

國立臺灣大學生命科學院生命科學系碩士班

碩士論文

Department of Life Science

College of Life Science

National Taiwan University

Master Thesis



DEAD-box Protein 6的分子生物學探討

Molecular Interrogation of DEAD-box Protein 6

黃若晞

Jo-Hsi Huang

指導教授：朱家瑩 博士

Advisor: Chia-Ying Chu, Ph.D.

中華民國 105 年 7 月

July, 2016



國立臺灣大學碩士學位論文
口試委員會審定書



對 DEAD-box Protein 6 的分子生物學研究
A Molecular Interrogation of DEAD-box Protein 6

本論文係 黃若晞 (R02B21008) 在國立臺灣大學生命科學系、所完成之碩士學位論文，於民國 105 年 07 月 14 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

朱家榮 (指導教授)

張典顯

黃水蒼

生命科學系 系主任

閔明源



Acknowledgement



I am deeply indebted to a great many people who helped me through this research and the completion of this thesis. My deepest gratitude goes to my advisor and mentor Prof. Chia-Ying Chu. Her guidance and encouragement steered my learning as a young scientist. I am especially grateful for the opportunities and freedom under her advisorship. What privilege it is to have been able to enjoy the resources and to learn and explore nearly at one's will. If I have seen any further than before, that is because I am on her shoulder. My gratitude also goes to Prof. Wei-Chi Ku who contributed tremendously with his expertise in quantitative proteomics. I thank Prof. Kai-Yin Lo, Ting-Chun Lu, and Yi-Ting Yang who lent us materials, equipment, as well as their precious expertise and time in the polysome profiling experiments. I thank Prof. Yi-Sheng Cheng whose comment on the protein structure is truly illuminating. I thank Prof. Chao-Ti Ting and Prof. Wei-Pang Huang who kindly provided the materials used in this study. I also owe my upmost gratitude to the specialists in TechComm, including Dr. Mark Yih-Hsien Chiang, Dr. Riva Jui-Hua Chu, Samuel Chun-Hsing Lu, Kuan-Yi Li, Chang-Heng Chang, Chun-Yi Hsiao, Yu-Yun Kao, Jung-Chun Chiang, and Yi-Chun Chuang. The technical trainings and supports for Next-Generation sequencing and imaging that I received from them is invaluable. I also have to thank Dr. Tien-Hsien Chang and Dr. Yi-Shuian Huang, my thesis committee, for their helpful comments. Without the guidance and supports of these people, this study would not have reached its current form. I am also grateful for the extraordinary partnership with my lab mates. I owe my thanks to Yen-Chun Chen and Chih-Chun Yeh for their instructions on the experimental techniques in cell biology and biochemistry. I thank Hung-Yi Yeh and Chao-Yi Chuang for their instructions on molecular cloning. I would give special credits to Chao-Yi Chuang for her help in confocal microscopy, and Shih-Hsuan Lin for his



contribution on optimizing the cell cycle synchronization protocol. I thank Jin-Wen Huang, Chih-Kai Chang, Ben-Lyue Tseng, Yu-Wen Hsieh, Yi-Jou Huang for their advises during my early days in the lab. I thank Yen-Chun Chen, Hung-Yi Yeh, Chao-Yi Chuang, Yu-Chi Lin, Yen-Han Wang, Kuang-Tse Wang, Shih-Hsuan Lin, and Yi-Hsuan Chang for the joy and insights that I received from the discussions with them. Chao-Yi Chuang and Yen-Han Wang have been my greatest allies during last year. I benefited so much from their dedication to the lab. I also appreciate the consistent care and encouragement from Kuan-Hung Lin. I owe my greatest appreciation to all the fellow graduate students and the faculties in the Life Science Building. I could never imagine how much worse things would have been after the catastrophic Typhoon Soudelor without their collective dedication in reversing the damage done. I am especially thankful to Dr. Shih-Ting Cha for her kind arrangement that allowed me to be able to concentrate on my writing during the power outage following the Typhoon Soudelor. I am thankful to my great companion and teacher Nancy Nai-Hsin Yeh. Besides everything, she is credited for a large proportion of my computer skills, which made my work way more efficient. Last but foremost, I am all thankful to my Mom and Dad. I owe everything to their effort and strife. I could not have done the same without them.

摘要



對生物而言，受控制的基因表現是至關重要的。基因表現的控制部分有賴後轉錄(Post-transcriptional)調控者，包括 DEAD-box protein 6 (DDX6)。DDX6 最廣為所知的是它參與 RNA 代謝(RNA turnover)以及轉譯抑制(translational repression)。本研究中，我們關注 DDX6 在基因表現調控中所扮演的角色與其在細胞內的分佈、分佈的調控機制、及其生物意義。於第一章，我們闡述了 DDX6 於基因表現及其調控中的意義，並由此展開 DDX6 的基礎生化性質、在細胞內的功能、以及其對表型的(phenotypic)影響。第二章包括我們對於 DDX6 的細胞核質(nucleocytoplasmic)分佈的研究。深入地檢視現行的假說，並以分子生物以及蛋白結構的證據予以反駁。我們的結果更顯示 DDX6 可以藉由另外兩種並存的機制分布到細胞核與細胞質之中。第三章描述我們對於 DDX6 的基因表現調控的研究。我們利用整合性的功能性基因體學的方法，來探討 DDX6 對於核糖核酸 (RNA)及蛋白質表現的調控以及其規模與局限。我們首先指出 DDX6 在細胞內能結合信使核糖核酸(messenger RNA)以及長片段非編碼核糖核酸(long noncoding RNA)。不若一般的期待，我們發現在 HeLa 細胞株中 DDX6 並不顯著地影響整體的核糖核酸及蛋白質表現量。相反地，我們觀察到 DDX6 調控基因表現雜訊(gene expression noise)的證據。因此，我們的實驗結果提供對於 DDX6 的基因表現調控的新觀點。第四章描述了我們對於 DDX6 的入核機制更進一步的實驗結果。這些結果或能引導未來的研究。

關鍵字：

基因表現、後轉錄調控、基因靜默、DEAD-box protein、DDX6、核質分布、
入核信號、出核信號、信使核糖核酸、長片段非編碼核糖核酸、基因表現雜訊



Abstract

Well-tuned gene expression is essential for life and is in part secured by the post-transcriptional regulators. DEAD-box protein 6 (DDX6) is one such regulator. DDX6 is best known for its participation in RNA turnover and translational repression. In this study, we focus on the roles of DDX6 in regulating gene expression and on the subcellular localization of DDX6, and its regulation, as well as its functional implications. In the first chapter, we elaborated a logical sequence to locate the niche of DDX6 in the context of gene expression and regulation. On this basis, we further expounded the fundamental biochemical properties of DDX6, its functional roles in the cells, and their phenotypic consequences. The second chapter includes our investigation on the governing mechanism of the nucleocytoplasmic distribution of DDX6. We scrutinized the working hypothesis and refuted it with molecular and structural evidence. We further demonstrated that two parallel mechanisms direct the nucleocytoplasmic distribution of DDX6. The third chapter includes our investigation on the regulation of gene expression by DDX6. We adopted an integrative functional genomics approach to investigate the size and the extent of the DDX6-mediated regulation on RNA and protein expression. We first demonstrated that DDX6 binds both the messenger RNA and the long noncoding RNA. In contrast our original expectation, we did not observe significant regulation over the RNA or the protein expression levels by DDX6 in the HeLa cell line. Notwithstanding, we provided evidence that DDX6 regulates gene expression noise. Thus, our results provide a new perspective on the gene regulation by DDX6. The fourth chapter includes the further experimental results related to the nuclear entry mechanism of DDX6. The results described therein may guide further investigations in the future.

Key words:

Gene expression, post-transcriptional regulation, DEAD-box proteins, DDX6, nucleocytoplasmic distribution, nuclear localization signal, nuclear export signal, messenger RNA, long noncoding RNA, gene expression noise

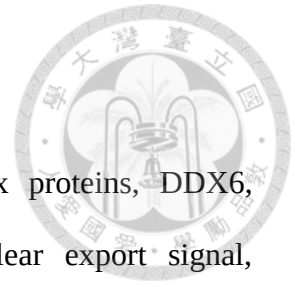


Table of contents



口試審定書	i
Acknowledgement	iii
摘要	v
Abstract	vii
I. Chapter 1	
1. Abstract	1
2. Gene expression and post-transcriptional regulation	1
2.1. Gene expression	1
2.2. Post-transcriptional regulation	3
3. DEAD-box proteins	6
3.1. Basic properties of the DEAD-box proteins	6
3.2. DEAD-box proteins as RNA chaperones and RNP remodelers	9
4. DEAD-box protein 6	12
4.1. A brief introduction to the DDX6 subfamily proteins	12
4.2. Identification and characterization of the DDX6 homologs	13
4.3. DDX6 in post-transcriptional gene silencing	20
4.4. The ATPase activity of DDX6	26
4.5. The RNA binding activity of DDX6	28
4.6. The cytoplasmic localization of DDX6	35
4.7. The nuclear localization of DDX6	36
4.8. Future perspectives	41



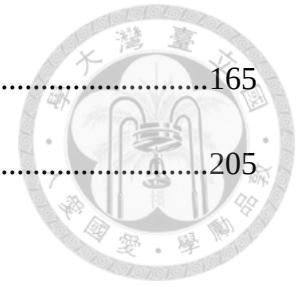
5. Tables.....	45
6. References.....	55

II. Chapter 2

1. Abstract.....	74
2. Introduction.....	74
3. Materials and methods	76
3.1. Cell culture and transfection.....	76
3.2. Drug treatments for CRM1 inhibition and cell cycle synchronization.....	77
3.3. Plasmid sources, molecular cloning, and fusion protein construction	77
3.4. Immunofluorescence (IF) and fluorescent in situ hybridization (FISH)	78
3.5. Microscopy, image processing, and analysis.....	79
3.6. Whole cell, nuclear, and cytoplasmic protein extraction.....	80
3.7. RNA co-immunoprecipitation (RIP)	81
3.8. Western blot (WB).....	82
3.9. RNA extraction, reverse transcription, and quantitative real-time PCR (qPCR)	82
3.10. Quantitative mass spectrometry (MS) and data analysis.....	83
3.11. Protein structures and NLS prediction.....	83
4. Results.....	84
4.1. DDX6 is present in the human cell nuclei.....	84
4.2. DDX6 binds the nuclear lncRNA MALAT1.....	85

4.3.	DDX6 is insensitive to the LMB treatment	85
4.4.	The putative NES of DDX6 is structurally inaccessible to CRM1	86
4.5.	The putative NLS, miRISC, and MALAT1 are dispensable for the nuclear entry of DDX6	88
4.6.	The CTD of DDX6 drives nuclear localization.....	89
4.7.	The CTD of DDX6 binds to the condensed chromosomes during mitosis	91
4.8.	The CTD of DDX6 facilitates nuclear entry by hitchhiking on the condensed chromosomes during mitosis	92
4.9.	4E-T can mediate piggyback shuttling of DDX6	92
5.	Discussion	94
5.1.	The widespread nuclear localization of DDX6	94
5.2.	On the nucleocytoplasmic shuttling model of DDX6	96
5.3.	The alternative mechanisms for the nucleocytoplasmic distribution of DDX6: A new synthesis	97
6.	Figures.....	100
7.	References	132
 III. Chapter 3		
1.	Abstract	138
2.	Introduction.....	138
3.	Materials and methods	140
3.1.	Cell culture and siRNA transfection.....	140
3.2.	RNA co-immunoprecipitation (RIP)	141

3.3.	RNA extraction, reverse transcription, and quantitative real-time PCR (qPCR)	141
3.4.	Protein extraction and Western blot (WB)	142
3.5.	Polysome fractionation and profiling	142
3.6.	RIP-seq and RNA-seq sample preparation and data acquisition	143
3.7.	RIP-seq and RNA-seq data analysis	144
3.8.	Quantitative MS sample preparation and data acquisition	145
3.9.	MS data analysis	146
4.	Results	146
4.1.	DDX6 binds both the mRNA and the lncRNA in human cell lines	146
4.2.	DDX6 binds the RNA localized in different subcellular compartments	147
4.3.	DDX6 has limited effect on the overall RNA expression level	149
4.4.	DDX6 regulates the expression noise of its target RNA	151
4.5.	DDX6 has limited effect on the overall protein expression level	152
4.6.	DDX6 regulates the protein expression noise of its bound mRNA	154
4.7.	DDX6 has limited effect on the polysome occupancy	155
4.8.	DDX6-regulated RNA and protein expression are weakly coupled	155
5.	Discussions	156
5.1.	A generalization of the concept of the DDX6's RNA binding activity	156
5.2.	On the role of DDX6 in regulating RNA and protein expression	158
5.3.	On the role of DDX6 in cell fate maintenance, specifications, and responses to stimuli: Under the "cell fate attractor" theory	162



6. Figures.....	165
7. References.....	205

IV. Chapter 4

1. Abstract.....	213
2. Introduction.....	213
3. Materials and methods	214
3.1. Molecular cloning.....	214
3.2. Cell culture and plasmid transfection	214
3.3. Sample preparation and microscopy	214
3.4. Image analysis	215
3.5. Protein sequence and structural analysis	216
4. Results.....	217
4.1. A semi-automatic workflow for the imaging-based analysis of the nucleocytoplasmic distribution.....	217
4.2. Time-dependent phenotypic shift in the cells expressing YFP-PKM-DDX6 CTD	218
4.3. The nucleocytoplasmic localization of the DDX6 homologs CTD.....	219
4.4. Mutational analysis of the DDX6 CTD motifs.....	220
5. Discussion	220
6. Figures.....	223
7. References.....	236

V. Appendices

1. List of quantitative PCR primer sets	238
2. List of small interfering RNA oligonucleotides.....	239
3. List of antibodies for Western blot	240
4. List of antibodies for immunofluorescence	241
5. List of antibodies for immunoprecipitation	242
6. List of expression vectors and constructs	243



Chapter 1

From Gene Expression
to the Multiple Faces DEAD-box Protein 6

A Literature Review





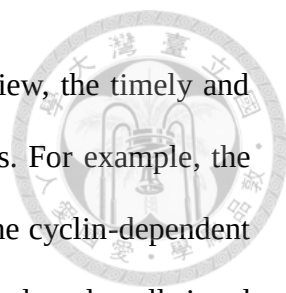
1. Abstract

Gene expression can be intervened by post-transcriptional regulation. The chief mediator of post-transcriptional regulation is the RNA-binding proteins, including the DEAD-box family proteins. In this study, we focus on one subfamily of the DEAD-box proteins, the DEAD-box proteins 6 (DDX6) subfamily. Members of the DDX6 subfamily are known as conserved and potent post-transcriptional regulators that function throughout a diverse set of different physiological/pathological conditions. In this chapter, we start from the extent and significance of gene expression and its regulation at post-transcriptional level. Then we turn to basic properties of the DEAD-box proteins and their functional relevance in post-transcriptional regulation. Subsequently, we take on and expand the subject of DDX6 with regards to its different facets, including its basic biochemical properties, molecular functions, physiological and pathological implications, subcellular localization and its functional relevance, and the promises of future research. We intend to locate the niche of DDX6 research in a broader context through presenting a logical sequence that connects the general knowledge of gene expression, post-transcriptional regulation, and the DEAD-box proteins to the current understanding of DDX6 in a hierarchical manner.

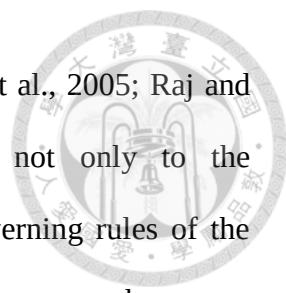
2. Gene expression and post-transcriptional regulation

2.1. Gene expression

Gene expression has far-reaching and multi-scaled significance in biology, ranging from the molecular dynamics to the cellular physiology and from the organismal



behaviors to social behaviors. From a molecular biology point of view, the timely and well-tuned gene expression underlies almost all biological processes. For example, the cell cycle is driven by the sequential expression of the cyclins and the cyclin-dependent kinases through time. At organismal level, the spatially organized and well-timed expression of the patterning genes is critical for pattern formation during the course of ontogenesis. Gene expression also governs individual behavior. This is widespread and manifest among the unicellular or the simple multicellular organisms, and also holds true for the more complex multicellular organisms, including us humans. For instance, our sleep/wake behavior is coupled to the circadian rhythm which is governed by the “circadian clock” (Lowrey and Takahashi, 2004). The circadian clock is in essence a complex molecular oscillator characterized by the persistent, oscillating, diurnal expression of the clock genes (Koike et al., 2012; Reppert and Weaver, 2001; Storch et al., 2002). Beyond the scope of the organismal biology, societies and ecosystems merge from the co-existence of the different individuals and the interactions among them. These interactions are the social behaviors. It is now known that the social behaviors can be influenced by differential patterns of gene expression, as have been demonstrated in the honey bees and other animals (Ament et al., 2008; Ament et al., 2012; Robinson et al., 2008; Robinson et al., 2005; Whitfield et al., 2003). Interestingly, the social behaviors can also alter gene expression reciprocally through epigenetic encoding (Robinson et al., 2008; Weaver et al., 2004). As such, gene expression connects different hierarchies of the biological systems through signaling and feedbacks, constituting a dynamic and complex network. From a systems biology perspective, the dynamics of the gene expression underlies the essential systems properties such as variability and stability, which in turn underlie the more macroscopic properties such as



population-wide heterogeneity or regularity (Huang, 2009; Huang et al., 2005; Raj and van Oudenaarden, 2008). Here, the term “dynamics” refers not only to the time-dependent evolution of the gene expression, but also the governing rules of the evolution of the systems; that is, the coupling, or the interactions, among the genes. These interactions, in the language of the molecular biologist, are the so-called “regulatory interactions” that altogether constitute a genome-wide gene regulatory network. It is now known that even the simple network motifs, containing 3 ~ 4 nodes, can create widely variable patterns of dynamics, including steady state, periodic oscillation, and even chaos (Zhang et al., 2012). Such network motifs have far-reaching implication in cell biology, developmental biology, and physiology. The circadian clock exemplifies the roles of the periodic oscillation in the physiology (Leloup et al., 2006; Scheper et al., 1999). Another striking example is the bi-repressive module which confers multi-stability on the fate of the progenitor cells (Huang et al., 2007). Together with the stochastic nature of gene expression, the bi-repressive module can poise and govern the cell fate bifurcation in the progenitor cells (Chang et al., 2008). The recurrent patterns of the bi-repressive modules and the other network motifs in the gene regulatory network thus indicate their pervasive participation in cell fate specification, maintenance of the cellular identity, and many other physiological processes.

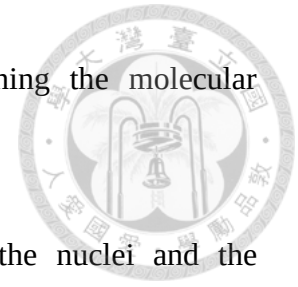
2.2. Post-transcriptional regulation

The process of gene expression can be simply described as the flow of the genetic information from DNA to RNA and then to proteins. In reality, however, owing to the stochastic nature of the molecular processes and the intervention of the intricate regulatory mechanisms, the genetic information in DNA does not always convey.

Likewise, the messages in the RNA do not always translate. Let alone the stochasticity, the eukaryotic cells possess a wide catalog of the regulatory mechanisms, such as the control of the chromatin accessibility, the transcriptional control, the post-transcriptional regulation of RNA, and the post-translational modification of proteins.

Here, we will focus on the post-transcriptional gene regulation of the eukaryotes. An overarching definition of post-transcriptional regulation is the gene regulation at RNA level. To be more specified, it is by convention to extensionally define post-transcriptional regulation as the collection of the regulatory pathways, including RNA splicing, mRNA surveillance, RNA interference (RNAi), translational regulation, and RNA turnover. Alternatively, post-transcriptional regulation can also be intentionally defined by simple criteria. A mechanism suffices for post-transcriptional regulation if it acts primarily on RNA and meets at least one of the following. First, it biochemically modifies the primary RNA sequences, such as alternative splicing and RNA editing. Second, it catalyzes the conformational changes of the RNA secondary and tertiary structures. Third, it modulates the RNA catalytic activities. Fourth, it controls the intermolecular interaction between RNA and other macromolecules. Fifth, it controls the subcellular localization of the RNA molecules, such as nuclear export. Sixth, it controls the RNA stability. Both definitions have their own merits. The extensional definition is more physiology-oriented. It illustrates a more holistic picture of the post-transcriptional regulation by tracing through the major pathways that intersect with the RNA life cycle. This can be especially helpful when dealing with problems on the cellular or the organismal levels. The intentional definition, on the other hand, is more mechanism-oriented. It describes the fundamental properties of the

post-transcriptional regulation. This can be helpful in determining the molecular functions of the RNA-binding proteins (RBPs).



Since the eukaryotic cells were compartmentalized in to the nuclei and the cytoplasm, the genetic materials and the transcription machinery are separated from the translation machinery. This compartmentalization detaches transcription from translation spatiotemporally, leaving substantial room for the intervention of the regulatory mechanisms at RNA level. For example, the mRNA surveillance mechanisms, including the nonsense-mediated decay, non-stop decay, and No-Go decay, destroy the aberrant messages to prevent their translation (Shoemaker and Green, 2012); the microRNA (miRNA) silences the protein expression noise (Schmiedel et al., 2015); and the mTOR complex 1, a biological sensor of the nutrient abundance, regulates translation through modulating several translational factors that directly act on the mRNA (Han et al., 2014). Collectively, these examples give a quick sketch of how the post-transcriptional regulatory mechanisms can integrate the intrinsic and the extrinsic as well as the genetic and the environmental information to tune gene expression.

