RLTR10-int transposon sequences (Fig. 4H). We did not observe any other changes in these sequenced transcriptomes. Taken together with the fact that the knockout mutants are fertile, this indicates that loss of NEF-sp is tolerated, perhaps via complementation by an unknown nuclease(s).

DISCUSSION

Here we examined the biochemical properties of the uncharacterized mammalian nuclease family member NEF-sp. We show that *Nef-sp* transcripts are detected exclusively in mouse testes. Our efforts to detect the endogenous protein in mouse testes extracts and visualize it in testes sections failed. Since our rabbit polyclonal antibodies detect transiently expressed HA-mNEF-sp in human cell cultures (data not shown), we conclude that perhaps abundance of the endogenous protein is below the detection limit of our antibodies. We demonstrate that hNEF-sp is a 3′ → 5′ exoribonuclease that is active on ssRNA substrates and likely functions in the nucleolar compartment. The importance of the two RNA-recognition motifs (RRMs) in the protein was not examined in this study, but these may perhaps help identify certain RNA substrates. Alternatively, potential protein complexes formed by the enzyme might help select the RNA targets. However, in vitro nuclease assays used RNAs with a variety of sequences and these were all efficiently processed.

Human NEF-sp is a highly conserved protein, with orthologs having a similar domain architecture being detected in amphibians, reptiles, birds, and mammals. Although proteins bearing similarity within the region encompassing the nuclease domain of NEF-sp are detected in insects, they lack the two tandem RRMs. A similar situation exists in the budding yeast *Saccharomyces cerevisiae*, where RNA exonuclease 1 (Rex1p)/Rnh70p/yGR276 (van Hoof et al. 2000) bears similarity to the nuclease domain of NEF-sp. Rex1p is a nuclear protein that is shown to use its 3′ → 5′ exonuclease activity to trim the extended 3′ end of precursors of 5S rRNA, $tRNA^{Arg}$, and $tRNA^{Met}$ (van Hoof et al. 2000; Ozanick et al. 2009). 5S rRNA is transcribed by RNA polymerase III as precursors with a 3′ extension. Such extended 5S rRNA or tRNA precursors accumulate in the *REX1* mutant yeast (van Hoof et al. 2000) and also become polyadenylated by nuclear surveillance machinery (Ozanick et al. 2009). However, this maturation process mediated by Rex1p is not essential for viability. In fact, the presence of the 3′-extended 5S rRNA precursors do not seem to affect the maturation of rRNA in ribosomes (Nariai et al. 2005), as sufficient mature rRNA and tRNA molecules get made due to the action of a number of redundant nucleases (van Hoof et al. 2000; Ozanick et al. 2009).

Our own analysis of the *Nef-sp* mouse knockout mutant revealed no obvious phenotype. Testis transcriptome analysis did not reveal any dramatic change in RNA species, except for the *Nef-sp* transcript itself (Fig. 4G). Some repeat transcripts were up-regulated in the mutant, but failure to silence them seems to have no consequence for mouse fertility (Fig. 4H). It is possible that precise trimming of the unknown substrate(s) may not be important for viability/fertility due to potential complementation by other nucleases. Nevertheless, our study provides the first biochemical and genetic characterization of the mammalian NEF-sp exoribonuclease.

MATERIALS AND METHODS

Clones and constructs

The coding sequences for human (LOC8169) and mouse (2610020H08Rik; MGI:1919402) RNA exonuclease 5 (Rexo5) or *Nef-sp* were amplified by reverse transcription PCR (RT-PCR) from adult human testes total RNA (Clontech, cat no. 636533) or mouse testes RNA, respectively. GenBank accession numbers are human *NEF-sp* (NM_001199053), mouse *Nef-sp* long isoform (KY853396), and mouse short isoform (KY853397). For mammalian cell culture expression, the following coding sequences were cloned into the SalI and NotI sites of the vector pCI-neo-N-HA (Pillai et al. 2005): hNEF-sp (1–774 aa), hNEF-$sp^{\Delta NLS}$ (58–774 aa), mNEF-sp (1–784 aa), mNEF-$sp^{\Delta NLS}$ (53–784 aa), and mNEF-sp short isoform (1–696 aa).