Post-transcriptional regulation is primarily mediated by the RBPs. The binding of RBPs to the target RNA generally precedes the biochemical reactions, which may be catalyzed by the RBPs themselves or by their interacting partners. In the first scenario, the RBPs must contain catalytic domains, and may be regulated by other co-factors. In the latter scenario, the RBPs essentially serve as the platforms for the establishment of specific regulatory ribonucleoproteins (RNPs) by recruiting their partners as downstream effectors. It is noteworthy that the two scenarios were not mutually exclusive.

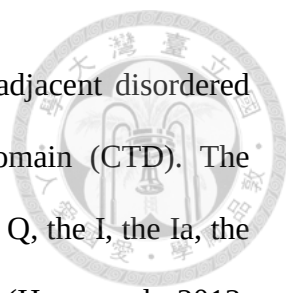
3. **DEAD-box proteins**

3.1. **Basic properties of the DEAD-box proteins**

In all the three taxonomical domains of life, there is a large, evolutionarily conserved family of RBPs known as the DEAD-box proteins (DBPs). The DBPs command a wide-range of functions that have emerged out of evolution which allow them to partake in multiple cellular processes. In the mammalian cells, the DBPs are involved in transcription, cellular/mitochondrial ribosome biogenesis, piwi-interacting RNA (piRNA) and miRNA biogenesis, spliceosome assembly, exon junction complex (EJC) assembly, nuclear export of mRNA, translational regulation, RNA storage, and RNA decay (Linder and Jankowsky, 2011). In these processes, the DBPs are known to serve as the molecular platforms for the assembly of the regulatory RNPs or as the RNA chaperones that modify the RNA structures or the RNP compositions (Linder and Jankowsky, 2011).

The basic architecture of the DBPs comprises of a structurally conserved helicase core flanked by the N-terminal and the C-terminal disordered regions. The helicase core consists of two contiguous globular domains interconnected by a flexible, disordered hinge region. The two globular domains of the DBPs, namely the DEXDc domain and the HELICc domain, both resemble the DNA recombination protein RecA topologically and are collectively referred to as the RecA-like domains. The two RecA-like domains are conventionally referred to as the N-terminal and the C-terminal RecA-like domains, or RecA-like domain 1 and 2, or more simply as RecA1 and RecA2 (Henn et al., 2012; Jarmoskaite and Russell, 2014; Linder and Jankowsky, 2011; Ozgur et al., 2015b). For

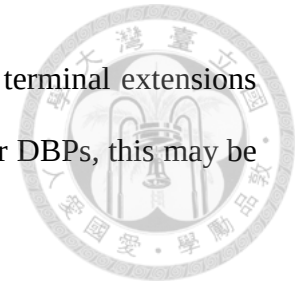




simplicity, we will refer to the two RecA-like domains and their adjacent disordered regions as the N-terminal domain (NTD) and the C-terminal domain (CTD). The helicase core also harbor 11 conserved functional motifs, namely the Q, the I, the Ia, the GG, the Ib, the II, the III, the IV, the QxxR, the V, and the VI motif (Henn et al., 2012; Ozgur et al., 2015b). These functional motifs are essential to the two fundamental activities common to the DBPs, namely the RNA binding activity and the ATPase activity. As such, these motifs are collectively referred to as the helicase motifs (Henn et al., 2012; Jarmoskaite and Russell, 2014; Linder and Jankowsky, 2011; Ozgur et al., 2015b).

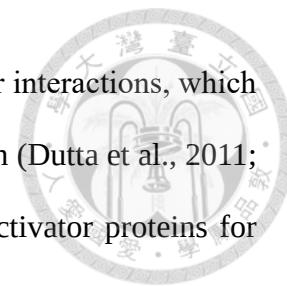
One crucial aspect of the RNA binding activity is the specificity towards RNA substrates. In general, the DBPs display similar affinity towards different RNA species that differ in sequence, indicating a lack of the sequence- and the structure-specificity per se. This lack of specificity is attributed to the structural features of the DBPs. More specifically, the DBPs bind on the sugar-phosphate backbone of their RNA substrates. The interaction between the DBPs and their RNA substrates is mediated primarily by hydrogen bonds and salt bridges between the helicase core and the RNA backbone instead of the nitrogenous bases (Henn et al., 2012; Linder and Jankowsky, 2011; Ozgur et al., 2015b). This lack of inherent specificity confers on the DBPs the robust tolerance for the widely variable RNA sequences or structures. For example, the deposition of eIF4AIII and the EJC would require robust tolerance for the variable sequences near the splice-junctions (Bono et al., 2006; Sauliere et al., 2012). However, a considerable fraction of the DBPs are highly dedicated to certain tasks, and thus require some sequence- or structure-specific tethers to direct them to their physiological targets. In

some DBPs, this may be achieved by the auxiliary domains or the terminal extensions that confer RNA binding specificity (Chang and Lee, 2012). In other DBPs, this may be achieved by their partner proteins (Young et al., 2013).



The other fundamental activity of the DBPs is the ATPase activity. As ATPases, the DBPs can convert the chemical energy stored in the ATP into the mechanical forces that drive both the conformational changes of the DBPs and their RNA/RNP substrates in each ATPase cycle (Henn et al., 2012; Jarmoskaite and Russell, 2014; Linder and Jankowsky, 2011; Ozgur et al., 2015b). The energetics of the ATPase cycle is coupled to the conformational changes of the DBPs (Henn et al., 2012; Linder and Jankowsky, 2011; Ozgur et al., 2015b). The conformations of the DBPs are determined by the relative positioning of the two RecA-like domains. Generally, the DBPs adopt similar ATPase-active conformations (also known as the “closed” conformation, or the “on” state), with the two RecA-like domains enclose on each other and stay roughly perpendicular to each other (Linder and Jankowsky, 2011; Ozgur et al., 2015b). By perpendicular, we refer to the directions pointed by the β -sheets in the two RecA-like domains (Ozgur et al., 2015b). The ATPase-inactive conformations (also known as the “open” conformation; or the “off” state) are more variable (Ozgur et al., 2015b). The conformational switch from the “off” state to the “on” state is characterized by the domain closure triggered by the concomitant binding of RNA and ATP. The domain closure brings the conserved functional motifs together to form the functional catalytic site. However, not all DBPs can be readily activated by the concomitant binding of RNA and ATP. For example, the helicase core of yeast eIF4A is too flexible to approach the “on” state conformation efficiently (Schutz et al., 2008). DDX6, on the other hand,

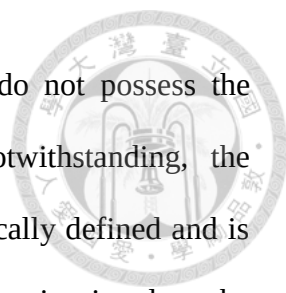
is locked in a rigid “off” state conformation by strong intramolecular interactions, which prevent the conformational switch required for the ATPase activation (Dutta et al., 2011; Mathys et al., 2014). These examples signify the demand of the activator proteins for some DBPs during the ATPase activation.



3.2. DEAD-box proteins as RNA chaperones and RNP remodelers

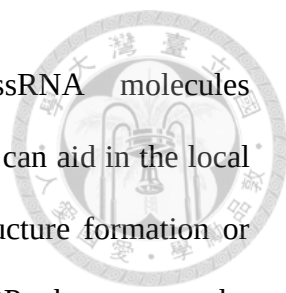
The DBPs are known to function as RNA chaperones and RNP remodelers. As chaperones, the DBPs catalyze the rearrangement of the RNA structures, which by and large dictate the RNA's functions. Thus, these functionalities have the biological significance of helping the RNA to achieve functional folding or to convert from the mis-folded conformations. As RNP remodelers, the DBPs can serve as platforms for RNP nucleation or can actively remodel the composition of the existing RNP (Henn et al., 2012; Jarmoskaite and Russell, 2014; Linder and Jankowsky, 2011).

The most studied RNA chaperoning function of the DBPs is RNA unwinding; that is, separating the two strands of an RNA duplex. Unlike the other helicases, the DBPs utilize a unique mechanism for RNA unwinding which is distinct from the conventional “translocation-displacement” mechanism. In fact, the DBPs are non-processive, meaning that they do not translocate along the RNA molecules (Henn et al., 2012; Jarmoskaite and Russell, 2014; Linder and Jankowsky, 2011). The DBPs are also non-directional, meaning that they do not distinguish between 3'-to-5' and 5'-to-3' directionality (Jarmoskaite and Russell, 2014; Ozgur et al., 2015b). Furthermore, the DBPs unwind RNA duplex internally with no dependence on initial 3'- or 5'-overhang (Henn et al., 2012; Jarmoskaite and Russell, 2014; Linder and Jankowsky, 2011; Ozgur



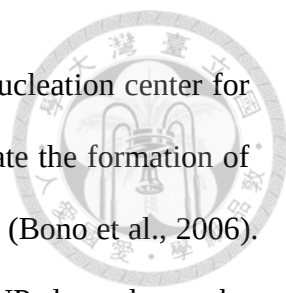
et al., 2015b). For the reasons above, some consider the DBPs do not possess the genuine helicase activity (Jarmoskaite and Russell, 2014). Notwithstanding, the mechanism by which the DBPs unwind RNA duplexes is biochemically defined and is referred to as local duplex separation. In essence, local duplex separation involves the RNA clamping and the subsequent RNA bending (Henn et al., 2012; Jarmoskaite and Russell, 2014; Linder and Jankowsky, 2011; Ozgur et al., 2015b). Before switching to the “on” state conformation, the NTD and CTD of the DBPs can bind ATP and RNA separately (Mallam et al., 2012). Upon changing to the “on” state, the helicase core of the DBP encloses and clamps on the sugar-phosphate backbone of the RNA substrate. In the case of single-stranded RNA (ssRNA) binding, the clamping simply bends the RNA substrates, whereas in the case of double-stranded RNA (dsRNA) binding, the clamping results in the bending of one strand (the one that is directly bound) and causes the duplex to separate locally (Henn et al., 2012; Jarmoskaite and Russell, 2014; Linder and Jankowsky, 2011; Mallam et al., 2012; Ozgur et al., 2015b). This indicates the DBPs can exert regulatory functions on the local secondary RNA structures through resolving the RNA helices. Notably, the local duplex separation is a direct consequence of the domain closure and precedes the ATP hydrolysis. That is, ATP hydrolysis is not required for the direct displacement of the targeted RNA strand (Henn et al., 2012; Liu et al., 2008), but for the release of the DBPs from their RNA substrates. Thus, ATPase is required for the completion of an ATPase cycle and for the consecutive ATPase cycles which may ultimately lead to full unwinding of the RNA substrate.

Besides RNA unwinding, research has also revealed additional RNA chaperoning functions for the DBPs. The first is the RNA annealing activity. This activity facilitates



the RNA duplex formation from the complementary ssRNA molecules (Uhlmann-Schiffler et al., 2006; Zhao and Jain, 2011). This activity can aid in the local structural rearrangement of RNA by facilitating the secondary structure formation or even fostering the tertiary structure formation. Interestingly, the DBPs that possess the RNA annealing activity also retain the RNA unwinding activity, albeit the two activities were observed under the different conditions in vitro. In addition, the RNA annealing activity is ATP-independent. These observations suggested the annealing and the unwinding activities are not mutually exclusive, and may be regulated in a switch-like manner. Thus the RNA chaperoning activities of the DBPs may be modulated by the biological contexts and serve for different purposes. The second is the RNA strand displacement/exchange activity. This activity promotes the exchanging of the individual strand of a given RNA duplex with the other complementary RNA molecules in an ATP-dependent manner (Stampfl et al., 2013; Zhao and Jain, 2011). This activity can facilitate the fast interchanging between different RNA conformations. In some research literature, this activity was simply referred to as the RNA strand displacement activity, but it should not be confused with the “translocation-displacement” unwinding mechanism of the conventional helicases. Notably, both the RNA annealing and the strand displacement/exchange activities are not sequence-specific. Third, it was recently discovered that two DBPs, CYT-19 and Ded1p, can disrupt tertiary RNA structure by capturing and resolving the transiently exposed RNA helices. This further prevents the restoration of the original tertiary structures (Pan et al., 2014). This activity is named “tertiary RNA structure disruption by helix capture”.

Besides RNA chaperoning, the DBPs can also exert regulatory effects by

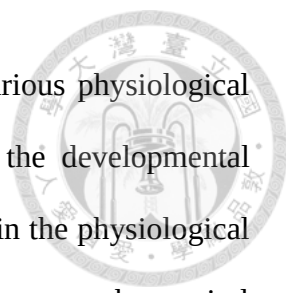


modifying the RNP composition. First, the DBPs can serve as the nucleation center for other RNP components. For example, DDX48 (eIF4AIII) can nucleate the formation of the EJC on the spliced RNA by recruiting the other EJC components (Bono et al., 2006). Notably, the exact mechanism by which a DBP nucleates the RNP depends on the identity of the given DBP and its interacting partners. Second, the DBPs can also remodel the existing RNP composition by displacing the preoccupying proteins. For example, yeast Dbp5p displaces Nab2p to facilitate mRNA export (Tran et al., 2007). Another example is that yeast Ded1p displaces U1 snRNP from RNA substrate (Bowers et al., 2006). Interestingly, both Dbp5p and Ded1p exhibited some degree of selectivity towards the displaced targets, indicating the DBPs may be programmed to modify specific RNPs.

4. DEAD-box protein 6

4.1. A brief introduction to the DDX6 subfamily proteins

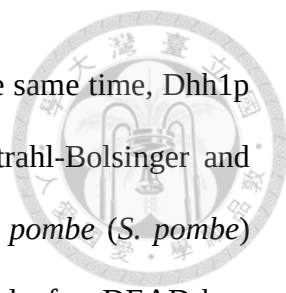
Members of the DDX6 subfamily are extremely well-conserved across the eukaryotes. The evolutionary conservation of the DDX6 subfamily is best exemplified by the amazing 68% amino acid identity in the helicase core of the human DDX6 and the *S. cerevisiae* Dhh1p. This high degree of conservation implies the conservation of the higher-order structural features. It also resonates with the essential roles of the DDX6 homologs in the fundamental physiological processes. DDX6 and its partner proteins have been given various names in different model organisms (Table. 1). For general reference, we adhere to the common nomenclature in human. But we occasionally use the organism-specific nomenclature to reflect the historic contexts.



The DDX6 homologs enjoy wide functional implication in various physiological conditions. For instance, the DDX6 homologs are implicated in the developmental processes such as gametogenesis, neurogenesis, and organogenesis, in the physiological conditions such as stress responses, and in the disease development such as viral infections, carcinogenesis and cancer progression. These seemingly diverse processes ultimately converged on one unifying theme: the roles of the DDX6 subfamily in regulating gene expression. The past research has crystalized into the understanding that DDX6 is a potent post-transcriptional regulator whose principal duty is to silence protein production. On this foundation, the recent research has furthered the roles for the DDX6 subfamily members in gene regulation, including the regulatory role for the long noncoding RNA (lncRNA) and the potential role in modulating transcription.

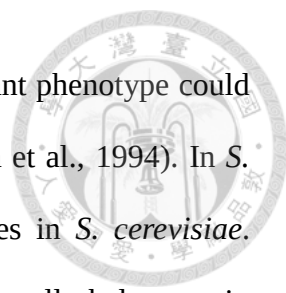
4.2. Identification and characterization of the DDX6 homologs

Since the 1990s, the DDX6 homologs have been identified and characterized in various biological models. The first DDX6 homolog was identified in the model fruit fly *Drosophila melanogaster* (hereafter referred to as *Drosophila*) and was named Me31B after its maternal (ovarian germline) expression and its chromosomal locus at region 31B (de Valoir et al., 1991). The human DDX6 was the second to be identified. It was found on the chromosomal breakpoint 11q23.3 in the human RC-K8 lymphoma cell line (Lu and Yunis, 1992). The DDX6 gene was later named as the RCK gene after the RC-K8 cell line (Akao et al., 1992). Following the identification of human DDX6, the Ddx6 in *mus musculus* (the house mouse, hereafter referred to as mouse) was also identified (Akao et al., 1995). The human and the mouse homologs were reported to share $\approx 98\%$ of amino acid identity and both produced ≈ 54 kDa protein products (Akao



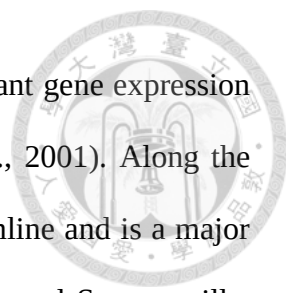
et al., 1995). Hence, DDX6 is often alias as RCK/p54. Around the same time, Dhh1p of the budding yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) (Strahl-Bolsinger and Tanner, 1993) and Ste13 of the fission yeast *Schizosaccharomyces pombe* (*S. pombe*) (Maekawa et al., 1994) were identified. The name Dhh1p stands for DEAD-box helicase homolog 1, and the name Ste13 came from the sterile (incapable of mating) phenotype of the Ste13-null mutants. Shortly after, Xp54, the DDX6 homolog in the African claw-frog *Xenopus laevis* (hereafter abbreviated as *Xenopus*), was identified in the female germline cells as a major component of the maternal mRNP (Ladomery et al., 1997). CGH-1, or conserved germline helicase 1, was identified as the DDX6 homolog in *Caenorhabditis elegans* (*C. elegans*) (Navarro et al., 2001). Trailing these earlier identifications, DjCBC-1, or chromatoid body component 1, was identified in the pluripotent stem cells in the planarian *Dugesia japonica* (*D. japonica*) (Yoshida-Kashikawa et al., 2007). Recently, our group identified the DDX6 homolog Zp54 in *Danio rerio* (zebrafish) and characterized its expression pattern during the early embryonic development (C. C. Yeh's master thesis). These and the ensuing studies have accumulated into a large body of knowledge from which several physiological themes have emerged.

First, the DDX6 homologs take part in the stress responses. In *S. pombe*, nitrogen starvation triggers the cell cycle arrest and the subsequent sexual development, which includes mating, meiosis, and sporulation. In the earliest study, the *S. pombe* Ste13 was simply identified as a positive regulator of the cell cycle arrest during nitrogen starvation and therefore an essential gene for the sexual development in response to such stress (Kitamura et al., 1990). It was not until later time that Ste13 was



characterized as a member of the DDX6 subfamily whose null-mutant phenotype could be suppressed by the heterologous expression of Me31B (Maekawa et al., 1994). In *S. cerevisiae*, Dhh1p takes part in the responses to metabolic stresses in *S. cerevisiae*. When deprived of glucose or amino acids, the WT *S. cerevisiae* normally halts protein synthesis. But the Δ Dhh1/ Δ Pat1 *S. cerevisiae* mutant was unable to attenuate translation under the same metabolic stresses (Coller and Parker, 2005). Dhh1p is also required for the resistance to heat (restrictive growing temperature) and DNA damaging agents, including hydroxyurea, methyl methane sulfonate, ultraviolet (UV), and γ -irradiation (Bergkessel and Reese, 2004; Carroll et al., 2011; Cheng et al., 2005; Dutta et al., 2011; Westmoreland et al., 2003). Similarly, CGH-1 moderates the DNA damage-induced apoptosis in *C. elegans* (Tomazella et al., 2012). Recently, our group has discovered that DjCBC-1 confers on *D. japonica* the resistance to γ -irradiation (Y. H. Wang's master thesis). Together, these results point to the conserved, positively regulatory roles of the DDX6 homologs in stress responses. However, a recent report displayed that the reduction in DDX6 expression level increases the resistance to chemotherapy and radiotherapy in the patient-derived glioblastoma cells (Cho et al., 2016). This divergence perplexed the exact roles of DDX6 in stress responses. One initial question to be answered is whether such phenomenon is restricted to glioblastoma or prevails among cancers. A secondary, yet broader in scope, question is whether such divergence represents the difference in normal physiological states and the cancerous states.

Second, the DDX6 homologs are involved in gametogenesis. In *Drosophila*, Me31b is highly expressed in the female gonad (de Valoir et al., 1991). Me31B prevents the premature translation of the maternal messages in the ovaries (Nakamura et al.,



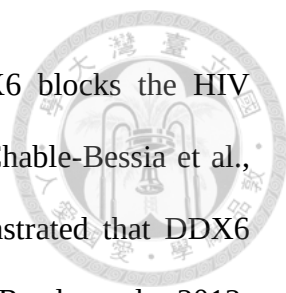
2001). The conditional knockout (KO) of Me31B leads to the aberrant gene expression in the ovaries and eventually abolishes oogenesis (Nakamura et al., 2001). Along the same line, Xp54 is also predominantly expressed in the female germline and is a major component of the maternal mRNP in the *Xenopus* oocytes (Smillie and Sommerville, 2002). In addition, Xp54 is capable of translationally silencing the reporter mRNA in the oocytes (Minshall et al., 2001). These findings suggest a role in controlling the maternal messages in the vertebrates. In *C. elegans*, CGH-1 is predominantly expressed in both the oogenic and the spermatogenic cells and is essential for fertility (Navarro et al., 2001). More specifically, CGH-1 loss-of-function leads to the production of crippled sperms and triggers apoptosis in the female germline (Navarro et al., 2001). Furthermore, CGH-1 is capable of binding, stabilizing, and controlling the localization of maternal messages during the oogenesis (Boag et al., 2008). In the mammals, the DDX6 homologs are expressed in both somatic and germline cells (Akao and Nakagawa, 2006; Kawahara et al., 2014; Matsumoto et al., 2005). In mouse, Ddx6 is expressed the spermatocytes and the oocytes (Matsumoto et al., 2005). More recently, an anatomical study revealed the conserved and continuous expression of the DDX6 homologs in spermatogenic cells during course of the spermatogenesis (Kawahara et al., 2014). This observation resonates with the results obtained in the other model organisms, suggesting a conserved role in gametogenesis for the DDX6 subfamily members. Interestingly, Ste13 of *S. pombe* is required for the commitment into meiosis (Maekawa et al., 1994), implicating broader functional conservation in sexual development in the eukaryotes. The commonalities in the roles of the DDX6 homologs in both the yeast meiosis and the metazoan gametogenesis thus represent a potential direction for future studies.



Third, the DDX6 homologs are required for the normal development in the metazoans. In *Drosophila*, Me31B KO leads to larval lethality (Nakamura et al., 2001). In early *C. elegans* embryos, partial CGH-1 silencing disrupts cell division (Audhya et al., 2005). In mouse embryos, disrupting the temporal pattern of Ddx6 expression leads to developmental delay in early embryonic stage and is ultimately lethal (Matsumoto et al., 2005). A recent study, utilizing the in vitro human epithelial organoid model, reported that DDX6 is required for maintaining the stemness of the epithelial progenitor cells and is therefore also required for the normal epithelium formation (Wang et al., 2015). Another recent study, utilizing the in vitro neural differentiation model, reported that the mouse Ddx6 is upregulated during the differentiation of the cultured neural stem cells, and is both sufficient and necessary to induce neural differentiation (Nicklas et al., 2015). Notably however, DDX6 does not force differentiation in all cell types with neuronal potency. For example, DDX6 overexpression (OE) in the cultured SH-SY5Y neuroblastoma cell line does not induce differentiation (data not shown, J. H. Huang's data). Notwithstanding, several in vivo studies of the neural development still support the roles of the DDX6 homologs in neurogenesis. In *Drosophila*, Me31B is required for the normal dendrite arborization, including both tiling and branching (Barbee et al., 2006). In *D. japonica*, DjCBC-1 knockdown (KD) disrupts the regeneration of the nervous system (Rouhana et al., 2010). Our group further displayed that the long-term DjCBC-1 KD also causes the mis-patterning of the nervous system, presumably through perturbing the fast tissue homeostasis of the planarians. Furthermore, in the newly regenerated planarian, DjCBC-1 KD causes defects ranging from the mis-patterning of the nervous system to lethality. Furthermore, higher lethality correlates with the smaller body size and the earlier RNAi treatment following

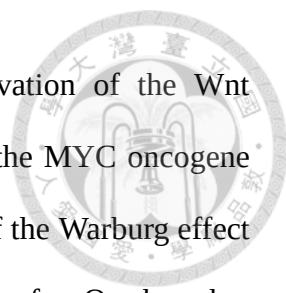
amputation. Both conditions represent smaller neoblast/progenitor reservoir and thus the lower regenerative potential (Y. H. Wang's master thesis). Together, these results point to the fundamental roles of the DDX6 homologs in the cell fate decision of the stem cell/progenitor cell populations and in the processes of organogenesis, neurogenesis, and ontogenesis.

Fourth, the DDX6 homologs are involved in the viral and the retrotransposon replication. The roles of the DDX6 homologs in viral replication are well exemplified by the positive-strand ssRNA (+ssRNA, or the group IV) viruses. One well-studied case is the hepatitis C virus (HCV), with several studies agreeing on DDX6's role in promoting HCV replication (Chatel-Chaix et al., 2013; Huys et al., 2013; Jangra et al., 2010; Scheller et al., 2009). An early clinical report had also remarked the positive correlations between DDX6 OE, HCV infections, and liver cancer development (Miyaji et al., 2003). In addition, DDX6 is required for the replication of the dengue virus (DENV) (Ward et al., 2011), and is also implicated the replication of the West Nile virus (WNV) (Chahar et al., 2013). HCV, DENV, and WNV all belong to the family *Flaviviridae*. The DENV and the WNV both belong to the genus *Flavivirus*. Noteworthily, the Zika virus (ZIKV) also belongs to the genus *Flavivirus*. Since these viruses are the major pathogens to humans, understanding how DDX6 promotes their replication can aid the prevention and the medication. In *S. cerevisiae*, Dhh1p is required for replication of the Brome mosaic virus (BMV) (Alves-Rodrigues et al., 2007) and the Flock House virus (FHV) (Gimenez-Barcons et al., 2013), both of which also are +ssRNA viruses. DDX6 is also involved in the replication of the retroviruses (ssRNA-RT, or the group VI), and the most studied example is the human



immunodeficiency virus (HIV). Two studies displayed that DDX6 blocks the HIV replication as an effector of the microRNA (miRNA) machinery (Chable-Bessia et al., 2009; Nathans et al., 2009). However, two counter-studies demonstrated that DDX6 promotes HIV replication through facilitating capsid assembly (Reed et al., 2012; Robinson et al., 2014). The studies in the other retroviruses also offered conflicting results. A study in the prototype foamy virus (PFV) displayed that DDX6 is required for the PFV encapsidation (Yu et al., 2011). Yet a counter-study showcased that Me31B represses the Rift Valley fever virus (RVFV) replication (Hopkins et al., 2013). Moreover, Me31B is involved in the silencing of the retrotransposons in the *Drosophila* germline (Liu et al., 2011). Yet again, two studies in *S. cerevisiae* demonstrated that Dhh1p is required for the assembly of the virus-like particles for the Ty1 and Ty3 retrotransposons (Bilanchone et al., 2015; Checkley et al., 2010). Together, these discrepancies demand further investigations into the exact roles of the DDX6 in the retrovirus and the retrotransposon replication.

Fifth, DDX6 is involved in the carcinogenesis and cancer progression. As mentioned earlier, DDX6 facilitates HCV replication. And DDX6 OE correlates with both hepatitis C and liver cancer development. Therefore, DDX6 likely partakes in the malignant transformation in the hepatocytes through fueling the HCV infection. DDX6 OE was also reported in colorectal tumors (Hashimoto et al., 2001; Nakagawa et al., 1999; Taniguchi et al., 2015). Along this line, DDX6 KD decreases the proliferation of the colorectal cancer LoVo cell line both in vitro and in xenografts (Lin et al., 2008). These results indicate DDX6 may have a direct role in the development of the colorectal cancer. Several mechanisms have been proposed for the involvement of DDX6 in



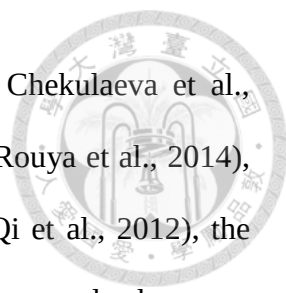
colorectal cancer development, including the transcriptional activation of the Wnt pathway (Lin et al., 2008), the post-transcriptional stabilization of the MYC oncogene (Hashimoto et al., 2001; Matsui et al., 2006), and the maintenance of the Warburg effect (Taniguchi et al., 2015), yet no clear consensus has been reached so far. On the other hand, a study demonstrated that DDX6 is required for the proliferation and the migration of the glioblastoma U87MG cell line (Baskin et al., 2015). Another study, however, displayed that the downregulation of DDX6 promotes the establishment of the resistance to chemotherapy and radiotherapy in the patient-derived glioblastoma cells (Cho et al., 2016). In breast cancer, the genomic locus of DDX6 gene is commonly amplified in the nodal metastases as compared to the primary tumors (Desouki et al., 2011). Recently, our group also found that the expression level of DDX6 affects the migration capability and the invasiveness in the paired lung cancer cell lines, CL1-0 and CL1-5 (Y. C. Chen's master thesis). Collectively, these findings evidence that DDX6 is highly relevant in the cancer biology. But since the cancer development and the progression are highly dynamic, and the cancer cells are heterogeneous in nature, further investigation is required to fully resolve the DDX6's role in cancer.

4.3. DDX6 in post-transcriptional gene silencing

Commanding changes in gene expression is one fundamental commonality among the seemingly diverse biological processes in which the DDX6 homologs are involved. Generally the physiological processes are secured by the normal activities of the DDX6 homologs, whereas the pathological processes are fueled by the deregulation of the DX6 homologs. This resonates with the roles of the DDX6 homologs in post-transcriptional regulation. In the following section, we summarize three lines of evidence that marks

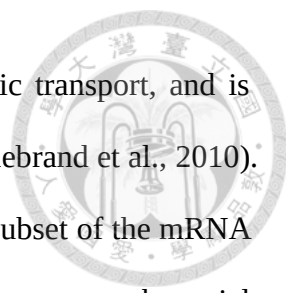
the roles of the DDX6 homologs in gene silencing at the RNA level.

The first line of evidence marks to the role of the DDX6 homologs in RNA decay. In the early studies in *S. cerevisiae*, Dhh1p was reported to interact with Caf1p (also known as Pop2p), Ccr4p, and Not1p (Coller et al., 2001; Hata et al., 1998; Maillet and Collart, 2002), the three core components of the CCR4-NOT complex, both of which were later known to function in deadenylation (Chen et al., 2002; Daugeron et al., 2001; Tucker et al., 2002; Tucker et al., 2001) in addition to their first-known functions in transcription (Collart and Struhl, 1994; Liu et al., 1998; Maillet et al., 2000; Oberholzer and Collart, 1999). In the following, Dhh1p was also reported to interact with the 5'-3' RNA decay machineries, including Xrn1p, Dcp1p, Pat1p, and Edc3p (Carroll et al., 2011; Fischer and Weis, 2002; Nissan et al., 2010). Consistently, Dhh1p was found to be required for the efficient mRNA decapping (Coller et al., 2001; Fischer and Weis, 2002), which follows deadenylation and precedes the 5'-3' exonuclease degradation by Xrn1p (Beelman and Parker, 1995; Decker and Parker, 1993). Since both the deadenylation and the decapping are general steps for RNA decay, Dhh1p is considered as a positive regulator of the general RNA turnover. In the following, the findings in *S. cerevisiae* were generalized by the similar findings in *Drosophila* and human (Fenger-Gron et al., 2005; Haas et al., 2010; Ozgur et al., 2010). This generalization demonstrates the interactome of the DDX6 homologs and their roles in RNA decay are evolutionarily conserved (Jonas and Izaurralde, 2013). Importantly, the direct interaction between the human DDX6 and CNOT1, the scaffolding component of the CCR4-NOT complex, was recently revealed by the crystallographic analyses (Chen et al., 2014; Mathys et al., 2014; Rouya et al., 2014). This finding established the structural basis of how the

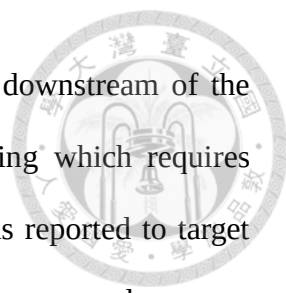


general deadenylation, the miRNA-mediated (Braun et al., 2011; Chekulaeva et al., 2011; Chen et al., 2014; Chu and Rana, 2006; Mathys et al., 2014; Rouya et al., 2014), the TTP-mediated (Fabian et al., 2013; Fenger-Gron et al., 2005; Qi et al., 2012), the NANOS-mediated (Bhandari et al., 2014; Raisch et al., 2016), and several other gene silencing pathways all converge on the CCR4-NOT/DDX6 axis. Consistently, Dhh1p can bypass deadenylation and mediate the decapping of the reporter mRNA to which it is artificially tethered (Carroll et al., 2011; Sweet et al., 2012). Notably however, it was demonstrated that Dhh1p, unlike Pat1p and Edc3p, is incapable of directly stimulating the activity of the purified Dcp1p/Dcp2p complex, the core decapping machine, in vitro (Nissan et al., 2010). As such, Dhh1p most probably activates decapping indirectly. One likely hypothesis is that DDX6 facilitates the deposition of PATL1 and EDC3 onto the targeted RNA to foster decapping. In addition, the recent research has also implicated the involvement of Dhh1p in the other RNA metabolic processes in *S. cerevisiae*. First, Dhh1p and Pat1p likely participate in the bulk degradation of the cellular mRNA in the face of glucose deprivation (Arribere et al., 2011; Collier and Parker, 2005; Sweet et al., 2012). Second, the control of mRNA stability by codon optimality requires Dhh1p and the other decay factors (Presnyak et al., 2015).

The second line of evidence marks the role of the DDX6 homologs in translational repression. In the *Xenopus* oocyte, where the decapping machinery is inactive (Zhang et al., 1999), Xp54 binds the stored mRNA without causing degradation (Ladomery et al., 1997; Minshall et al., 2001). Similarly, Me31B silence the translation of the maternal mRNA in the developing *Drosophila* ovaries without causing degradation (Nakamura et al., 2001). Also, Me31B participates in the neuronal staufen granules (Barbee et al.,

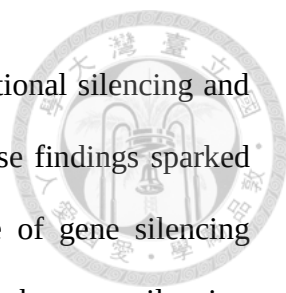


2006), a class of the repressive RNA granules involved in dendritic transport, and is implicated in the translational repression of the *CaMKII* mRNA (Hillebrand et al., 2010). In *C. elegans*, CGH-1 binds, stabilizes, and presumably represses a subset of the mRNA during oogenesis (Boag et al., 2008). In *S. cerevisiae*, Dhh1p OE causes substantial polysome reduction and growth inhibition presumably independent of decapping (Coller and Parker, 2005). More importantly, Dhh1p, Xp54, and the human DDX6 all block protein synthesis without causing mRNA degradation in vitro (Coller and Parker, 2005; Tanaka et al., 2006). Interestingly, the ability of Xp54 to repress translation is enhanced by xRAPA (Tanaka et al., 2006), a direct binding partner of Xp54. Consistently, CGH-1 and CAR-1 together repress the *glp-1* mRNA translation in *C. elegans* (Noble et al., 2008). In addition, 4E-T also interacts directly with DDX6 and functions downstream to the CCR4-NOT/DDX6 axis to repress translation (Nishimura et al., 2015; Ozgur et al., 2015a; Waghray et al., 2015). These findings established the role of the DDX6 homologs in translational repression. The fundamental question is how DDX6 repress translation. Since Xp54 cannot repress the translation mediated by several types of the viral internal ribosomal binding site (IRES) in vivo (Minshall et al., 2009; Waghray et al., 2015), it is unlikely that Xp54 represses translation at elongation. Furthermore, while Xp54 is unable to repress the translation mediated by the classical swine fever virus (CSFV) IRES (Minshall et al., 2009; Waghray et al., 2015), it is capable of repressing the translation mediated by the polio virus (PV) IRES (Waghray et al., 2015). As compared to the CSFV IRES, the PV IRES requires eIF4A and eIF4G in addition to other common factors (Jackson et al., 2010; Walsh and Mohr, 2011). Consistently, 4E-T was found to repress translational through an eIF4E-independent mechanism downstream of the CCR4-NOT complex (Kamenska et al., 2014; Sweet et al., 2012;



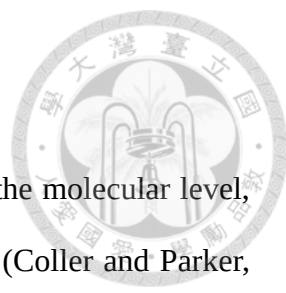
Waghray et al., 2015). In conjunction, the translational repression downstream of the CCR4-NOT complex is independent of the 43S ribosomal scanning which requires eIF4A (Kuzuoglu-Ozturk et al., 2016). Most importantly, Scd6p was reported to target eIF4G with their RGG motifs and formed complex with it on the targeted reporter mRNA (Rajyaguru et al., 2012). Together, these findings suggest DDX6 targets eIF4G to repress translational initiation. Interestingly, it was recently concluded that PATL1, EDC3, 4E-T, LSM14A, and their homologs utilize the same binding surface on the DDX6 CTD (Kuzuoglu-Ozturk et al., 2016; Ozgur et al., 2015a; Ozgur and Stoecklin, 2013; Sharif et al., 2013; Tritschler et al., 2009; Tritschler et al., 2008), indicating that DDX6 can elicit decapping and translational repression through parallel mechanisms.

The third line of evidence marks the role of the DDX6 homologs in the processing body (PB) assembly, which in turn highlights their roles in gene silencing. The PBs were first discovered in the cytoplasm of *S. cerevisiae* as the granular foci enriched with the mRNA decay intermediates and the RNA decay factors (Sheth and Parker, 2003). This discovery can be partially attributed to the idea the decay factors could occupy certain subcellular loci, which in turn stemmed from the dual roles of the DDX6 homologs in decapping and in aggregating into the mRNP granules that are visible under microscopes (Sheth and Parker, 2003). Shortly after, the homologous structures of the *S. cerevisiae* PB were also discovered in the human cells (Cougot et al., 2004; Eystathioy et al., 2003) and the *Drosophila* Schneider 2 (S2) cell line (Eulalio et al., 2007a). In the following, it was found that the PB sizes and numbers vary with the cellular translational status, the stress stimuli, decay factors, and the RNA content within the PBs (Brenques et al., 2005; Teixeira and Parker, 2007; Teixeira et al., 2005).

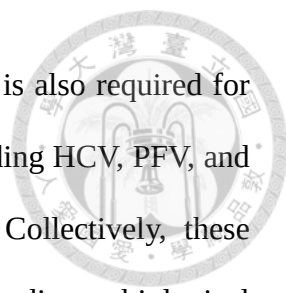


It was further demonstrated that the PB assembly required the functional silencing and the decay pathways, but not vice versa (Eulalio et al., 2007a). These findings sparked the notion that PB assembly is a consequence but not the cause of gene silencing (Eulalio et al., 2007a). Thus, to a large extent, the PBs can reflect the gene silencing status of the host cells. In this sense, the loss of PB following the loss of the DDX6 homologs in human and *Drosophila* cells not only indicates that the metazoan PB assembly depends on the DDX6 homologs, but also highlights their roles in gene silencing. Recently, a report displayed that DDX6, 4E-T, and LSM14A are strictly required for PB assembly, even under cytotoxic stress (Ayache et al., 2015). Consistently, these three proteins are the effectors of translational repression. Notwithstanding, the DDX6 homologs not only promotes the PB assembly by modulating the cell's translational status. One study has demonstrated that Dhh1p can nucleate PB assembly on the targeted RNA in *S. cerevisiae* (Carroll et al., 2011), indicating a direct role for DDX6 in the PB assembly. Consistently, the DDX6 homologs contribute to the PB assembly through maintaining normal PB dynamics (Carroll et al., 2011; Hubstenberger et al., 2013; Jangra et al., 2010). In *C. elegans*, it was observed that CAR-1 crystalizes into sheet-like solid in the absence of CGH-1 (Hubstenberger et al., 2013). In *S. cerevisiae* and human cells, the ATPase-inactive mutants of Dhh1p and DDX6 displayed aberrant PB dynamics (Carroll et al., 2011; Jangra et al., 2010), indicating potential involvement of the ATPase-dependent RNP remodeling in normal PB dynamics. However, the mechanism of the PB assembly and the DDX6's role in it remains to be investigated.

4.4. The ATPase activity of DDX6



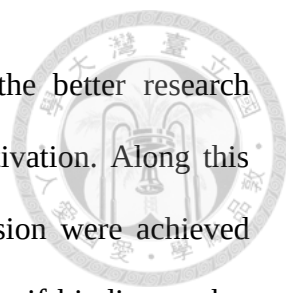
The ATPase activity of DDX6 is essential to its functions. On the molecular level, the DDX6's ATPase activity is required for translational repression (Coller and Parker, 2005; Kuzuoglu-Ozturk et al., 2016; Mathys et al., 2014; Minshall et al., 2009; Minshall and Standart, 2004; Minshall et al., 2001). As for the activation of decapping, both supporting and opposing evidence for the need of the DDX6's ATPase activity have both been provided and no clear consensus has been reached (Carroll et al., 2011; Sweet et al., 2012). On the supra-molecular level, the ATPase activity of the DDX6 homologs is necessary for the assembly and the proper dynamics of the PBs. In *S. cerevisiae*, the ATPase-dead Dhh1p mutant formed constitutive PBs and was unable to exchange with the cytoplasmic pool of Dhh1p (Carroll et al., 2011). In human cells, the ATPase-dead DDX6 mutant failed to rescue the loss of PBs in the absence of WT DDX6 (Jangra et al., 2010; Minshall et al., 2009). Similarly, in myotonic dystrophy type 1 (DM1)-affected cells, the ATPase-dead DDX6 mutant was unable to decrease the number of the CUG-expanded RNA foci number like WT DDX6 (Pettersson et al., 2014). On the physiological level, the ATPase activity was required for the DDX6's role in controlling cell fate decision in neural and epidermal progenitor cells (Nicklas et al., 2015; Wang et al., 2015). In neural progenitor cells, the ATPase-dead DDX6 mutant failed to induce neural differentiation like the WT DDX6 (Nicklas et al., 2015). In epidermal progenitor cells, the ATPase-dead DDX6 mutant failed to rescue the premature differentiation triggered by the DDX6 KD (Wang et al., 2015). In the Δ Dhh1p *S. cerevisiae*, the ATPase-dead Dhh1p mutant could not rescue the sensitivity to restrictive temperature and the DNA damages (Carroll et al., 2011; Cheng et al., 2005; Dutta et al., 2011).



Beyond the physiological conditions, the ATPase activity of DDX6 is also required for the efficient replication of the ssRNA viruses in the host cells, including HCV, PFV, and HIV (Jangra et al., 2010; Reed et al., 2012; Yu et al., 2011). Collectively, these observations mark the significance of DDX6's ATPase activity across diverse biological contexts.