Cell culture and antibodies

Human epithelioid cervix carcinoma (HeLa) and mouse NIH/3T3 cell lines growing on coverslips in 12-well plates were transfected with 2.5 µg of expression plasmids using X-tremeGENE HP (DNA Transfection Reagent, Roche; 06366236001). Subcellular

localization of HA-tagged proteins was examined using the anti-HA as primary antibody, followed with staining by secondary antibody anti-mouse-IgG coupled Alexa Fluor 488 (Invitrogen). Cells were examined under an inverted confocal microscope Leica TCS SP2 AOBS.

Recombinant human NEF-sp production

For production of recombinant proteins using the eukaryotic expression systems we used the following ovary-derived insect cells: Sf21 from the Fall Army worm *Spodoptera frugiperda* or High Five (Hi5) from the cabbage looper, *Trichoplusia ni*. The full-length human NEF-sp (1–774 aa) or a mutant version (hNEF-sp^{Mut}) carrying a point mutation (D313A) were cloned into the NheI and KpnI sites of the vector pACEBac2 (6xHis-Strep-SUMO-TEV-fusion) (Bieniossek et al. 2012). Proteins were purified by Ni^{2+}-affinity chromatography in lysis buffer (25 mM Tris, pH 7.5, 500 mM NaCl, 5 mM of β-mercaptoethanol, 0.5% Tween, 20 mM imidazole, 5 mM $MgCl_2$, 10% of glycerol) supplemented with protease inhibitor (Roche; Complete EDTA-free). The protein was further purified on a Heparin column (HI-Trap; GE Healthcare) and monodisperse fractions collected by gel filtration chromatography (25 mM Tris, pH 7.5, 500 mM NaCl, 5 mM of β-mercaptoethanol, 0.5% Tween, and 10% of glycerol) (Superdex 200; GE healthcare). Proteins were used in nuclease assays without removal of the tags.

Nuclease assay

The ssRNA (Microsynth) and ssDNA (Invitrogen) substrates (see list below) were labeled at the 5′ end with [γ-^{32}P]ATP (PerkinElmer; NEG002A001MC) and T4 polynucleotide kinase (Thermo Scientific; EK0031). The labeled probes were resolved in 15% denaturing urea–polyacrylamide gels, and full-length labeled RNAs were gel-eluted with 0.3 M NaCl followed by phenol–chloroform extraction, precipitation and resuspension in 20 µL of RNase-free water.

RNA and DNA duplexes were made as described: 100 µM of the top strand was labeled at the 5′ end with [γ-^{32}P]ATP and resolved in 20% in denaturing urea–polyacrylamide gels. The full-length labeled oligonucleotides were eluted in 600 µL of elution buffer (300 mM NaOAc, 1 mM EDTA, and 0.5% SDS) overnight at 4°C followed by precipitation using 1.8 mL of 100% ethanol and 2 µL of glycogen. The pellets were resuspended in 16 µL of water and annealing reactions were performed using 2 µL of 100 µM unlabeled complementary strand and 2 µL of 10× duplex-annealing buffer (100 mM MOPS, pH 6.5, 10 mM EDTA, and 0.5 M KCl) followed by heating the samples to 95°C and gradually cooled down to room temperature. The oligonucleotide duplexes were resolved on 15% nondenaturing PAGE followed by gel elution and precipitation as described above.

We mixed 1 µL of RNA/DNA substrate with 1 µM of purified recombinant human NEF-sp proteins (wild-type or point mutant [D313A]) and incubated at 37°C for 1 h in buffer (25 mM Tris–HCl, pH 7.5, 150 mM NaCl, 2 mM DTT). Divalent metal ions needed for the reaction were provided by supplementing the reaction with 25 mM $MgCl_2$ or 2.5 mM $MnCl_2$. When required, 150 mM EDTA was added to chelate Mg^{2+} ions. Treatment with Proteinase K was used to terminate the reaction, followed by phenol–chloroform extraction of the RNA and precipitation with ethanol. The samples were resolved by electrophoresis on 15% denaturing urea–polyacrylamide gels and exposed overnight to a Phosphor Storage screen (GE Healthcare Life Sciences). Screens were scanned using a Typhoon Scanner (GE Healthcare Life Sciences). ssRNA1 was used for the nuclease assays of Figure 2, B and C. Figure 2D was made using ssRNA1, ssRNA2, and ssRNA3. The nuclease assay with 2-*O*-Met modification at the 3′ end was done using ssRNA2 (without modification) and ssRNA 2-*O*-Met_1 (Fig. 2E). dsRNA was obtained by annealing ssRNA4 and ssRNA5 (Fig. 2F), and dsDNA by annealing ssDNA1 and ssDNA2 (Fig. 2G). Hairpin 1 (without overhangs) and Hairpin 2 (with 10-nt 3′ overhang) were used for the nuclease assays of Figure 2H.