Surprisingly, both Dhh1p and the human DDX6 have low or even barely detectable ATPase activity in vitro (Dutta et al., 2011). This is attributed to the auto-inhibition caused by the strong intramolecular interactions within the helicase cores of the DDX6 homologs, which prevent the conformational switch from the “off” state to the “on” state (Dutta et al., 2011). These auto-inhibitory interactions are disrupted by CNOT1, the scaffolding component of the multifunctional CCR4-NOT complex (Mathys et al., 2014). CNOT1 further provide structural constraints that favor the ATPase-active conformation of DDX6. Together, these findings indicate the ATPase activity of the DDX6 homologs is usually kept dormant in the absence of activators. In other words, the ATPase activity of the DDX6 homologs is under tight regulation, even though it is widely demanded in various kinds to biological processes. Except for CNOT1, it is possible that there exist other activator proteins, but this possibility has yet to be investigated.

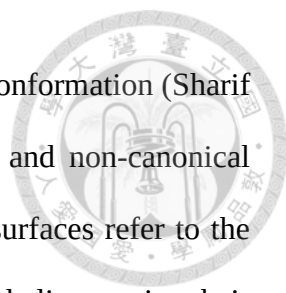
The CNOT1 binding model has provided us a mechanistic understanding of the DDX6's ATPase activation. With this knowledge, the problem now narrows down to how the activation of DDX6 translates into biological functions. As compared to the physiological outcomes, the molecular processes in translational repression, activation of decapping, and PB assembly are more likely to be the immediate consequences to



ATP hydrolysis. Therefore, these molecular processes represent the better research targets for dissecting the exact reactions following the ATPase activation. Along this line, a recent report demonstrated that similar levels of de-repression were achieved when mutations were introduced to the DEAD motif and the FDF motif-binding pocket (Kuzuoglu-Ozturk et al., 2016). The FDF motif-binding pocket is the common binding site for the other decapping activators and translational repressors, namely, EDC3, PATL1, LSM14A/B, and 4E-T (Kuzuoglu-Ozturk et al., 2016; Ozgur et al., 2015a; Ozgur and Stoecklin, 2013; Sharif et al., 2013; Tritschler et al., 2009; Tritschler et al., 2008). This result indicates the possibility that DDX6 rearranges the targeted mRNP in an ATP-dependent manner and fosters the deposition of the downstream effector proteins. Furthermore, since PATL1, LSM14A, and 4E-T all contribute substantially to the PB assembly (Ayache et al., 2015; Ozgur and Stoecklin, 2013), the DDX6's ATPase may likewise contribute to the PB assembly through RNP remodeling. As such, the molecular mechanisms that lie herein are at the core of the DDX6's function in gene silencing.

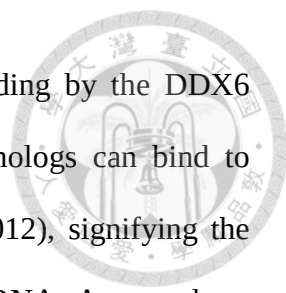
4.5. The RNA binding activity of DDX6

The RNA binding activity is the other fundamental activity of the DDX6. One crucial aspect of the RNA binding activity is the RNA binding surface. Like the other DBPs, several conserved helicase motifs are known to contribute to Dhh1p's RNA binding (Cheng et al., 2005; Dutta et al., 2011). Additionally, the motif I and IV, which are generally known to promote ATP binding, also aid in Dhh1p's RNA binding (Cheng et al., 2005; Dutta et al., 2011). Surprisingly, with the power of the RNA crosslinking followed by mass spectrometry (RNA XL-MS), one study demonstrated that Dhh1p

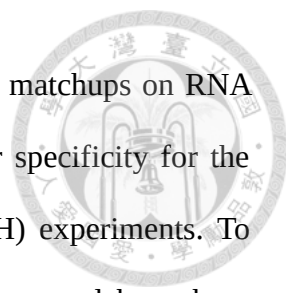


relies primarily on its CTD in RNA binding in the ATPase-inactive conformation (Sharif et al., 2013). In addition, the Dhh1p CTD utilizes both canonical and non-canonical surfaces in RNA binding (Sharif et al., 2013). Here, the canonical surfaces refer to the RNA binding surfaces commonly used by other DEAD-box helicases in their ATPase-active conformation, including the helicase motifs. Notably, these non-canonical surfaces coincide with the FDF motif-binding pocket which mediates the binding with Pat1p and Edc3p (Sharif et al., 2013). In Me31B and the human DDX6, the FDF motif-binding pocket also facilitates the binding of LSM14A and 4E-T (Ozgun et al., 2015a; Triteschler et al., 2009; Triteschler et al., 2008). This highlights the mutually exclusive binding of RNA and the translational repressors/decapping activators. Consistently, the integrity of the FDF motif-binding pocket is essential for DDX6 mediated gene silencing (Kuzuoglu-Ozturk et al., 2016). Together, these data suggests that substantial RNP remodeling is required for DDX6 to transit between the RNA-bound state to repressors/decapping activators-bound states. Notably however, the ATPase-inactive conformation differs between Dhh1p and the human DDX6 (Ozgun et al., 2015b). In this sense, DDX6 may utilize different RNA binding surfaces in its ATPase-inactive conformation, which may not coincide with the FDF motif-binding pocket. Currently, this remains a gap in our knowledge.

Substrate specificity is another essential aspect to the RNA binding activity. Previous studies have demonstrated that the DDX6 homologs can bind single-stranded poly(U), poly(G), partial dsRNA, and dsRNA of sufficient length (Cheng et al., 2005; Dutta et al., 2011; Ernoult-Lange et al., 2012). It has also been shown the human DDX6 can bind polyadenylated mRNA as efficiently as other in vitro synthesized RNA

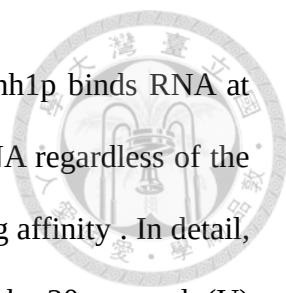


(Ernoul-Lange et al., 2012). These results indicate the RNA binding by the DDX6 homologs is not sequence-specific. Furthermore, The DDX6 homologs can bind to multiple sites on the same targeted RNA (Ernoul-Lange et al., 2012), signifying the DDX6 homologs can tolerate the varying sequences on the same RNA. Among these studies, one utilized the RNA crosslinking followed by immunoprecipitation and deep sequencing (CLIP-seq) but found no census target sequence for Dhh1p (Mitchell et al., 2013). The same study also showed that Dhh1p has little positional preference (in terms of 5'-3' directionality) along the targeted RNA (Mitchell et al., 2013). More importantly, since the activated DDX6 adopts the canonical ATPase-active conformation (Mathys et al., 2014), it is likely that the activated DDX6 interacts with the sugar-phosphate backbone of its RNA substrate but not with the nitrogenous bases. Collectively, these findings indicate that DDX6 does not bind RNA with evident sequence-specificity. Nonetheless, several studies claimed DDX6 selectively binds to several distinct RNA structures, namely, the DB1 and DB2 pseudoknots of DENV RNA (Ward et al., 2011), the MYC mRNA IRES (Matsui et al., 2006), the HIF1A mRNA IRES (Saito et al., 2011), the VEGFA mRNA IRES (de Vries et al., 2013), and the CUG trinucleotide-expanded repeats of DMPK mRNA (Pettersson et al., 2014). These cases can be appreciated in that they illustrated how DDX6 is capable of binding the structured and the noncoding RNA regions. However, the concept of specificity in RNA binding is not universally and rigorously defined, leaving sufficient space for differential interpretations. In these studies, DDX6 was considered to bind these RNA “specifically” as the control group did not achieve the same level of enrichment after affinity purifications (de Vries et al., 2013; Pettersson et al., 2014; Saito et al., 2011; Ward et al., 2011). However, the choices for control RNA were arbitrary and largely

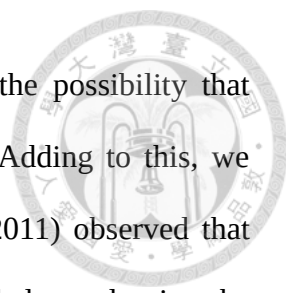


failed to match up on RNA length and nucleotide content. Yet both matchups on RNA length and nucleotide content are routinely demanded as proof for specificity for the biological probes such as those used the in situ hybridization (ISH) experiments. To resolve this discrepancy, we could borrow some insights from the rigorous debate about the RNA binding specificity of the polycomb repressive complex 2 (PRC2) (Cifuentes-Rojas et al., 2014; Davidovich et al., 2015; Davidovich et al., 2013). In this debate, different qualities of RNA binding are defined, including specific, promiscuous, and non-specific binding. Accordingly, the RNA binding by the DDX6 homologs can be best described as promiscuous (Davidovich et al., 2015), as they bind many RNA substrates without defined target motifs and without large difference in affinity (presumably within one order of magnitude). Also, for promiscuous RNA binding proteins, the RNA length can largely determine the overall affinity and will ultimately affect the apparent specificity (Davidovich et al., 2015; Davidovich et al., 2013). However, promiscuity and specificity are not mutually exclusive. It is first because the in vitro measurement for specificity is alterable on different experimental settings (Davidovich et al., 2015). It is also because the definition of promiscuity actually allows for moderately different affinity between different RNA substrates (Davidovich et al., 2015). On top of this, in the true biological context, the RNA binding of DDX6 may be inherently promiscuous, as its RNA substrates are at least partially determined by the regulators with higher specificity that act upstream to the CCR4-NOT/DDX6 axis.

The other crucial aspect about the RNA binding activity is the general affinity towards RNA. This can be measured on the absolute scales and is often expressed in the form of the dissociation constant (K_d). This has been systematically explored in Dhh1p

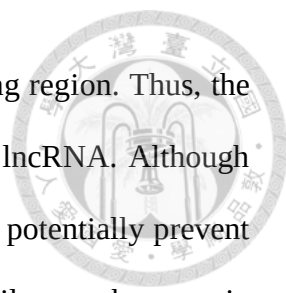


and the human DDX6 by two studies. Dutta et al. reported that Dhh1p binds RNA at very high affinity, displaying $K_d \approx 2$ nM for the 20 mer poly(U) RNA regardless of the ATP. The same study also reported the length dependence of binding affinity. In detail, Dhh1p displayed up to 500 fold difference in K_d for the 5 mer and the 20 mer poly(U) RNA. Furthermore, the affinity increased abruptly (~ 15 fold) from the 10 mer to the 12 mer poly(U) RNA (Dutta et al., 2011), suggesting Dhh1p binds preferentially to 12 mer or longer RNA. Ernoult-Lange et al. also reported human DDX6 binds RNA at very high affinity, with $K_d \approx 1$ nM regardless of the ATP. Furthermore, Ernoult-Lange et al. also showed that the RNA binding by DDX6 is cooperative, with the Hill coefficient ≈ 2.1 . Notably, human DDX6 binds tRNA with significantly lower affinity (Ernoult-Lange et al., 2012), even lower than dsDNA, suggesting tRNA is not a physiological target for DDX6. These quantitative measurements have widespread functional implications for DDX6 in vivo. First, since DDX6 in its ATPase-inactive conformation binds RNA with high affinity in vivo, the cellular reservoir of inactive DDX6 may largely be RNA-bound. Second, considering DDX6 binds RNA promiscuously, it can bind a large number of different RNA in vivo (Chap. 3). Exceptions may include the other short and extensively structured RNA like tRNA (Ernoult-Lange et al., 2012). Third, Ernoult-Lange et al. estimated that DDX6 vastly outnumbered its potential RNA substrates in vivo, thus the repressed targets are very likely to be bound by multiple DDX6 proteins (Ernoult-Lange et al., 2012). Interestingly, both Dutta et al. and Ernoult-Lange et al. reported that Dhh1p and DDX6 also binds DNA (both ssDNA and dsDNA) of sufficient length with high affinity, albeit at $1 \sim 2$ orders of magnitude (K_d between $10 \sim 100$ nM) lower than the affinity with RNA (Dutta et al., 2011; Ernoult-Lange et al., 2012). Although both studies interpreted their results



as RNA being the preferential target, both cannot fully dismiss the possibility that DDX6 may bind to DNA under certain physiological conditions. Adding to this, we (Chap. 2) and several studies (Ohta et al., 2010; Torrente et al., 2011) observed that DDX6 can bind to chromatin or mitotic chromosomes. Nonetheless, despite the chromatin is the largest cellular reservoir for DNA, it is also highly rich in RNA and proteins. Therefore, the exact elements that mediate the chromatin binding of DDX6 require further inspection.

It has been proposed that the DBPs can participate in the regulation of the lncRNA (Jarmoskaite and Russell, 2014). This idea well-applies to the DDX6 homologs. In fact, we view the DDX6's promiscuity in RNA binding as a clear indication for its role in the regulation of the lncRNA. The lncRNA is a diverse class of the ncRNA that is arbitrarily defined by length longer than 200 nucleotide (Guttman and Rinn, 2012; Mercer and Mattick, 2013; Rinn and Chang, 2012), except for the ribosomal RNA (rRNA) and the other previously known ncRNA. The lncRNA genes are prevalent in the mammalian genomes and are widely expressed. Most lncRNA genes are transcribed by RNA polymerase II (RNAP II) similar to the protein-coding genes. Consequently, the lncRNA shares common anatomy with the mRNA. Both the lncRNA and the mRNA are 5'-capped, often spliced, and sometimes polyadenylated. The defining distinction between the lncRNA and the mRNA is the presence of the complete coding sequences (CDS). Yet the RNA binding activity of DDX6 is independent of the translational status of its RNA substrates (Drummond et al., 2011). In addition, a complete CDS often spans hundreds to thousands of nucleotides in length and is far beyond the cognitive capability of the non-processive enzymes such as DDX6 itself. In this sense, we argue that DDX6

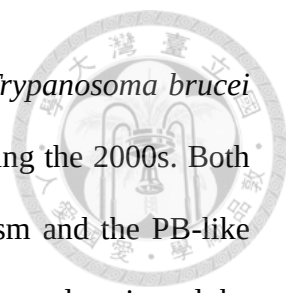


actually perceives no distinction between the CDS and the noncoding region. Thus, the lack of a complete CDS should not barricade DDX6 from binding lncRNA. Although the lncRNA can form extensive secondary or tertiary structures that potentially prevent the DDX6 binding, they can also adopt modular structures similar to the protein domains, which will leave some unstructured linker regions available for DDX6 binding (Mercer and Mattick, 2013). The core of our argument is, on local scales (for example, tens of nucleotide in length), the lncRNA can be indistinguishable from the mRNA to the DDX6 molecules. This indistinguishability therefore justifies the feasibility of the lncRNA binding by DDX6 and our treatment for the lncRNA as the near equivalence of the mRNA in terms of the RNA binding by DDX6. Supporting this idea, it was found that DDX6 can bind to noncoding region of the mRNA (de Vries et al., 2013; Pettersson et al., 2014; Saito et al., 2011; Ward et al., 2011). It is also supported by the two studies that reported specific lncRNA binding by DDX6 (Iio et al., 2013; Yoon et al., 2012). Consistently, we discovered the pervasive binding of lncRNA by DDX6 in human cell lines (Chap. 3). On a broader scale, it has also been shown that the miRNA can target lncRNA (Tay et al., 2014). Consequently, DDX6 can be recruited to the targeted lncRNA as a component of the silencing complex. Since many of the lncRNA genes are transcribed by RNAP II (Mercer and Mattick, 2013), their products are polyadenylated and capped and therefore succumb to deadenylation and decapping. In this sense, DDX6 may contribute to lncRNA decay through facilitating decapping. DDX6 may also participate in lncRNA repression. A study reported that DDX6 can relax the targeted RNA into linear conformation in vitro (Ernoult-Lange et al., 2012). This observation resonates with the disruption of RNA tertiary structures by helix capture. Thus, it is possible that DDX6 represses the lncRNA through disrupting the functional RNA

structures. Moreover, the deposition of DDX6 on the targeted lncRNA may also trigger the assembly of the PBs, wherein the incorporated lncRNA molecules can be maintained in a prolonged repressed state. Together, these possibilities illustrate the potential involvement of DDX6 in lncRNA regulation.

4.6. The cytoplasmic localization of DDX6

The subcellular localization of the DDX6 subfamily members were characterized in the latter half of the 1990s and the 2000s, coinciding with the time of their functional characterization. In the late 1990s, the predominant cytoplasmic staining of the human DDX6 in the Raji B cell line was the first to be reported (Akao et al., 1995). This was followed by the predominant cytoplasmic staining of Xp54 in the developing *Xenopus* oocytes and the cultured tadpole XTC cell line (Ladomery et al., 1997). In the early 2000s, the predominant cytoplasmic staining of Dhh1p in *S. cerevisiae* was reported (Fischer and Weis, 2002). Shortly after, the PB was identified in both the *S. cerevisiae* and the human cells, with Dhh1p and DDX6 being the conserved PB components (Cougot et al., 2004; Sheth and Parker, 2003). Following these findings, the DDX6 homologs gained popularity as viable markers for the PBs in various organisms. As a result, their cytoplasmic and PB localization were widely confirmed in many subsequent studies. During the 2000s, several studies reported the predominant Me31B staining in the cytoplasm, the PBs, and other RNA granules in multiple *D. melanogaster* cell types, including the developing oocytes, the nurse cells, neurons, and the culture S2 cell line (Barbee et al., 2006; Eulalio et al., 2007b; Hillebrand et al., 2010; Nakamura et al., 2001). Likewise, in *C. elegans* and *D. japonica*, CGH-1 and DjCBC-1 were also reported with predominant staining in the cytoplasm and the granular structures

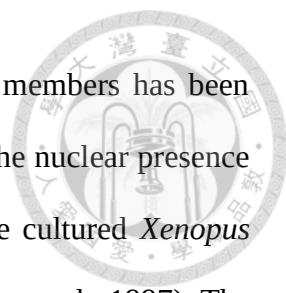


(Navarro et al., 2001; Yoshida-Kashikawa et al., 2007). DHH1 in *Trypanosoma brucei* and *Trypanosoma cruzi* were identified as the DDX6 homologs during the 2000s. Both of the *T. brucei* and *T. cruzi* DHH1 were expressed in the cytoplasm and the PB-like granules (Holetz et al., 2007; Tkacz et al., 2008). More recently, the cytoplasmic and the PB localization of Ste13 in *S. pombe* were reported (Wang et al., 2013). Collectively, these studies yielded a universal consensus on the predominant cytoplasmic distribution and the PB localization of the DDX6 homologs.

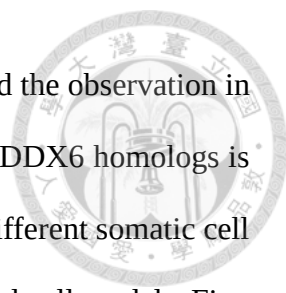
The functional relevance of the cytoplasmic distribution of DDX6 is explicit. First, the predominant cytoplasmic distribution of DDX6 is linked to its role in translational repression, for the mass of active translation machineries also localize in the cytoplasm. As described in the previous sections, there is accumulating evidence supporting that DDX6 indeed targets the cytoplasmic pool of translational machineries. Second, in the female germline cells, maternal messages can accumulate in the cytoplasm of the developing oocytes and the matured eggs before being actively translated. The cytoplasmic DDX6 homologs serve to mask these maternal messages. Third, the predominant cytoplasmic distribution and PB localization also coincide with the localization of other decay factors. Co-localizing in the same subcellular compartment conceivably grants more efficiency to the general RNA turnover.

4.7. The nuclear localization of DDX6

In contrast to the cytoplasmic localization and functions of the DDX6 homologs, the nuclear localization of the DDX6 homologs and their functional relevance are less understood. Yet the research literature describing the nuclear presence of DDX6 was not

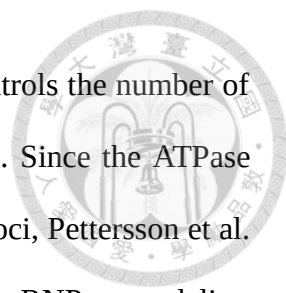


obsolete. In fact, the nuclear localization of the DDX6 subfamily members has been documented in various biological models since the early research. The nuclear presence of Xp54 was first displayed by the nuclear staining of Xp54 in the cultured *Xenopus* tadpole XTC cell line and the developing *Xenopus* oocytes (Ladomery et al., 1997). The nuclear localization of Xp54 was confirmed by Western blot in the nuclear extracts prepared from the early stage but not the late stage oocytes (Smillie and Sommerville, 2002). These results were further complemented by the positive staining of Xp54 on the isolated lampbrush chromosomes from the early stage oocytes (Smillie and Sommerville, 2002). Following observation of the nuclear Xp54, the nuclear localization of the DDX6 homologs were observed in the clam *Spisula solidissima* (*Spisula*) and the planarian *D. japonica*. Substantial nuclear staining of p47, the *Spisula* homolog of DDX6, was observed in the nuclei of the 2 cell- and 8 cell-stage embryos (Minshall et al., 2001). This and the observation in the *Xenopus* oocytes together imply the nuclear localization of the DDX6 homologs can be developmentally regulated. In *D. japonica*, the nuclear localization of DjCBC-1 in the neoblasts was revealed by immuno-electron microscopy (IEM). Surprisingly, 11 out of 14 (~80%) examined neoblasts contained speckled nuclear staining of DjCBC-1 (Yoshida-Kashikawa et al., 2007), suggesting the nuclear localization of DjCBC-1 is widespread among the planarian neoblasts. The DDX6 homologs were also observed in the mammalian cells. With IEM, the nuclear localization of the DDX6 homologs was observed in the various cell types within the male gonads of the common laboratory rodents (Kawahara et al., 2014). More specifically, the nuclear staining of DDX6 was observed in both spermatogenic cells and somatic cells (Kawahara et al., 2014). This and the observation in the *Xenopus* oocytes imply the nuclear localization of the vertebrate DDX6 can be



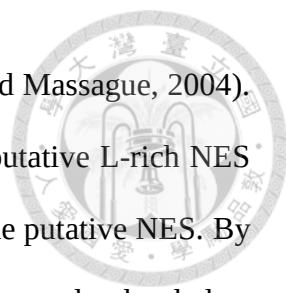
widespread in the germline cells during gametogenesis. Also, this and the observation in the tadpole XTC cells also imply that the nuclear localization of the DDX6 homologs is not restricted in the germline cells, but can be prevalent among the different somatic cell types. In human, DDX6 has been observed in the cell nuclei of several cell models. First, DDX6 was identified in the chromatin extract of the HeLa-S3 cell line by mass spectrometry (MS) (Torrente et al., 2011). Second, DDX6 was detected in the nuclear extract of the gastric cancer MKN45 cell line by Western blot (Iio et al., 2013). Third, the punctate nuclear staining of DDX6 was observed in the cells affected by DM1 and coincided with the pathological CUG trinucleotide-expanded DMPK mRNA foci (Pettersson et al., 2014). In addition to these observations, we also confirmed the nuclear localization of DDX6 in other human cell lines (Chap. 2). In *S. cerevisiae*, the nuclear localization of Dhh1p was implicated by the association with chromatin. More detailedly, the chromatin binding of Dhh1p at the promoter regions and across the gene bodies were reported by two independent studies (Haimovich et al., 2013; Kruk et al., 2011). Collectively, these results indicate that the nuclear localization can be prevalent across the eukaryotes, and can also be widespread among various cell types under different physiological conditions in the metazoans.

So far, several potential nuclear functions for the DDX6 homologs have been proposed. Sommerville and colleagues proposed that nuclear Xp54 binds to the nascent transcripts and nucleates stored mRNP within the cell nuclei in the early stage *Xenopus* oocyte (Smillie and Sommerville, 2002). Since the transcribed mRNA is largely stored in the cytoplasm during the oocyte maturation, they further proposed that Xp54 can facilitate the export of the stored mRNP (Smillie and Sommerville, 2002). In the nuclei

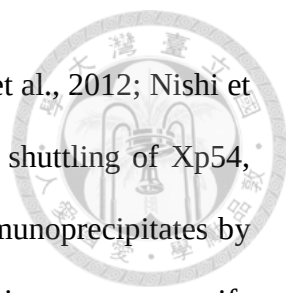


of the DM1-affected cells, Pettersson et al. proposed that DDX6 controls the number of the pathological CUG-expanded RNA foci (Pettersson et al., 2014). Since the ATPase activity was required for the reduction of the CUG-expanded RNA foci, Pettersson et al. argued that DDX6 disseminates the pathological foci through RNP remodeling (Pettersson et al., 2014). In *S. cerevisiae*, it was proposed that Dhh1p can positively regulate transcription. Kruk et al. demonstrated that Dhh1p binds RNAP II as a component of the CCR4-NOT complex (Kruk et al., 2011). The CCR4-NOT complex, in turn, aids the transcription elongation by releasing the “backtracking” RNAP II (Kruk et al., 2011). On the other hand, Haimovich et al. demonstrated that Dhh1p, as well as Xrn1p, Dcp2p, Lsm1p, and Pat1p, can bind to promoters and stimulate transcription. Haimovich et al. also showed that the 5’-3’ decay factors take part in the transcription elongation in similar manners to the CCR4-NOT complex. Thus, Haimovich et al. proposed that the CCR4-NOT complex, the 5’ -3’ decay factors, and Dhh1p can act in the same “mega” protein complex to regulate transcription (Haimovich et al., 2013).

One fundamental issue about the nucleocytoplasmic localization of the DDX6 homologs is the governing mechanism behind the distribution. Sommerville and colleagues, based on their findings in the Xp54, proposed an influential model (Smillie and Sommerville, 2002; Weston and Sommerville, 2006), which we will refer to as the shuttling model. According to the shuttling model, the vertebrate DDX6 homologs possess both a putative lysine/arginine-rich nuclear localization signal (K/R-rich NLS) and a putative leucine-rich nuclear export signals (L-rich NES) in their NTD (Smillie and Sommerville, 2002; Weston and Sommerville, 2006). The K/R-rich NLS and the L-rich NES label their host proteins as cargoes for the importin- α/β nuclear import



pathway and the CRM1 nuclear export pathway, respectively (Xu and Massague, 2004). For simplicity, we will refer to the putative K/R-rich NLS and the putative L-rich NES of the vertebrate DDX6 homologs simply as the putative NLS and the putative NES. By implication, the vertebrate DDX6 homologs are actively and continuously shuttled to and fro between the cytoplasm and the cell nuclei. In this sense, the nucleocytoplasmic distribution of the vertebrate DDX6 homologs represents the equilibrium of the innate shuttling dynamics. However, the validity of the shuttling is greatly undermined by several major concerns. First, the putative NLS lacks conclusive supporting evidence. Second, attributing the shuttling capacity to the putative NLS and NES is inherently problematic. More detailedly, the argument that Xp54 being a shuttling protein was primarily established on the observation that the 1-164 segment of Xp54, which contains both the putative NLS and NES, can shuttle nucleocytoplasmically in a CRM1-dependent manner (Smillie and Sommerville, 2002). However, the full length Xp54 cannot shuttle in the same way as the 1-164 segment (Kirli et al., 2015; Smillie and Sommerville, 2002), implying the reductionist logic that dictated such reasoning cannot apply. To mend this contradiction, Sommerville and colleagues proposed a putative cytoplasm-retention signal (CRS) within the CTD of Xp54 as an ad hoc hypothesis (Smillie and Sommerville, 2002). However, this putative CRS lacks theoretical basis and had not been supported by any evidence since its proposal. Furthermore, we have just observed the opposite. The DDX6 CTD enters the nuclei through a novel mechanism (Chap. 2). Third, the CRM1-dependent shuttling of the DDX6 homologs has been challenged by the conflicting data presented by the following studies. While one study claimed decreasing nuclear DDX6 after LMB treatment (Iio et al., 2013), we (Chap. 2) and several other studies observed no alteration in the



nucleocytoplasmic distribution of DDX6 (Kirli et al., 2015; Marnef et al., 2012; Nishi et al., 2013). Interestingly, while not observing the CRM1-dependent shuttling of Xp54, Kirli et al. detected DDX6, Xp54, and Dhh1p, in the CRM1 co-immunoprecipitates by quantitative MS (Kirli et al., 2015). Notably, Dhh1p does not contain sequence motifs similar to the L-rich NES. As such, this finding suggested the DDX6 homologs could interact with CRM1 indirectly via other factors. Consistently, we confirmed that the human DDX6 can be transported by the CRM1 pathway via 4E-T (Chap. 2). However, our results refuted the notion of inherent shuttling of DDX6 through the importin- α/β and the CRM1 pathways (Chap. 2). Collectively, the Chap. 2 of this study presents a new synthesis on the governing mechanism of the DDX6's nucleocytoplasmic distribution.

4.8. Future perspectives

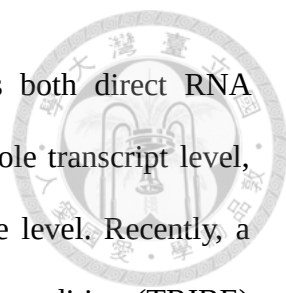
Twenty-five years have lapsed since the first discovery of the DDX6 subfamily protein Me31B. Great progress has been made in the understanding of the DDX6 subfamily. But there still are fundamental questions that await answers.

The most anticipated answers would be the exact mechanisms of how DDX6 elicits translational repression and RNA decay. Since DDX6 is known to form complexes with multiple interacting partners, and many of the interactions are mutually exclusive, this would require the resolution of how DDX6 forms and possibly alternates between the different complexes. Ingrained to this problem are the roles of the ATPase and the conformational regulation of DDX6. Despite remaining a tough challenge, this can surely be resolved down into great details, with the increasingly exquisite reporter



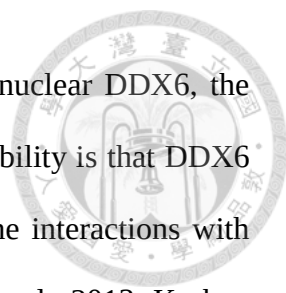
systems and the well-establishment of the structural biology approach. Relatedly, we now command the powerful technologies that could assess the extent of the regulatory functions of DDX6, which can provide a more holistic/realistic view of its role in tuning gene expression. With the next-generation sequencing technologies, we now can resolve the changes in the RNA expression level and the change in the translational status in impressive details. Direct profiling of the ribosome occupancy, or Ribo-seq, has become possible (Guo et al., 2010; Pelechano et al., 2015). In combination with the conventional RNA-seq, this strategy has proven to be effective in distinguishing the effects of RNA decay and translational repression (Eichhorn et al., 2014; Guo et al., 2010; Pelechano et al., 2015). With the advancing of the imaging and computational technologies, the translation dynamics can now be assessed at single cell resolution (Han et al., 2014). Given this advance, it is possible to obtain a dynamic view of the impact of DDX6 on translation. Moreover, since our functional analyses are dictated by the logic of the reverse genetics, the advent of the CRISPR technique can potentially improve the quality of our assessments for it allows for complete loss-of-function and lower perturbation to the systems as compared to the conventional RNAi methods.

The second most anticipated answer would be the determination of the physiological targets of DDX6. Even though DDX6 may not be sequence-specific per se, its RNA occupancy is influenced by the upstream factors and the RNA stoichiometry. That being said, the physiological targets of DDX6 may change according to the biological context. This requires the physiological binding targets to be explicitly determined by experiments under different conditions or different time points. This has been made possible by the RNA pulldown followed by deep sequencing methods,



including RIP-seq and CLIP-seq. The RIP-seq technique detects both direct RNA binding and indirect binding via other RNP components on the whole transcript level, whereas the CLIP-seq provides resolution down to the binding site level. Recently, a new sequencing-based technique called targets of RBPs identified by editing (TRIBE) has emerged (McMahon et al., 2016). TRIBE utilizes the ADAR catalytic domain as a protein tag for the RBP of interest, which results in the site-specific RNA editing of the RBP targets that can be detected by RNA-seq (McMahon et al., 2016). With the CRISPR tagging technique (Lackner et al., 2015), the ADAR tag can potentially be fused to the RBP of interest at its genomic locus, thereby minimizing perturbation to the studied systems. Furthermore, since the TRIBE technique utilized the standard RNA-seq methods, this technique has the promise to be coupled with the single-cell RNA-seq technique (McMahon et al., 2016). As such, the TRIBE technique can shed lights on the task of exploring the dynamics and the heterogeneity of the RNA binding by DDX6 across the cell population.

The other major questions include the potential role in regulating the lncRNA and the nuclear function of DDX6. For the regulation of the lncRNA, the advances in the tools of the reverse genetics, including CRISPR and RNA-seq, have poised for the in-depth interrogation of the impact of DDX6 on the expression level of the lncRNA. Furthermore, the recent advancement in the sequencing-based analysis of RNA structures has also poised for the assessment of the impact of DDX6 on RNA folding (Kwok et al., 2015). In conjunction, the smFRET approach has also generated invaluable insights to the structural regulation of RNA by the DBPs (Pan et al., 2014). These techniques have forearmed the functional exploration of DDX6 beyond the



facilitation of translational repression and RNA decay. As for the nuclear DDX6, the mission at hand is to elucidate its functional significance. One possibility is that DDX6 participates in the modulation of the transcription machinery via the interactions with the CCR4-NOT complex or the 5'-3' decay machinery (Haimovich et al., 2013; Kruk et al., 2011). This possibility has been explored in *S. cerevisiae* and awaits further investigations in the other biological models.

Although we concentrated on the molecular function of DDX6, there is no doubt that the genetic analyses of DDX6 in the different physiological or pathological conditions are equally important, and can contribute to the understanding of DDX6 or even to the discovery of the novel functions. Yet the ultimate goal of the molecular interrogation of DDX6 is to achieve the comprehensive understanding of the biochemical and the physical properties of DDX6 as well as its roles in the cellular physiology as viewed from the dynamical systems perspective. Ultimately, this knowledge should form a predictive model that could help us extrapolate the diverse physiological roles of the DDX6 homologs.

5. Tables

Table 1. The homologs of DDX6 and its interacting partners in the conventional model organisms

<i>Hsa</i>	<i>Mmu</i>	<i>Dre</i>	<i>Xla</i>	<i>Dme</i>	<i>Cel</i>	<i>Sce</i>
DDX6	Ddx6	Zp54	Xp54	Me31B	CGH-1	Dhh1p
CNOT1	Cnot1	Cnot1	Cnot1	Not1	LET-711	Not1p
4E-T ^{a1}	4e-t ^{a2}	4e-t ^{a3}	4e-t ^{a4}	4E-T	IFET-1	- ^d
PATL1 ^{b1}	Patl1	Patl1	Patl1	HPat ^{b2}	PATR-1	Pat1p
LSM14A ^{c1}	Lsm14a	Lsm14aa/ Lsm14ab	xRAPA ^{c3}	Tral	CAR-1	Scd6p
LSM14B ^{c2}	Lsm14b	Lsm14b	xRAPB ^{c4}	Tral	CAR-1	Scd6p
EDC3	Edc3	Edc3	Edc3	Edc3	- ^d	Edc3p
DCP1A	Dcp1a	Dcp1a	Dcp1a	Dcp1	DCAP-1	Dcp1p
DCP1B	Dcp1b	Dcp1b	Dcp1b	Dcp1	DCAP-1	Dcp1p
DCP2	Dcp2	Dcp2	Dcp2	Dcp2	DCAP-2	Dcp2p
XRN1	Xrn1	Xrn1	Xrn1	Pcm	XRN-1	Xrn1p

Hsa: *Homo sapiens*

Mmu: *Mus musculus*

Dre: *Danio rerio*

Xla: *Xenopus laevis*

Dme: *Drosophila melanogaster*

Cel: *Caenorhabditis elegans*

Sce: *Saccharomyces cerevisiae*

a1-4, Official gene symbol: EIF4ENIF1 (a1), Eif4enif1 (a2), Eif4enif1 (a3), Eif4enif1 (a4)

b1, Also alias as PAT1B; b2, official gene symbol: Patr-1

c1 and 2, Also aliases as RAP55A and RAP55B

c3 and c4, also aliases as xRAP55A and xRAP55B, official gene symbols: Lsm14a and Lsm14b

d, No known homolog

Table 2. The homologs of DDX6 in other model organisms

<i>Hsa</i>	<i>Dja</i>	<i>Sso</i>	<i>Tbr</i>	<i>Tcr</i>	<i>Spo</i>
DDX6	CBC-1	p47	DHH1	DHH1	Ste13

Hsa: *Homo sapiens*

Dja: *Dugesia japonica*

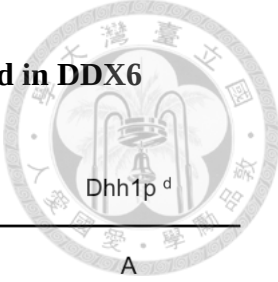
Sso: *Spisula solidissima*

Tbr: *Trypanosoma brucei*

Tcr: *Trypanosoma cruzi*

Spo: *Schizosaccharomyces pombe*

Table 3. Functions of the helicase motifs in general DBPs and in DDX6



Motif	DBP ^a	DBP ^b	Dhh1p ^c	Dhh1p ^d
Q	A	A	-	A
I	A	A	R	A/R
Ia	R	R	-	-
GG	R	R	-	-
Ib	R	R	-	-
II	A	A	-	A
III	A/R	A/R	-	-
IV	R	R	-	-
QxxR	R	R	-	-
V	R	A/R	R	A
VI	A	A	R	R

a-d, A: ATP binding; R: RNA binding

a, Linder and Jankowsky, Nat. Rev. Mol. Cell Biol. (2011)

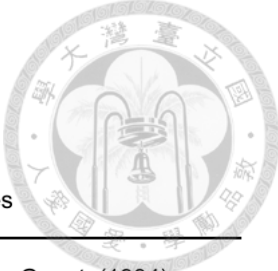
b, Henn et al., Annu. Rev. Biophys. (2012)

c, Cheng et al., RNA (2005); only tested RNA binding

d, Dutta et al., J. Biol. Chem. (2011); tested both RNA and ATP binding

c and d, -: No difference or not tested

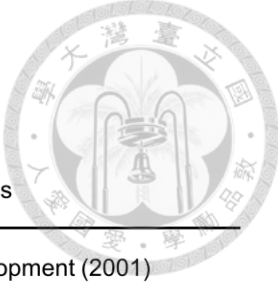
Table 4. Roles of the DDX6 homologs in stress responses



Protein	Stress	Regulation	References
Ste13	Nitrogen starvation	+	Maekawa et al., Mol. Gen. Genet. (1994)
Dhh1p	Restrictive temperature	+	Bergkessel and Reese, Genetics (2004) Cheng et al., RNA (2005) Carroll et al., J. Cell Biol. (2011) Dutta et al., J. Biol. Chem. (2011)
	DNA damage ^a	+	Bergkessel and Reese, Genetics (2004) Dutta et al., J. Biol. Chem. (2011)
	Starvation	+	Westmoreland et al, J. Surg. Res. (2003) ^b Coller and Parker, Cell (2005) ^c
CGH-1	DNA damage	+	Tomazella et al., J. Proteome Res. (2012)
CBC-1	γ -irradiation	+	Y. H. Wang's thesis (Chu lab unpublished data)
DDX6	Chemo/radio-therapy ^d	-	Cho et al., Int. J. Oncol. (2016)

a, Various damaging agents: hydroxyurea, methane sulfonate, UV, and γ -irradiation
b, Nitrogen starvation
c, Glucose and AA starvation
d, In patient derived glioblastoma cells

Table 5. Roles of the DDX6 homologs in sexual development



Protein	Process/ Location	Regulation	References
Me31B	Oogenesis	+	Nakamura et al., Development (2001)
CGH-1	Oogenesis Spermatogenesis	+	Navarro et al., Development (2001)
Xp54	Oocyte	NA ^a	Smillie and Sommerville, J. Cell Sci. (2002)
Ddx6	Oocyte Spermatocyte	NA ^b	Matsumoto et al., Dev. Dynam. (2005) Kawahara et al., Histochem. Cell Biol. (2014)
Ste13	Meiosis	+	Maekawa et al., Mol. Gen. Genet. (1994)
a and b, Not assessed .			