Mouse mutant generation

The coding sequence in the *Nef-sp* locus of the mouse genome was disrupted using the RNA-guided Cas9 endonuclease (Cas9 nickase [Cas9n]) and a pair of guide RNAs (gRNAs) (Mouse Biology Program, University of California, Davis). The *Cas9n* mRNA and the in vitro transcribed gRNAs (gRNA L1 and R4) were used for injection into C57B1/6J host embryos for mouse generation using one-cell stage injection (Fig. 4B). The founder females carrying a 23-bp out-of-frame deletion were crossed with wild-type C57Bl/6J Rj males (Janvier Labs, France) to establish the mutant line.

Tail genomic DNA isolation and genotyping PCR conditions

Tails were digested in buffer (50 mM Tris–HCl, pH 8.0, 100 mM EDTA, 100 mM NaCl, 1% SDS with 25 µg of Proteinase K) at 55°C overnight. DNA was precipitated with isopropanol, washed in 70% (v/v) ethanol and resuspended in 10 mM Tris–HCl, pH 8.0. To identify the F1 animals we used a pair of primers, the forward primer (represented in blue) complementary to the region spanning the deleted region and the reverse primer around 500-bp downstream from the deletion (Fig. 4D). To distinguish between heterozygous and homozygous mice, an additional PCR was done using a forward primer that only anneals to the wild-type allele (represented in orange) (Fig. 4D).

Genotyping primers: SS323 (5′-GGTTTCATAACTTGAC TGTG-3′), SS330 (5′-CAGATAGCAGTCCCCTCTTT-3′), and SS336 (5′-GGATTACGGAAACTCAAAGTG-3′). The PCR reactions were performed in 1× Taq Buffer (Fermentas), 200 µM dNTPs, 0.25 µM primers, 2.5 µM of $MgCl_2$, 2 µL of DNA template, and 0.5 µL of Taq DNA polymerase in a final volume of 25 µL. PCR conditions were 95°C for 5 min, then 40 cycles of 30 sec at 95°C, 30 sec at 60°C and 30 sec at 72°C, and a final step at 72°C for 5 min.

Total RNA libraries and bioinformatics

Total testicular RNA was isolated by TRIzol Reagent (Invitrogen cat. no. 15596026). Strand-specific RNA-seq libraries were prepared using the TruSeq Stranded Total RNA Sample Preparation Kit (Illumina) after removal of abundant ribosomal RNAs with Ribo-Zero (mouse). Libraries were sequenced with the Illumina HiSeq 2000 platform (EMBL Heidelberg Gene Core facility) for 50 cycles.

Reads were sorted into individual libraries based on the barcodes; the 3′ adapter sequences were removed and mapped to the mouse genome (mm9). The software used for processing the data (genomic coordinates, etc.) from the raw data files are in-house tools

developed by the Sachidanandam laboratory (Olson et al. 2008). Only reads perfectly matching the genome were kept for further analysis. The reads were divided into groups based on their annotation (repeat sense, repeat antisense, gene exon sense, gene exon antisense, gene intron sense, gene intron antisense, ncRNA sense, ncRNA antisense, miRNA sense, miRNA antisense, none) and their counts were compared between *Nef-sp*$^{-/-}$ and *Nef-sp*$^{+/-}$. DESeq2 bioconductor package (Love et al. 2014) was used to search for differentially expressed transcripts. An adjusted *P*-value of 0.05 was used as a threshold for statistical significance.

RNA secondary structures were predicted using the Mfold web server (Zuker 2003).

DATA DEPOSITION

Deep sequencing data sets generated in this study are deposited with the Gene Expression Omnibus (GEO) under accession number GSE97235.

ACKNOWLEDGMENTS

We thank the Genomics core facility of EMBL. S.S. received a fellowship from the EMBL PhD program. This work was supported by grants from the Swiss National Science Foundation: ERC Transfer Grant (GermMethylation), Project Grant (Origin-of-pi), and NCCR RNA & Disease. Funding is also acknowledged from the National Institutes of Health (1R01HD069592-01A1) to R.S.P. Work in the Pillai laboratory is supported by the Republic and Canton of Geneva.

Received January 9, 2017; accepted May 22, 2017.