Table 6. Roles of the DDX6 homologs in somatic development

Class	Process	Protein	References
Ontogenesis & Regeneration	Embryogenesis	Ddx6	Matsumoto et al., Dev. Dynam. (2005)
		CGH-1	Audhya et al., J. Cell Biol. (2005)
	Larval development	Me31B	Nakamura et al., Development (2001)
	Regeneration	CBC-1	Rouhana et al., Dev. Biol. (2010) Y. H. Wang's thesis (Chu lab unpublished data)
	Epithelium formation	DDX6	Wang et al., Mol. Cell (2015)
Neurogenesis	Neural differentiation	Ddx6	Nicklas et al., Nucleic Acids Res. (2015)
		CBC-1	Y. H. Wang's thesis (Chu lab unpublished data)
	Dendrite aborization	Me31B	Barbee et al., Neuron (2006)

Table 7. Roles of DDX6 in RNA virus and retrotransposon replication

Class	Name	Protein	Regulation	References
Group IV +ssRNA	HCV	DDX6	+	Scheller et al., Proc. Acad. Natl. Sci. USA (2009) Jangra et al., J. Virol. (2010) Huys et al., PLoS ONE (2013) Chatel-Chaix et al., J. Virol. (2013)
	DENV		+	Ward et al., RNA Biol. (2011)
	WNV		+	Chahar et al., Virology (2013)
	BMV	Dhh1p	+	Alves-Rodrigues et al., J. Virol. (2007)
	FHV		+	Gimenez-Barcons et al., J. Virol. (2013)
Group VI ssRNA-RT	HIV	DDX6	+	Reed et al., J. Cell Biol. (2012) Robinson et al., J. Virol. (2014)
			-	Chable-Bessia et al., Retrovirology (2009) Nathans et al., Mol. Cell (2009)
	PFV		+	Yu et al., PLoS Pathog. (2011)
	RVFV	Me31B	-	Hopkins et al., Genes Dev. (2013)
Retro- transposon	Several	Me31B	-	Liu et al., Development (2011)
	Ty1	Dhh1p	+	Checkley et al., Mol. Cell Biol. (2010)
	Ty3		+	Bilanchone et al., PLoS Genet. (2015)

Table 8. Roles of DDX6 in cancer development



Cancer	Context	Involvement	References
Liver cancer	Liver	HCV infection	Miyaji et al., J. Viral Hepatitis (2003)
Colon cancer	Colon	Over-expression	Nakagawa et al., Brit. J. Cancer (1999) Hashimoto et al., Carcinogenesis (2001) Taniguchi et al., Biochim. Biophys. Acta (2015)
	LoVo cell line	Proliferation	Lin et al., Cancer Biol. Ther. (2008)
	DLD-1 cell line	Survival	Taniguchi et al., Biochim. Biophys. Acta (2015)
	WiDr cell line		
Glio-blastoma	U87MG cell line	Proliferation & Migration	Baskin et al., Sci. Rep. (2015)
Breast cancer	Primary tumor/ metastases	Gene duplication	Desouki et al., Cancer Invest. (2011)
Lung cancer	CL1-0 & CL1-5 cell lines	Migration & Invasion	Y. C. Chen's thesis (Chu lab unpublished data)

Table 9. The requirement of the ATPase activity in different biological context

Function	Protein	ATPase requirement	References
Translational repression	DDX6	+	Mathys et al., Mol. Cell (2014) Kuzuoglu-Ozturk et al., EMBO J. (2016)
	Xp54	+	Minshall et al., RNA (2001) Minshall and Standart, Nucleic Acids Res. (2004) Minshall et al., Mol. Biol. Cell (2009)
	Dhh1p	+	Coller and Parker, Cell (2005)
Facilitating RNA decapping	Dhh1p	+	Sweet et al., PLoS Biol. (2012)
		-	Carroll et al., J. Cell Biol. (2011)
PB formation	DDX6	+	Minshall et al., Mol. Biol. Cell (2009) Jangra et al., J. Virol. (2010)
	Dhh1p	+	Carroll et al., J. Cell Biol. (2011)
Decreasing CUG foci	DDX6	+	Pettersson et al., Nucleic Acids Res. (2014)
Epithelium formation	DDX6	+	Wang et al., Mol. Cell (2015)
Neural differentiation	Ddx6	+	Nicklas et al., Nucleic Acids Res. (2015)
Stress response	Dhh1p	+	Cheng et al., RNA (2005) ^{a1} Carroll et al., J. Cell Biol. (2011) ^{a2} Dutta et al., J. Biol. Chem. (2011) ^{a3}
ssRNA virus replication	DDX6	+	Jangra et al., J. Virol. (2010) ^{b1} Yu et al., PLoS Pathog. (2011) ^{b2} Reed et al., J. Cell Biol. (2012) ^{b3}
Meiosis	Ste13	+	Maekawa et al., Mol. Gen. Genet. (1994)

a1, Heat

a2, Heat

a3, Hydroxyurea, methane sulfoxide, UV, and heat

b1, HCV

b2, PFV

b3, HIV

Table 10. The nucleocytoplasmic distribution of the DDX6 homologs



Protein	Cyto.	Nuc.	References
DDX6	+	+	Torrente et al., PLoS ONE (2011) ^a lio et al., Biochim. Biophys. Acta (2013) Pettersson et al., Nucleic Acids Res. (2014) Chap. 3 of this work
Ddx6	+	+	Kawahara et al., Histochem. Cell Biol. (2014)
Xp54	+	+	Ladomery et al., Nucleic Acids Res. (1997) Smillie and Sommerville, J. Cell Sci. (2002)
p47	+	+	Minshall et al., RNA (2001)
CBC-1	+	+	Yoshida-Kashikawa et al., Dev. Dynam. (2007)
Dhh1p	+	+	Kruk et al., Genes Dev. (2011) ^b Haimovich et al., Cell (2013) ^c
a-c, Inferred by indirect evidence			

6. References

Akao, Y., Marukawa, O., Morikawa, H., Nakao, K., Kamei, M., Hachiya, T., and Tsujimoto, Y. (1995). The rck/p54 candidate proto-oncogene product is a 54-kilodalton D-E-A-D box protein differentially expressed in human and mouse tissues. *Cancer Res.* 55, 3444-3449.

Akao, Y., and Nakagawa, Y. (2006). Expression of the DEAD-box/RNA helicase rck/p54 in mouse tissues: implications for heterogeneous protein expression. *J. Histochem. Cytochem.* 54, 955-960.

Akao, Y., Seto, M., Yamamoto, K., Iida, S., Nakazawa, S., Inazawa, J., Abe, T., Takahashi, T., and Ueda, R. (1992). The RCK gene associated with t(11;14) translocation is distinct from the MLL/ALL-1 gene with t(4;11) and t(11;19) translocations. *Cancer Res.* 52, 6083-6087.

Alves-Rodrigues, I., Mas, A., and Diez, J. (2007). *Xenopus* Xp54 and human RCK/p54 helicases functionally replace yeast Dhh1p in brome mosaic virus RNA replication. *J. Virol.* 81, 4378-4380.

Ament, S.A., Corona, M., Pollock, H.S., and Robinson, G.E. (2008). Insulin signaling is involved in the regulation of worker division of labor in honey bee colonies. *Proc. Natl. Acad. Sci. USA* 105, 4226-4231.

Ament, S.A., Wang, Y., Chen, C.C., Blatti, C.A., Hong, F., Liang, Z.S., Negre, N., White, K.P., Rodriguez-Zas, S.L., Mizzen, C.A., *et al.* (2012). The transcription factor ultraspiracle influences honey bee social behavior and behavior-related gene expression. *PLoS Genet.* 8, e1002596.

Arribere, J.A., Doudna, J.A., and Gilbert, W.V. (2011). Reconsidering movement of eukaryotic mRNAs between polysomes and P bodies. *Mol. Cell* 44, 745-758.

Audhya, A., Hyndman, F., McLeod, I.X., Maddox, A.S., Yates, J.R., 3rd, Desai, A., and Oegema, K. (2005). A complex containing the Sm protein CAR-1 and the RNA helicase CGH-1 is required for embryonic cytokinesis in *Caenorhabditis elegans*. *J. Cell Biol.* 171, 267-279.

Ayache, J., Benard, M., Ernoult-Lange, M., Minshall, N., Standart, N., Kress, M., and

Weil, D. (2015). P-body assembly requires DDX6 repression complexes rather than decay or Ataxin2/2L complexes. *Mol. Biol. Cell* 26, 2579-2595.

Barbee, S.A., Estes, P.S., Cziko, A.M., Hillebrand, J., Luedeman, R.A., Collier, J.M., Johnson, N., Howlett, I.C., Geng, C., Ueda, R., *et al.* (2006). Staufen- and FMRP-containing neuronal RNPs are structurally and functionally related to somatic P bodies. *Neuron* 52, 997-1009.

Baskin, R., Woods, N.T., Mendoza-Fandino, G., Forsyth, P., Egan, K.M., and Monteiro, A.N. (2015). Functional analysis of the 11q23.3 glioma susceptibility locus implicates PHLDB1 and DDX6 in glioma susceptibility. *Sci. Rep.* 5, 17367.

Beelman, C.A., and Parker, R. (1995). Degradation of mRNA in eukaryotes. *Cell* 81, 179-183.

Bergkessel, M., and Reese, J.C. (2004). An essential role for the *Saccharomyces cerevisiae* DEAD-box helicase DHH1 in G1/S DNA-damage checkpoint recovery. *Genetics* 167, 21-33.

Bhandari, D., Raisch, T., Weichenrieder, O., Jonas, S., and Izaurralde, E. (2014). Structural basis for the Nanos-mediated recruitment of the CCR4-NOT complex and translational repression. *Genes Dev.* 28, 888-901.

Bilanchone, V., Clemens, K., Kaake, R., Dawson, A.R., Matheos, D., Nagashima, K., Sitlani, P., Patterson, K., Chang, I., Huang, L., *et al.* (2015). Ty3 Retrotransposon Hijacks Mating Yeast RNA Processing Bodies to Infect New Genomes. *PLoS Genet.* 11, e1005528.

Boag, P.R., Atalay, A., Robida, S., Reinke, V., and Blackwell, T.K. (2008). Protection of specific maternal messenger RNAs by the P body protein CGH-1 (Dhh1/RCK) during *Caenorhabditis elegans* oogenesis. *J. Cell Biol.* 182, 543-557.

Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* 126, 713-725.

Bowers, H.A., Maroney, P.A., Fairman, M.E., Kastner, B., Luhrmann, R., Nilsen, T.W., and Jankowsky, E. (2006). Discriminatory RNP remodeling by the DEAD-box protein DED1. *RNA* 12, 903-912.

Braun, J.E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Mol. Cell* 44, 120-133.

Brengues, M., Teixeira, D., and Parker, R. (2005). Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies. *Science* 310, 486-489.

Carroll, J.S., Munchel, S.E., and Weis, K. (2011). The DExD/H box ATPase Dhh1 functions in translational repression, mRNA decay, and processing body dynamics. *J. Cell Biol.* 194, 527-537.

Chable-Bessia, C., Meziane, O., Latreille, D., Triboulet, R., Zamborlini, A., Wagschal, A., Jacquet, J.M., Reynes, J., Levy, Y., Saib, A., *et al.* (2009). Suppression of HIV-1 replication by microRNA effectors. *Retrovirology* 6, 26.

Chahar, H.S., Chen, S., and Manjunath, N. (2013). P-body components LSM1, GW182, DDX3, DDX6 and XRN1 are recruited to WNV replication sites and positively regulate viral replication. *Virology* 436, 1-7.

Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* 453, 544-547.

Chang, L.C., and Lee, F.J. (2012). The RNA helicase Dhh1p cooperates with Rbp1p to promote porin mRNA decay via its non-conserved C-terminal domain. *Nucleic Acids Res.* 40, 1331-1344.

Chatel-Chaix, L., Germain, M.A., Motorina, A., Bonneil, E., Thibault, P., Baril, M., and Lamarre, D. (2013). A host YB-1 ribonucleoprotein complex is hijacked by hepatitis C virus for the control of NS3-dependent particle production. *J. Virol.* 87, 11704-11720.

Checkley, M.A., Nagashima, K., Lockett, S.J., Nyswaner, K.M., and Garfinkel, D.J. (2010). P-body components are required for Ty1 retrotransposition during assembly of retrotransposition-competent virus-like particles. *Mol. Cell. Biol.* 30, 382-398.

Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. *Nat. Struct. Mol. Biol.* 18,

1218-1226.

Chen, J., Chiang, Y.C., and Denis, C.L. (2002). CCR4, a 3'-5' poly(A) RNA and ssDNA exonuclease, is the catalytic component of the cytoplasmic deadenylase. *EMBO J.* 21, 1414-1426.

Chen, Y., Boland, A., Kuzuoglu-Ozturk, D., Bawankar, P., Loh, B., Chang, C.T., Weichenrieder, O., and Izaurralde, E. (2014). A DDX6-CNOT1 complex and W-binding pockets in CNOT9 reveal direct links between miRNA target recognition and silencing. *Mol. Cell* 54, 737-750.

Cheng, Z., Collier, J., Parker, R., and Song, H. (2005). Crystal structure and functional analysis of DEAD-box protein Dhh1p. *RNA* 11, 1258-1270.

Cho, Y.J., Kang, W., Kim, S.H., Sa, J.K., Kim, N., Paddison, P.J., Kim, M., Joo, K.M., Hwang, Y.I., and Nam, D.H. (2016). Involvement of DDX6 gene in radio- and chemoresistance in glioblastoma. *Int. J. Oncol.* 48, 1053-1062.

Chu, C.Y., and Rana, T.M. (2006). Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. *PLoS Biol.* 4, e210.

Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K., and Lee, J.T. (2014). Regulatory interactions between RNA and polycomb repressive complex 2. *Mol. Cell* 55, 171-185.

Collart, M.A., and Struhl, K. (1994). NOT1(CDC39), NOT2(CDC36), NOT3, and NOT4 encode a global-negative regulator of transcription that differentially affects TATA-element utilization. *Genes Dev.* 8, 525-537.

Collier, J., and Parker, R. (2005). General translational repression by activators of mRNA decapping. *Cell* 122, 875-886.

Collier, J.M., Tucker, M., Sheth, U., Valencia-Sanchez, M.A., and Parker, R. (2001). The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes. *RNA* 7, 1717-1727.

Cougot, N., Babajko, S., and Seraphin, B. (2004). Cytoplasmic foci are sites of mRNA decay in human cells. *J. Cell Biol.* 165, 31-40.

Daugeron, M.C., Mauxion, F., and Seraphin, B. (2001). The yeast POP2 gene encodes a

nuclease involved in mRNA deadenylation. *Nucleic Acids Res.* 29, 2448-2455.

Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K.J., Gooding, A.R., Lee, J.T., and Cech, T.R. (2015). Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Mol. Cell* 57, 552-558.

Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* 20, 1250-1257.

de Valoir, T., Tucker, M.A., Belikoff, E.J., Camp, L.A., Bolduc, C., and Beckingham, K. (1991). A second maternally expressed *Drosophila* gene encodes a putative RNA helicase of the "DEAD box" family. *Proc. Natl. Acad. Sci. USA* 88, 2113-2117.

de Vries, S., Naarmann-de Vries, I.S., Urlaub, H., Lue, H., Bernhagen, J., Ostareck, D.H., and Ostareck-Lederer, A. (2013). Identification of DEAD-box RNA helicase 6 (DDX6) as a cellular modulator of vascular endothelial growth factor expression under hypoxia. *J. Biol. Chem.* 288, 5815-5827.

Decker, C.J., and Parker, R. (1993). A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes Dev.* 7, 1632-1643.

Desouki, M.M., Liao, S., Conroy, J., Nowak, N.J., Shepherd, L., Gaile, D.P., and Geradts, J. (2011). The genomic relationship between primary breast carcinomas and their nodal metastases. *Cancer Invest.* 29, 300-307.

Drummond, S.P., Hildyard, J., Firczuk, H., Reamtong, O., Li, N., Kannambath, S., Claydon, A.J., Beynon, R.J., Eyers, C.E., and McCarthy, J.E. (2011). Diauxic shift-dependent relocalization of decapping activators Dhh1 and Pat1 to polysomal complexes. *Nucleic Acids Res.* 39, 7764-7774.

Dutta, A., Zheng, S., Jain, D., Cameron, C.E., and Reese, J.C. (2011). Intermolecular interactions within the abundant DEAD-box protein Dhh1 regulate its activity in vivo. *J. Biol. Chem.* 286, 27454-27470.

Eichhorn, S.W., Guo, H., McGeary, S.E., Rodriguez-Mias, R.A., Shin, C., Baek, D., Hsu, S.H., Ghoshal, K., Villen, J., and Bartel, D.P. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues.

Mol. Cell 56, 104-115.

Ernault-Lange, M., Baconnais, S., Harper, M., Minshall, N., Souquere, S., Boudier, T., Benard, M., Andrey, P., Pierron, G., Kress, M., *et al.* (2012). Multiple binding of repressed mRNAs by the P-body protein Rck/p54. *RNA* 18, 1702-1715.

Eulalio, A., Behm-Ansmant, I., Schweizer, D., and Izaurralde, E. (2007a). P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol. Cell. Biol.* 27, 3970-3981.

Eulalio, A., Rehwinkel, J., Stricker, M., Huntzinger, E., Yang, S.F., Doerks, T., Dorner, S., Bork, P., Boutros, M., and Izaurralde, E. (2007b). Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. *Genes Dev.* 21, 2558-2570.

Eystathiou, T., Jakymiw, A., Chan, E.K., Seraphin, B., Cougot, N., and Fritzler, M.J. (2003). The GW182 protein colocalizes with mRNA degradation associated proteins hDcp1 and hLSm4 in cytoplasmic GW bodies. *RNA* 9, 1171-1173.

Fabian, M.R., Frank, F., Rouya, C., Siddiqui, N., Lai, W.S., Karetnikov, A., Blackshear, P.J., Nagar, B., and Sonenberg, N. (2013). Structural basis for the recruitment of the human CCR4-NOT deadenylase complex by tristetraprolin. *Nat. Struct. Mol. Biol.* 20, 735-739.

Fenger-Gron, M., Fillman, C., Norrild, B., and Lykke-Andersen, J. (2005). Multiple processing body factors and the ARE binding protein TTP activate mRNA decapping. *Mol. Cell* 20, 905-915.

Fischer, N., and Weis, K. (2002). The DEAD box protein Dhh1 stimulates the decapping enzyme Dcp1. *EMBO J.* 21, 2788-2797.

Gimenez-Barcons, M., Alves-Rodrigues, I., Jungfleisch, J., Van Wynsberghe, P.M., Ahlquist, P., and Diez, J. (2013). The cellular decapping activators LSM1, Pat1, and Dhh1 control the ratio of subgenomic to genomic Flock House virus RNAs. *J. Virol.* 87, 6192-6200.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466, 835-840.

Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding

RNAs. *Nature* 482, 339-346.

Haas, G., Braun, J.E., Igreja, C., Triteschler, F., Nishihara, T., and Izaurralde, E. (2010). HPat provides a link between deadenylation and decapping in metazoa. *J. Cell Biol.* 189, 289-302.

Haimovich, G., Medina, D.A., Causse, S.Z., Garber, M., Millan-Zambrano, G., Barkai, O., Chavez, S., Perez-Ortin, J.E., Darzacq, X., and Choder, M. (2013). Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell* 153, 1000-1011.

Han, K., Jaimovich, A., Dey, G., Ruggero, D., Meyuhas, O., Sonenberg, N., and Meyer, T. (2014). Parallel measurement of dynamic changes in translation rates in single cells. *Nat. Methods* 11, 86-93.

Hashimoto, K., Nakagawa, Y., Morikawa, H., Niki, M., Egashira, Y., Hirata, I., Katsu, K., and Akao, Y. (2001). Co-overexpression of DEAD box protein rck/p54 and c-myc protein in human colorectal adenomas and the relevance of their expression in cultured cell lines. *Carcinogenesis* 22, 1965-1970.

Hata, H., Mitsui, H., Liu, H., Bai, Y., Denis, C.L., Shimizu, Y., and Sakai, A. (1998). Dhh1p, a putative RNA helicase, associates with the general transcription factors Pop2p and Ccr4p from *Saccharomyces cerevisiae*. *Genetics* 148, 571-579.

Henn, A., Bradley, M.J., and De La Cruz, E.M. (2012). ATP utilization and RNA conformational rearrangement by DEAD-box proteins. *Annu Rev Biophys* 41, 247-267.

Hillebrand, J., Pan, K., Kokaram, A., Barbee, S., Parker, R., and Ramaswami, M. (2010). The Me31B DEAD-Box Helicase Localizes to Postsynaptic Foci and Regulates Expression of a CaMKII Reporter mRNA in Dendrites of *Drosophila* Olfactory Projection Neurons. *Front Neural Circuits* 4, 121.

Holetz, F.B., Correa, A., Avila, A.R., Nakamura, C.V., Krieger, M.A., and Goldenberg, S. (2007). Evidence of P-body-like structures in *Trypanosoma cruzi*. *Biochem. Biophys. Res. Commun.* 356, 1062-1067.

Hopkins, K.C., McLane, L.M., Maqbool, T., Panda, D., Gordesky-Gold, B., and Cherry, S. (2013). A genome-wide RNAi screen reveals that mRNA decapping restricts

bunyaviral replication by limiting the pools of Dcp2-accessible targets for cap-snatching. *Genes Dev.* 27, 1511-1525.

Huang, S. (2009). Reprogramming cell fates: reconciling rarity with robustness. *BioEssays* 31, 546-560.

Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D.E. (2005). Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* 94, 128701.

Huang, S., Guo, Y.P., May, G., and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* 305, 695-713.

Hubstenberger, A., Noble, S.L., Cameron, C., and Evans, T.C. (2013). Translation repressors, an RNA helicase, and developmental cues control RNP phase transitions during early development. *Dev. Cell* 27, 161-173.

Huys, A., Thibault, P.A., and Wilson, J.A. (2013). Modulation of hepatitis C virus RNA accumulation and translation by DDX6 and miR-122 are mediated by separate mechanisms. *PLoS ONE* 8, e67437.

Iio, A., Takagi, T., Miki, K., Naoe, T., Nakayama, A., and Akao, Y. (2013). DDX6 post-transcriptionally down-regulates miR-143/145 expression through host gene NCR143/145 in cancer cells. *Biochim. Biophys. Acta* 1829, 1102-1110.

Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11, 113-127.

Jangra, R.K., Yi, M., and Lemon, S.M. (2010). DDX6 (Rck/p54) is required for efficient hepatitis C virus replication but not for internal ribosome entry site-directed translation. *J. Virol.* 84, 6810-6824.

Jarmoskaite, I., and Russell, R. (2014). RNA helicase proteins as chaperones and remodelers. *Annu. Rev. Biochem.* 83, 697-725.

Jonas, S., and Izaurralde, E. (2013). The role of disordered protein regions in the assembly of decapping complexes and RNP granules. *Genes Dev.* 27, 2628-2641.

Kamenska, A., Lu, W.T., Kubacka, D., Broomhead, H., Minshall, N., Bushell, M., and Standart, N. (2014). Human 4E-T represses translation of bound mRNAs and enhances microRNA-mediated silencing. *Nucleic Acids Res.* 42, 3298-3313.

Kawahara, C., Yokota, S., and Fujita, H. (2014). DDX6 localizes to nuage structures and the annulus of mammalian spermatogenic cells. *Histochem. Cell Biol.* 141, 111-121.

Kirli, K., Karaca, S., Dehne, H.J., Samwer, M., Pan, K.T., Lenz, C., Urlaub, H., and Gorlich, D. (2015). A deep proteomics perspective on CRM1-mediated nuclear export and nucleocytoplasmic partitioning. *eLife* 4.

Kitamura, K., Nakagawa, T., and Shimoda, C. (1990). Novel Sterile Mutants of the Fission Yeast *Schizosaccharomyces-Pombe* Which Are Defective in Their Response to Starvation. *Curr. Genet.* 18, 315-321.

Koike, N., Yoo, S.H., Huang, H.C., Kumar, V., Lee, C., Kim, T.K., and Takahashi, J.S. (2012). Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science* 338, 349-354.