REFERENCES

Allmang C, Mitchell P, Petfalski E, Tollervey D. 2000. Degradation of ribosomal RNA precursors by the exosome. *Nucleic Acids Res* **28:** 1684–1691.

Bellvé AR, Cavicchia JC, Millette CF, O'Brien DA, Bhatnagar YM, Dym M. 1977. Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization. *J Cell Biol* **74:** 68–85.

Bieniossek C, Imasaki T, Takagi Y, Berger I. 2012. MultiBac: expanding the research toolbox for multiprotein complexes. *Trends Biochem Sci* **37:** 49–57.

Colgan DF, Manley JL. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes Dev* **11:** 2755–2766.

Dadoune JP, Siffroi JP, Alfonsi MF. 2004. Transcription in haploid male germ cells. *Int Rev Cytol* **237:** 1–56.

Eddy EM. 1995. 'Chauvinist genes' of male germ cells: gene expression during mouse spermatogenesis. *Reprod Fertil Dev* **7:** 695–704.

Ghildiyal M, Zamore PD. 2009. Small silencing RNAs: an expanding universe. *Nat Rev Genet* **10:** 94–108.

Hilscher B, Hilscher W, Bülthoff-Ohnolz B, Krämer U, Birke A, Pelzer H, Gauss G. 1974. Kinetics of gametogenesis. I. Comparative histological and autoradiographic studies of oocytes and transitional prospermatogonia during oogenesis and prespermatogenesis. *Cell Tissue Res* **154:** 443–470.

Ji L, Chen X. 2012. Regulation of small RNA stability: methylation and beyond. *Cell Res* **22:** 624–636.

Kuramochi-Miyagawa S, Kimura T, Yomogida K, Kuroiwa A, Tadokoro Y, Fujita Y, Sato M, Matsuda Y, Nakano T. 2001. Two mouse *piwi*-related genes: *miwi* and *mili*. *Mech Dev* **108:** 121–133.

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

Makino DL, Halbach F, Conti E. 2013. The RNA exosome and proteasome: common principles of degradation control. *Nat Rev Mol Cell Biol* **14:** 654–660.

Mühlemann O, Jensen TH. 2012. mRNP quality control goes regulatory. *Trends Genet* **28:** 70–77.

Nariai M, Tanaka T, Okada T, Shirai C, Horigome C, Mizuta K. 2005. Synergistic defect in 60S ribosomal subunit assembly caused by a mutation of Rrs1p, a ribosomal protein L11-binding protein, and 3′-extension of 5S rRNA in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **33:** 4553–4562.

Olson AJ, Brennecke J, Aravin AA, Hannon GJ, Sachidanandam R. 2008. Analysis of large-scale sequencing of small RNAs. *Pac Symp Biocomput:* 126–136.

Ozanick SG, Wang X, Costanzo M, Brost RL, Boone C, Anderson JT. 2009. Rex1p deficiency leads to accumulation of precursor initiator $tRNA^{Met}$ and polyadenylation of substrate RNAs in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **37:** 298–308.

Pillai RS, Bhattacharyya SN, Artus CG, Zoller T, Cougot N, Basyuk E, Bertrand E, Filipowicz W. 2005. Inhibition of translational initiation by Let-7 microRNA in human cells. *Science* **309:** 1573–1576.

Schümperli D, Pillai RS. 2004. The special Sm core structure of the U7 snRNP: far-reaching significance of a small nuclear ribonucleoprotein. *Cell Mol Life Sci* **61:** 2560–2570.

Simpson JC, Wellenreuther R, Poustka A, Pepperkok R, Wiemann S. 2000. Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep* **1:** 287–292.

van Hoof A, Lennertz P, Parker R. 2000. Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast. *EMBO J* **19:** 1357–1365.

Yang W, Lee JY, Nowotny M. 2006. Making and breaking nucleic acids: two-Mg^{2+}-ion catalysis and substrate specificity. *Mol Cell* **22:** 5–13.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31:** 3406–3415.

Zuo Y, Deutscher MP. 2001. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res* **29:** 1017–1026.

Characterization of the mammalian RNA exonuclease 5/NEF-sp as a testis-specific nuclear 3′ → 5′ exoribonuclease

Sara Silva, David Homolka and Ramesh S. Pillai

RNA 2017 23: 1385-1392 originally published online May 24, 2017
Access the most recent version at doi:10.1261/rna.060723.117