Kruk, J.A., Dutta, A., Fu, J., Gilmour, D.S., and Reese, J.C. (2011). The multifunctional Ccr4-Not complex directly promotes transcription elongation. *Genes Dev.* 25, 581-593.

Kuzuoglu-Ozturk, D., Bhandari, D., Huntzinger, E., Fauser, M., Helms, S., and Izaurralde, E. (2016). miRISC and the CCR4-NOT complex silence mRNA targets independently of 43S ribosomal scanning. *EMBO J.* 35, 1186-1203.

Kwok, C.K., Tang, Y., Assmann, S.M., and Bevilacqua, P.C. (2015). The RNA structurome: transcriptome-wide structure probing with next-generation sequencing. *Trends Biochem. Sci.* 40, 221-232.

Lackner, D.H., Carre, A., Guzzardo, P.M., Banning, C., Mangena, R., Henley, T., Oberndorfer, S., Gapp, B.V., Nijman, S.M., Brummelkamp, T.R., *et al.* (2015). A generic strategy for CRISPR-Cas9-mediated gene tagging. *Nat Commun* 6, 10237.

Ladomery, M., Wade, E., and Sommerville, J. (1997). Xp54, the *Xenopus* homologue of human RNA helicase p54, is an integral component of stored mRNP particles in oocytes. *Nucleic Acids Res.* 25, 965-973.

Leloup, J.-C., Gonze, D., and Goldbeter, A. (2006). Chapter 13 - Computational Models

for Circadian Rhythms: Deterministic Versus Stochastic Approaches. In *Computational Systems Biology* (Burlington: Academic Press), pp. 249-291.

Lin, F., Wang, R., Shen, J.J., Wang, X., Gao, P., Dong, K., and Zhang, H.Z. (2008). Knockdown of RCK/p54 expression by RNAi inhibits proliferation of human colorectal cancer cells in vitro and in vivo. *Cancer Biol. Ther.* 7, 1669-1676.

Linder, P., and Jankowsky, E. (2011). From unwinding to clamping - the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* 12, 505-516.

Liu, F., Putnam, A., and Jankowsky, E. (2008). ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc. Natl. Acad. Sci. USA* 105, 20209-20214.

Liu, H.Y., Badarinarayana, V., Audino, D.C., Rappsilber, J., Mann, M., and Denis, C.L. (1998). The NOT proteins are part of the CCR4 transcriptional complex and affect gene expression both positively and negatively. *EMBO J.* 17, 1096-1106.

Liu, L., Qi, H., Wang, J., and Lin, H. (2011). PAPI, a novel TUDOR-domain protein, complexes with AGO3, ME31B and TRAL in the nuage to silence transposition. *Development* 138, 1863-1873.

Lowrey, P.L., and Takahashi, J.S. (2004). Mammalian circadian biology: elucidating genome-wide levels of temporal organization. *Annu. Rev. Genom. Hum. Genet.* 5, 407-441.

Lu, D., and Yunis, J.J. (1992). Cloning, expression and localization of an RNA helicase gene from a human lymphoid cell line with chromosomal breakpoint 11q23.3. *Nucleic Acids Res.* 20, 1967-1972.

Maekawa, H., Nakagawa, T., Uno, Y., Kitamura, K., and Shimoda, C. (1994). The *ste13⁺* gene encoding a putative RNA helicase is essential for nitrogen starvation-induced G1 arrest and initiation of sexual development in the fission yeast *Schizosaccharomyces pombe*. *Mol. Gen. Genet.* 244, 456-464.

Maillet, L., and Collart, M.A. (2002). Interaction between Not1p, a component of the Ccr4-not complex, a global regulator of transcription, and Dhh1p, a putative RNA helicase. *J. Biol. Chem.* 277, 2835-2842.

Maillet, L., Tu, C., Hong, Y.K., Shuster, E.O., and Collart, M.A. (2000). The essential function of Not1 lies within the Ccr4-Not complex. *J. Mol. Biol.* 303, 131-143.

Mallam, A.L., Del Campo, M., Gilman, B., Sidote, D.J., and Lambowitz, A.M. (2012). Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p. *Nature* 490, 121-125.

Marnef, A., Weil, D., and Standart, N. (2012). RNA-related nuclear functions of human Pat1b, the P-body mRNA decay factor. *Mol. Biol. Cell* 23, 213-224.

Mathys, H., Basquin, J., Ozgur, S., Czarnocki-Cieciura, M., Bonneau, F., Aartse, A., Dziembowski, A., Nowotny, M., Conti, E., and Filipowicz, W. (2014). Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. *Mol. Cell* 54, 751-765.

Matsui, T., Hogetsu, K., Usukura, J., Sato, T., Kumasaka, T., Akao, Y., and Tanaka, N. (2006). Structural insight of human DEAD-box protein rck/p54 into its substrate recognition with conformational changes. *Genes Cells* 11, 439-452.

Matsumoto, K., Kwon, O.Y., Kim, H., and Akao, Y. (2005). Expression of rck/p54, a DEAD-box RNA helicase, in gametogenesis and early embryogenesis of mice. *Dev. Dynam.* 233, 1149-1156.

McMahon, A.C., Rahman, R., Jin, H., Shen, J.L., Fieldsend, A., Luo, W., and Rosbash, M. (2016). TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. *Cell* 165, 742-753.

Mercer, T.R., and Mattick, J.S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* 20, 300-307.

Minshall, N., Kress, M., Weil, D., and Standart, N. (2009). Role of p54 RNA helicase activity and its C-terminal domain in translational repression, P-body localization and assembly. *Mol. Biol. Cell* 20, 2464-2472.

Minshall, N., and Standart, N. (2004). The active form of Xp54 RNA helicase in translational repression is an RNA-mediated oligomer. *Nucleic Acids Res.* 32, 1325-1334.

Minshall, N., Thom, G., and Standart, N. (2001). A conserved role of a DEAD box

helicase in mRNA masking. *RNA* 7, 1728-1742.

Mitchell, S.F., Jain, S., She, M., and Parker, R. (2013). Global analysis of yeast mRNPs. *Nat. Struct. Mol. Biol.* 20, 127-133.

Miyaji, K., Nakagawa, Y., Matsumoto, K., Yoshida, H., Morikawa, H., Hongou, Y., Arisaka, Y., Kojima, H., Inoue, T., Hirata, I., *et al.* (2003). Overexpression of a DEAD box/RNA helicase protein, rck/p54, in human hepatocytes from patients with hepatitis C virus-related chronic hepatitis and its implication in hepatocellular carcinogenesis. *J. Viral Hepatitis* 10, 241-248.

Nakagawa, Y., Morikawa, H., Hirata, I., Shiozaki, M., Matsumoto, A., Maemura, K., Nishikawa, T., Niki, M., Tanigawa, N., Ikegami, M., *et al.* (1999). Overexpression of rck/p54, a DEAD box protein, in human colorectal tumours. *Br. J. Cancer* 80, 914-917.

Nakamura, A., Amikura, R., Hanyu, K., and Kobayashi, S. (2001). Me31B silences translation of oocyte-localizing RNAs through the formation of cytoplasmic RNP complex during *Drosophila* oogenesis. *Development* 128, 3233-3242.

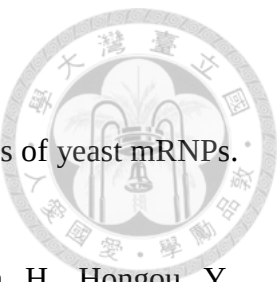
Nathans, R., Chu, C.Y., Serquina, A.K., Lu, C.C., Cao, H., and Rana, T.M. (2009). Cellular microRNA and P bodies modulate host-HIV-1 interactions. *Mol. Cell* 34, 696-709.

Navarro, R.E., Shim, E.Y., Kohara, Y., Singson, A., and Blackwell, T.K. (2001). cgh-1, a conserved predicted RNA helicase required for gametogenesis and protection from physiological germline apoptosis in *C. elegans*. *Development* 128, 3221-3232.

Nicklas, S., Okawa, S., Hillje, A.L., Gonzalez-Cano, L., Del Sol, A., and Schwamborn, J.C. (2015). The RNA helicase DDX6 regulates cell-fate specification in neural stem cells via miRNAs. *Nucleic Acids Res.* 43, 2638-2654.

Nishi, K., Nishi, A., Nagasawa, T., and Ui-Tei, K. (2013). Human TNRC6A is an Argonaute-navigator protein for microRNA-mediated gene silencing in the nucleus. *RNA* 19, 17-35.

Nishimura, T., Padamsi, Z., Fakim, H., Milette, S., Dunham, W.H., Gingras, A.C., and Fabian, M.R. (2015). The eIF4E-Binding Protein 4E-T Is a Component of the mRNA Decay Machinery that Bridges the 5' and 3' Termini of Target mRNAs. *Cell Rep.* 11,



1425-1436.

Nissan, T., Rajyaguru, P., She, M., Song, H., and Parker, R. (2010). Decapping activators in *Saccharomyces cerevisiae* act by multiple mechanisms. *Mol. Cell* 39, 773-783.

Noble, S.L., Allen, B.L., Goh, L.K., Nordick, K., and Evans, T.C. (2008). Maternal mRNAs are regulated by diverse P body-related mRNP granules during early *Caenorhabditis elegans* development. *J. Cell Biol.* 182, 559-572.

Oberholzer, U., and Collart, M.A. (1999). In vitro transcription of a TATA-less promoter: negative regulation by the Not1 protein. *Biol. Chem.* 380, 1365-1370.

Ohta, S., Bukowski-Wills, J.C., Sanchez-Pulido, L., Alves Fde, L., Wood, L., Chen, Z.A., Platani, M., Fischer, L., Hudson, D.F., Ponting, C.P., *et al.* (2010). The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. *Cell* 142, 810-821.

Ozgur, S., Basquin, J., Kamenska, A., Filipowicz, W., Standart, N., and Conti, E. (2015a). Structure of a Human 4E-T/DDX6/CNOT1 Complex Reveals the Different Interplay of DDX6-Binding Proteins with the CCR4-NOT Complex. *Cell Rep.* 13, 703-711.

Ozgur, S., Buchwald, G., Falk, S., Chakrabarti, S., Prabu, J.R., and Conti, E. (2015b). The conformational plasticity of eukaryotic RNA-dependent ATPases. *FEBS J.* 282, 850-863.

Ozgur, S., Chekulaeva, M., and Stoecklin, G. (2010). Human Pat1b connects deadenylation with mRNA decapping and controls the assembly of processing bodies. *Mol. Cell. Biol.* 30, 4308-4323.

Ozgur, S., and Stoecklin, G. (2013). Role of Rck-Pat1b binding in assembly of processing-bodies. *RNA Biol.* 10, 528-539.

Pan, C., Potratz, J.P., Cannon, B., Simpson, Z.B., Ziehr, J.L., Tijerina, P., and Russell, R. (2014). DEAD-box helicase proteins disrupt RNA tertiary structure through helix capture. *PLoS Biol.* 12, e1001981.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread Co-translational RNA

Decay Reveals Ribosome Dynamics. *Cell* 161, 1400-1412.

Pettersson, O.J., Aagaard, L., Andrejeva, D., Thomsen, R., Jensen, T.G., and Damgaard, C.K. (2014). DDX6 regulates sequestered nuclear CUG-expanded DMPK-mRNA in dystrophia myotonica type 1. *Nucleic Acids Res.* 42, 7186-7200.

Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., *et al.* (2015). Codon optimality is a major determinant of mRNA stability. *Cell* 160, 1111-1124.

Qi, M.Y., Wang, Z.Z., Zhang, Z., Shao, Q., Zeng, A., Li, X.Q., Li, W.Q., Wang, C., Tian, F.J., Li, Q., *et al.* (2012). AU-rich-element-dependent translation repression requires the cooperation of tristetraprolin and RCK/P54. *Mol. Cell. Biol.* 32, 913-928.

Raisch, T., Bhandari, D., Sabath, K., Helms, S., Valkov, E., Weichenrieder, O., and Izaurralde, E. (2016). Distinct modes of recruitment of the CCR4-NOT complex by *Drosophila* and vertebrate Nanos. *EMBO J.* 35, 974-990.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135, 216-226.

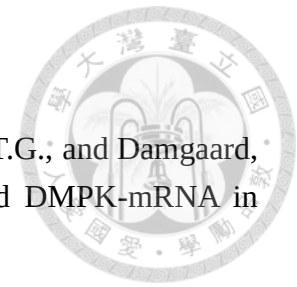
Rajyaguru, P., She, M., and Parker, R. (2012). Scd6 targets eIF4G to repress translation: RGG motif proteins as a class of eIF4G-binding proteins. *Mol. Cell* 45, 244-254.

Reed, J.C., Molter, B., Geary, C.D., McNevin, J., McElrath, J., Giri, S., Klein, K.C., and Lingappa, J.R. (2012). HIV-1 Gag co-opts a cellular complex containing DDX6, a helicase that facilitates capsid assembly. *J. Cell Biol.* 198, 439-456.

Reppert, S.M., and Weaver, D.R. (2001). Molecular analysis of mammalian circadian rhythms. *Annu. Rev. Physiol.* 63, 647-676.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81, 145-166.

Robinson, B.A., Reed, J.C., Geary, C.D., Swain, J.V., and Lingappa, J.R. (2014). A temporospatial map that defines specific steps at which critical surfaces in the Gag MA and CA domains act during immature HIV-1 capsid assembly in cells. *J. Virol.* 88, 5718-5741.



Robinson, G.E., Fernald, R.D., and Clayton, D.F. (2008). Genes and social behavior. *Science* 322, 896-900.

Robinson, G.E., Grozinger, C.M., and Whitfield, C.W. (2005). Sociogenomics: social life in molecular terms. *Nat. Rev. Genet.* 6, 257-270.

Rouhana, L., Shibata, N., Nishimura, O., and Agata, K. (2010). Different requirements for conserved post-transcriptional regulators in planarian regeneration and stem cell maintenance. *Dev. Biol.* 341, 429-443.

Rouya, C., Siddiqui, N., Morita, M., Duchaine, T.F., Fabian, M.R., and Sonenberg, N. (2014). Human DDX6 effects miRNA-mediated gene silencing via direct binding to CNOT1. *RNA* 20, 1398-1409.

Saito, K., Kondo, E., and Matsushita, M. (2011). MicroRNA 130 family regulates the hypoxia response signal through the P-body protein DDX6. *Nucleic Acids Res.* 39, 6086-6099.

Sauliere, J., Murigneux, V., Wang, Z., Marquenet, E., Barbosa, I., Le Tonqueze, O., Audic, Y., Paillard, L., Roest Crolius, H., and Le Hir, H. (2012). CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *Nat. Struct. Mol. Biol.* 19, 1124-1131.

Scheller, N., Mina, L.B., Galao, R.P., Chari, A., Gimenez-Barcons, M., Noueiry, A., Fischer, U., Meyerhans, A., and Diez, J. (2009). Translation and replication of hepatitis C virus genomic RNA depends on ancient cellular proteins that control mRNA fates. *Proc. Natl. Acad. Sci. USA* 106, 13517-13522.

Scheper, T., Klinkenberg, D., Pennartz, C., and van Pelt, J. (1999). A mathematical model for the intracellular circadian rhythm generator. *J. Neurosci.* 19, 40-47.

Schmiedel, J.M., Klemm, S.L., Zheng, Y., Sahay, A., Bluthgen, N., Marks, D.S., and van Oudenaarden, A. (2015). Gene expression. MicroRNA control of protein expression noise. *Science* 348, 128-132.

Schutz, P., Bumann, M., Oberholzer, A.E., Bieniossek, C., Trachsel, H., Altmann, M., and Baumann, U. (2008). Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proc. Natl. Acad. Sci. USA*

105, 9564-9569.

Sharif, H., Ozgur, S., Sharma, K., Basquin, C., Urlaub, H., and Conti, E. (2013). Structural analysis of the yeast Dhh1-Pat1 complex reveals how Dhh1 engages Pat1, Edc3 and RNA in mutually exclusive interactions. *Nucleic Acids Res.* *41*, 8377-8390.

Sheth, U., and Parker, R. (2003). Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* *300*, 805-808.

Shoemaker, C.J., and Green, R. (2012). Translation drives mRNA quality control. *Nat. Struct. Mol. Biol.* *19*, 594-601.

Smillie, D.A., and Sommerville, J. (2002). RNA helicase p54 (DDX6) is a shuttling protein involved in nuclear assembly of stored mRNP particles. *J. Cell Sci.* *115*, 395-407.

Stampfl, S., Doetsch, M., Beich-Frandsen, M., and Schroeder, R. (2013). Characterization of the kinetics of RNA annealing and strand displacement activities of the *E. coli* DEAD-box helicase CsdA. *RNA Biol.* *10*, 149-156.

Storch, K.F., Lipan, O., Leykin, I., Viswanathan, N., Davis, F.C., Wong, W.H., and Weitz, C.J. (2002). Extensive and divergent circadian gene expression in liver and heart. *Nature* *417*, 78-83.

Strahl-Bolsinger, S., and Tanner, W. (1993). A yeast gene encoding a putative RNA helicase of the "DEAD"-box family. *Yeast* *9*, 429-432.

Sweet, T., Kovalak, C., and Collier, J. (2012). The DEAD-box protein Dhh1 promotes decapping by slowing ribosome movement. *PLoS Biol.* *10*, e1001342.

Tanaka, K.J., Ogawa, K., Takagi, M., Imamoto, N., Matsumoto, K., and Tsujimoto, M. (2006). RAP55, a cytoplasmic mRNP component, represses translation in *Xenopus* oocytes. *J. Biol. Chem.* *281*, 40096-40106.

Taniguchi, K., Sugito, N., Kumazaki, M., Shinohara, H., Yamada, N., Matsushashi, N., Futamura, M., Ito, Y., Otsuki, Y., Yoshida, K., *et al.* (2015). Positive feedback of DDX6/c-Myc/PTB1 regulated by miR-124 contributes to maintenance of the Warburg effect in colon cancer cells. *Biochim. Biophys. Acta* *1852*, 1971-1980.

Tay, Y., Rinn, J., and Pandolfi, P.P. (2014). The multilayered complexity of ceRNA crosstalk and competition. *Nature* 505, 344-352.

Teixeira, D., and Parker, R. (2007). Analysis of P-body assembly in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 18, 2274-2287.

Teixeira, D., Sheth, U., Valencia-Sanchez, M.A., Brengues, M., and Parker, R. (2005). Processing bodies require RNA for assembly and contain nontranslating mRNAs. *RNA* 11, 371-382.

Tkacz, I.D., Cohen, S., Salmon-Divon, M., and Michaeli, S. (2008). Identification of the heptameric Lsm complex that binds U6 snRNA in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 160, 22-31.

Tomazella, G.G., Kassahun, H., Nilsen, H., and Thiede, B. (2012). Quantitative proteome analysis reveals RNA processing factors as modulators of ionizing radiation-induced apoptosis in the *C. elegans* germline. *J. Proteome Res.* 11, 4277-4288.

Torrente, M.P., Zee, B.M., Young, N.L., Baliban, R.C., LeRoy, G., Floudas, C.A., Hake, S.B., and Garcia, B.A. (2011). Proteomic interrogation of human chromatin. *PLoS ONE* 6, e24747.

Tran, E.J., Zhou, Y., Corbett, A.H., and Wente, S.R. (2007). The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA:protein remodeling events. *Mol. Cell* 28, 850-859.

Tritschler, F., Braun, J.E., Eulalio, A., Truffault, V., Izaurralde, E., and Weichenrieder, O. (2009). Structural basis for the mutually exclusive anchoring of P body components EDC3 and Tral to the DEAD box protein DDX6/Me31B. *Mol. Cell* 33, 661-668.

Tritschler, F., Eulalio, A., Helms, S., Schmidt, S., Coles, M., Weichenrieder, O., Izaurralde, E., and Truffault, V. (2008). Similar modes of interaction enable Trailer Hitch and EDC3 to associate with DCP1 and Me31B in distinct protein complexes. *Mol. Cell. Biol.* 28, 6695-6708.

Tucker, M., Staples, R.R., Valencia-Sanchez, M.A., Muhlrads, D., and Parker, R. (2002). Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in *Saccharomyces cerevisiae*. *EMBO J.* 21, 1427-1436.

Tucker, M., Valencia-Sanchez, M.A., Staples, R.R., Chen, J., Denis, C.L., and Parker, R. (2001). The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in *Saccharomyces cerevisiae*. *Cell* 104, 377-386.

Uhlmann-Schiffler, H., Jalal, C., and Stahl, H. (2006). Ddx42p--a human DEAD box protein with RNA chaperone activities. *Nucleic Acids Res.* 34, 10-22.

Waghray, S., Williams, C., Coon, J.J., and Wickens, M. (2015). *Xenopus* CAF1 requires NOT1-mediated interaction with 4E-T to repress translation in vivo. *RNA* 21, 1335-1345.

Walsh, D., and Mohr, I. (2011). Viral subversion of the host protein synthesis machinery. *Nat. Rev. Microbiol.* 9, 860-875.

Wang, C.Y., Chen, W.L., and Wang, S.W. (2013). Pdc1 functions in the assembly of P bodies in *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* 33, 1244-1253.

Wang, Y., Arribas-Layton, M., Chen, Y., Lykke-Andersen, J., and Sen, G.L. (2015). DDX6 Orchestrates Mammalian Progenitor Function through the mRNA Degradation and Translation Pathways. *Mol. Cell* 60, 118-130.

Ward, A.M., Bidet, K., Yinglin, A., Ler, S.G., Hogue, K., Blackstock, W., Gunaratne, J., and Garcia-Blanco, M.A. (2011). Quantitative mass spectrometry of DENV-2 RNA-interacting proteins reveals that the DEAD-box RNA helicase DDX6 binds the DB1 and DB2 3' UTR structures. *RNA Biol.* 8, 1173-1186.

Weaver, I.C., Cervoni, N., Champagne, F.A., D'Alessio, A.C., Sharma, S., Seckl, J.R., Dymov, S., Szyf, M., and Meaney, M.J. (2004). Epigenetic programming by maternal behavior. *Nat. Neurosci.* 7, 847-854.

Westmoreland, T.J., Olson, J.A., Saito, W.Y., Huper, G., Marks, J.R., and Bennett, C.B. (2003). DHH1 regulates the G1/S-checkpoint following DNA damage or BRCA1 expression in yeast. *J. Surg. Res.* 113, 62-73.

Weston, A., and Sommerville, J. (2006). Xp54 and related (DDX6-like) RNA helicases: roles in messenger RNP assembly, translation regulation and RNA degradation. *Nucleic Acids Res.* 34, 3082-3094.

Whitfield, C.W., Cziko, A.M., and Robinson, G.E. (2003). Gene expression profiles in the brain predict behavior in individual honey bees. *Science* 302, 296-299.

Xu, L., and Massague, J. (2004). Nucleocytoplasmic shuttling of signal transducers. *Nat. Rev. Mol. Cell Biol.* 5, 209-219.

Yoon, J.H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., Huarte, M., Zhan, M., Becker, K.G., and Gorospe, M. (2012). LincRNA-p21 suppresses target mRNA translation. *Mol. Cell* 47, 648-655.

Yoshida-Kashikawa, M., Shibata, N., Takechi, K., and Agata, K. (2007). DjCBC-1, a conserved DEAD box RNA helicase of the RCK/p54/Me31B family, is a component of RNA-protein complexes in planarian stem cells and neurons. *Dev. Dynam.* 236, 3436-3450.

Young, C.L., Khoshnevis, S., and Karbstein, K. (2013). Cofactor-dependent specificity of a DEAD-box protein. *Proc. Natl. Acad. Sci. USA* 110, E2668-2676.

Yu, S.F., Lujan, P., Jackson, D.L., Emerman, M., and Linial, M.L. (2011). The DEAD-box RNA helicase DDX6 is required for efficient encapsidation of a retroviral genome. *PLoS Pathog.* 7, e1002303.

Zhang, S., Williams, C.J., Wormington, M., Stevens, A., and Peltz, S.W. (1999). Monitoring mRNA decapping activity. *Methods* 17, 46-51.

Zhang, Z., Ye, W., Qian, Y., Zheng, Z., Huang, X., and Hu, G. (2012). Chaotic motifs in gene regulatory networks. *PLoS ONE* 7, e39355.

Zhao, X., and Jain, C. (2011). DEAD-box proteins from *Escherichia coli* exhibit multiple ATP-independent activities. *J. Bacteriol.* 193, 2236-2241.





Chapter 2

On the Nucleocytoplasmic Localization
of DEAD-box Protein 6

A New Synthesis Superseding the Old Hypothesis



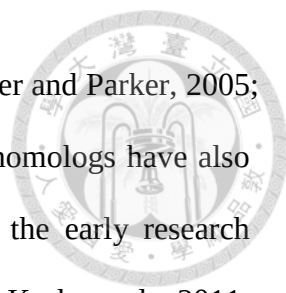


1. **Abstract**

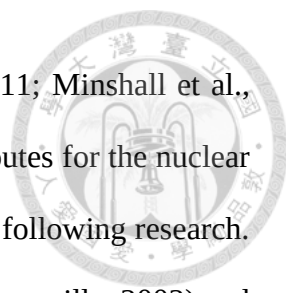
The nuclear localization and the potential nuclear functions of the DDX6 homologs have long been overshadowed by their cytoplasmic localization and functions. To gain a comprehensive understanding of the nuclear properties of the DDX6 homologs, we verified the nuclear localization of the DDX6 in human cell models. Furthermore, we assessed the validity of the shuttling model, the current model for the nucleocytoplasmic distribution of the DDX6 homologs. Our results refuted this model. We showed the proposed nuclear localization signal is not required for the nuclear entry. In conjunction, we demonstrated the proposed nuclear export signal was structurally inaccessible to the exportation machinery. Instead, we showed that the nucleocytoplasmic localization of DDX6 is governed by two parallel mechanisms, for which we termed the “hitchhiking” mechanism and the “piggyback” mechanism. The hitchhiking mechanism involves DDX6 binding the condensed chromosomes during mitosis via its C-terminal domain and entering the newly formed nuclei as a passenger protein on the mitotic chromosome. The piggyback mechanism involves the indirect shuttling of DDX6 by the importin- α/β and the CRM1 pathways through the mediator protein 4E-T. The two mechanisms can dominate under different biological context and can potentially complement or buffer the effect of one another.

2. **Introduction**

The DDX6 subfamily of the DEAD-box proteins (DBPs) is conserved across the eukaryotes. The members of the DDX6 subfamily are best known for their cytoplasmic functions, namely facilitating RNA turnover, translational repression, and processing



body (PB) assembly (Ayache et al., 2015; Chu and Rana, 2006; Collier and Parker, 2005; Collier et al., 2001; Minshall et al., 2001). But, in fact, the DDX6 homologs have also been observed in the cell nuclei in various model systems since the early research (Haimovich et al., 2013; Iio et al., 2013; Kawahara et al., 2014; Kruk et al., 2011; Lodomery et al., 1997; Minshall et al., 2001; Pettersson et al., 2014; Smillie and Sommerville, 2002; Torrente et al., 2011; Yoshida-Kashikawa et al., 2007). The descriptions for the nuclear localization and the potential functions for the DDX6 homologs were not obsolete, but indeed were overshadowed by the predominant cytoplasmic localization and functions. Notwithstanding, several potential nuclear functions have been proposed for the nuclear DDX6 homologs, including the masking of the nascent transcripts (Smillie and Sommerville, 2002), the structural remodeling of the nuclear-retained pathological RNA (Pettersson et al., 2014), and the modulation of the transcription machinery (Haimovich et al., 2013; Kruk et al., 2011). However, it is unclear if the nuclear localization of DDX6 is a common phenomenon in the human cells. Moreover, the comprehensive understanding of the nuclear functions of the DDX6 in parts depends on the understanding of the underlying mechanism governing its nucleocytoplasmic distribution. Currently, this is described by the shuttling model which argues that the vertebrate DDX6 homologs contain a pair of the lysine/arginine-rich nuclear localization signal (K/R-rich NLS) and the leucine-rich nuclear export signal (L-rich NES) and thus are shuttled by the importin- α/β and the CRM1-dependent pathways, respectively (Smillie and Sommerville, 2002; Weston and Sommerville, 2006). However, this model is undermined by several inconsistencies. First, the use of the putative K/R-rich NLS and the importin- α/β pathway has yet to receive experimental support. Second, the non-vertebrate DDX6 homologs were also



observed in the cell nuclei (Haimovich et al., 2013; Kruk et al., 2011; Minshall et al., 2001; Yoshida-Kashikawa et al., 2007), suggesting there are other routes for the nuclear entry. Third, the use of the putative L-rich NES was disputed by the following research. More specifically, both supporting (Iio et al., 2013; Smillie and Sommerville, 2002) and opposing (Kirli et al., 2015; Marnef et al., 2012; Nishi et al., 2013; Smillie and Sommerville, 2002) evidence are present in the current literature. To resolve these discrepancies, we sought to verify the nuclear localization of DDX6 in the human cell models and scrutinized the shuttling model with a combinatorial approach. Furthermore, we looked beyond the framework of the shuttling model to investigate for other potential mechanisms that govern the nucleocytoplasmic distribution of DDX6.

3. Materials and methods

3.1. Cell culture and transfection

The HeLa cells and the HEK 293T cells were cultured in DMEM (GE healthcare) supplied with 10% (v/v) fetal bovine serum (HyClone, GE healthcare; Gibco, Thermo Fisher Scientific), 1% penicillin-streptomycin mixture (Gibco, Thermo Fisher Scientific) and with constant atmospheric 5% CO₂ under 37°C. The Lipofectamine 2000 reagent (Thermo Fisher Scientific) and the Lipofectamine RNAiMAX reagent (Thermo Fisher Scientific) were used for transfection experiments. For DDX6 and AGO2 knockdown (KD), cells were transfected with siRNA at 25 nM final concentrations with the 2000 reagent and the RNAiMAX reagent respectively. For MALAT1 KD, cells were transfected with siRNA at 50 nM final concentration using the RNAiMAX reagent. For plasmid transfection, the 2000 reagent was used. The transfection procedure followed



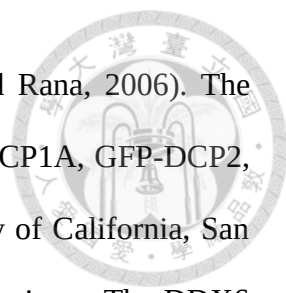
the supplier's protocol except for reducing the dosage to one fourth to reduce cytotoxicity. For the 2000 reagent, the transfection media were replaced with fresh, antibiotics-free media 6 hours after transfection. For the RNAiMAX reagent, the transfected cells were kept in the transfection media until harvest. For expression level analyses, the cells were harvested at 48 hr post-transfection. For image analysis, the cells were typically processed at 24 hr post-transfection. For the synchronized cells, the cells were processed at 46 hr post-transfection, accounting for the time needed for synchronization.

3.2. Drug treatments for CRM1 inhibition and cell cycle synchronization

For the inhibition of the CRM1-dependent export, we used leptomycin B (LMB; Sigma-Aldrich). The LMB stock solution was prepared at 5 mg/mL concentration in 70% (v/v) methanol. For the LMB treatment, the stock solution was directly diluted in the condition media to the desired concentration. For each assay, the same volume of the solvent to the highest dose LMB was used as control treatment. We treated the cell with LMB for 5 hr prior to downstream processes. For the cell synchronization at prometaphase, we used nocodazole (Sigma-Aldrich). The nocodazole stock solution was prepared at 10 mg/mL concentration in 100% dimethyl sulfoxide (DMSO). The stock solution was directly diluted in the condition media to 50 ng/mL concentration. The cells were treated for up to 16 hr to synchronize cell cycle. Under this condition, approximately 50% of the cultured cells will be synchronized.

3.3. Plasmid sources, molecular cloning, and fusion protein construction

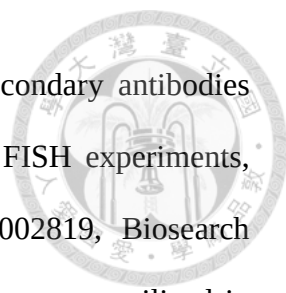
The cloning of the human DDX6, 4E-T, AGO2 into the pEYFP-C1 expression



vector (Clontech Laboratories) was described previously (Chu and Rana, 2006). The plasmids, pEYFP-C1, YFP-DDX6, YFP-4E-T, YFP-AGO2, GFP-DCP1A, GFP-DCP2, and pECFP-C1 were kind gifts from Prof. Tariq Rana in University of California, San Diego. GFP-CNOT1 was purchased from the Addgene plasmid repository. The DDX6 mutants used in this cloned by PCR amplification of DNA fragments using specific primers and then inserted back to the pEYFP-C1 vector. YFP-DDX6 NTD and CTD were constructed by Y. C. Chen (Y. C. Chen's master thesis). All other mutants were constructed by J. H. Huang. PATL1, EDC3, LSM14A, PKM, and the DBR1 CTD were amplified from the cDNA prepared from the HeLa cells with specific primers, and were then inserted into pEYFP-C1. The chimeric proteins were constructed by inserting the DNA fragments of the C-terminal peptide to the 3'-end of the DNA encoding the N-terminal peptide. CFP-DDX6 was constructed by inserting the PCR amplified DDX6 DNA fragment into the pECFP-C1 vector. Similarly, myc-DDX6 FL and CTD was constructed by inserting the PCR amplified DDX6 DNA fragment into the pKR43 vector with one copy of the N-terminal myc tag. The empty pKR43 vector was prepared from the myc-CEPB-2×MS2, a kind gift from Dr. Yi-Shuian Huang in Academia Sinica.

3.4. Immunofluorescence (IF) and fluorescent in situ hybridization (FISH)

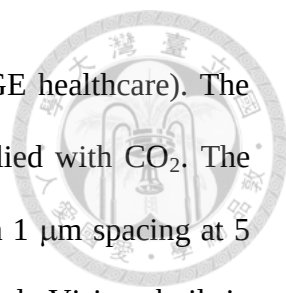
For the IF experiments, the cells were fixed with 4% (v/v) paraformaldehyde (PFA) in PBS at room temperature for 10 min before penetration with 0.25% (v/v) Triton X-100 in PBS at room temperature for 5 min. The cells were subsequently blocked with 2% (w/v) BSA/0.1% (v/v) Triton X-100 in PBS and were incubated with primary antibody recognizing DDX6 (A300-460A, Bethyl Laboratories) or with anti-myc tag



(clone 4A6, EMD Millipore) and Alexa Fluor 488 conjugated secondary antibodies (A-21202 and A-21206, respectively, Life Technologies). For the FISH experiments, the custom Stellaris FISH Probes recognizing MALAT1 (NR_002819, Biosearch Technologies) were directly labeled with Quasar 570. Probe set sequences utilized in the experiments have been previously described. The HeLa cells were hybridized with the MALAT1 Stellaris FISH Probe set, following the manufacturer's instructions. For anti-DDX6 and anti-phosphorylated SC35 IF (ab11826, Abcam) after MALAT1 FISH, the cells were washed with 50% FISH wash buffer in PBS, PBS and 0.1% Triton X-100 in PBS sequentially before proceeding on with IF procedures from blocking with 2% BSA/0.1% Triton X-100 in PBS.

3.5. Microscopy, image processing, and analysis

For the general purpose imaging, we routinely used the DMI3000 B wide field fluorescent microscope (Leica). More specifically, the DMI3000 B microscope accounted for all the images herein except for the confocal and the time-lapse imaging. Images acquired with the DMI3000 B microscope were pseudo-colored. Statistics of phenotypic counting were done with ImageJ cell counter plugin. For each construct, at least 3 biological replicates were counted. To capture high resolution images of the subnuclear structures and the condensed chromosomes, the samples were imaged with the LSM780 confocal microscope (Zeiss). Images were pseudo-colored. For Z-stacking, stacks of images were taken with spacing 0.63 μm or 1 μm along the Z axis, and were maximally projected on the XY-plane (Z projection) or projected 3-dimensionally (3D projection). The Z projection was done using ImageJ Z projection function. The 3D projection was done with the mesh 3D function in ICY image analysis platform. For



time-lapse imaging, we used DeltaVision Core imaging systems (GE healthcare). The cells were kept warm (34°C ~ 37°C) in the heating chamber supplied with CO₂. The cells were imaged across a 10 ~ 20 µm range along the Z axis with 1 µm spacing at 5 minute time interval for at least 3 hours. We utilize the DeltaVision built-in deconvolution function to improve the image resolution post-acquisition. We applied enhanced ratio (aggressive) mode of blind deconvolution with 10 iterations on each image in the original Z stacks. The channels and time points were included as extract dimensions in the original Z stacks using ImageJ. The resulting image hyperstacks were Z projected (sum slice projection) for each time point.

3.6. Whole cell, nuclear, and cytoplasmic protein extraction

The whole cell lysate, or the total cell lysate (TCE), were prepared from lysing the cells in the RIPA buffer (0.1% [w/v] SDS, 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% Nonidet P-40, 1% Triton X-100). The cytoplasmic and the nuclear fractions were prepared from the HeLa and the HEK 293T cells using either the NE-PER nuclear and cytoplasmic extraction reagents (Thermo Fisher Scientific) or the Gagnon-Li method. For the NE-PER fractionation, we generally followed manufacturer's protocol, except for gently washing the nuclei pellet with PBS before resuspending the nuclei with the NER reagent. For nuclear lysis, we had two rounds of additional freeze-thawing with liquid nitrogen after four rounds of the vortexing described in the supplier's protocol. For the Gagnon-Li fractionation, we followed the procedure for preparing cytoplasmic extract and total nuclear extract. In brief, the cytoplasmic fraction was released by lysing cell in the hypotonic lysis buffer (HLB; 10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl₂, 0.3% [v/v] Nonidet P-40, and 10% [v/v] glycerol). The resulting nuclei

pellet was washed with the HLB before resuspending the nuclei in the nuclear lysis buffer (NLB; 20 mM Tris-HCl pH 7.5, 150 mM KCl, 3 mM MgCl₂, 0.3% [v/v] Nonidet P-40, and 10% [v/v] glycerol). The nuclei were then lysed by sonication using the Digital Sonifier 450 system (Branson) at 40% power for a total of 45 s intercepted by 1 min cooling for every 15 s. The crude nuclear extract was then centrifuged at 1,200 ×g for 5 min. The resulting supernatant was collected as the total nuclear extract.

3.7. RNA co-immunoprecipitation (RIP)

RIP experiments were performed as described previously (Chu and Rana, 2006). Briefly, the TCE was prepared by lysing the cells in the IP lysis buffer (0.2 mM EDTA, 20 mM HEPES pH 7.9, 0.35% Triton X-100, 10 mM NaCl, 1 mM MgCl₂) for 10 min on ice and then centrifuged at 12,000 ×g for 10 min to remove the cell debris. For the anti-DDX6 or the control IgG IP, the supernatants were subsequently incubated with Dynabeads M-280 Sheep anti-rabbit IgG (Thermo Fisher Scientific) and anti-DDX6 (A300-460A, Bethyl Laboratories) or normal rabbit IgG (Santa Cruz) at 4°C for 3 hr, respectively. After incubation, the beads were washed with IP lysis buffer and IP wash buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.3% [v/v] Nonidet P-40). The beads were then split into 2 fractions; one for protein elution and the other for RNA extraction. For protein elution, beads were boiled at 95°C with the 4× SDS sample buffer (200 mM Tris-HCl pH 6.8, 400 mM dithiothreitol [DTT], 8% [w/v] sodium dodecyl sulfate [SDS], 0.4% [w/v] bromophenol blue, 40% [v/v] glycerol). As for RNA extraction, beads were resuspended in TRIzol reagent (Thermo Fisher Scientific).

3.8. Western blot (WB)

The whole cell lysate, the subcellular fractions, and the IP eluents were first separated by SDS-PAGE and transferred to PVDF membranes (Bio-Rad). To probe the specific proteins, we used the primary antibodies recognizing the GFP tag (Living Colors, Clontech), myc tag (clone 4A6, EMD Millipore), DDX6 (A300-460A, Bethyl Laboratories), LMNA (ab108595, Abcam), GAPDH (GTX100118, GeneTex), CALR (GTX111627, GeneTex; or, 06-661, EMD Millipore), AGO2 (H00027161-M01, Abnova), and the goat anti-rabbit or the goat anti-mouse IgG-HRP secondary antibodies (SC-2004 or SC-2005, Santa Cruz). For the WB following the IP experiments, the Rabbit Trueblot anti-rabbit IgG-HRP (18-8816-33, Rockland) secondary antibody was used to avoid detecting denatured IgG heavy chain near the 50 KDa position.

3.9. RNA extraction, reverse transcription and quantitative real-time PCR (qPCR)

The total RNA from the HeLa cells and the co-IPed RNA from the anti-DDX6 IP experiments were extracted using the TRIzol reagent. For the RNA extraction from RIP samples, glycogen (molecular biology grade, Roche) was used for promoting RNA precipitation according to the supplier's recommendation. The RNA samples were reverse-transcribed into complementary DNA (cDNA) using the SuperScript III first-strand synthesis system (Thermo Fisher Scientific) with random hexamer priming according to the suppliers' protocols. The resulting cDNA were further analyzed by qPCR using the iQ SYBR Green Supermix (Bio-Rad) and the gene-specific primer sets. Fold changes in gene expression level were calculated by the $\Delta\Delta C(t)$ method, with the

GAPDH mRNA serving as the reference for normalization. Fold enrichments in RIP-qPCR were calculated by $2^{[C_{\text{IgG}}(t) - C_{\text{anti-DDX6}}(t)]}$, the $\Delta C(t)$ method.



3.10. Quantitative mass spectrometry (MS) and data analysis

The details about the sample preparation, data acquisition, and raw data processing were described in the next chapter. In this chapter, we re-analyzed the MS data obtained by Y. C. Chen (replicate 1 and 2). To better estimate the DDX6 abundance, we only used the control siRNA (siDDX6-mm)-treated samples. The summed-up extracted ion chromatogram (XIC) intensities were used to estimate the protein abundances. To compare the relative protein abundance in the cytoplasmic and the nuclear fractions, we first median-centered the log₁₀-transformed XIC intensity distributions (Fig. 3). The resulting data were then back transformed to the natural scale where the XIC intensity ratios of the nuclear and the cytoplasmic fractions for each protein were taken.

3.11. Protein structures and NLS prediction

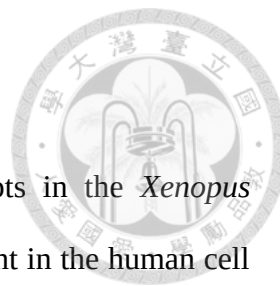
The published crystal structures of DDX6, including the apo state DDX6 (4CT5), DDX6 co-crystalized with the CNOT1 MIF4G domain (4CT4), and the DDX6 NTD (1VEC), are deposited in the Protein Data Bank (PDB) of the Research Collaboratory of Structural Bioinformatics (RCSB). The three crystal structures were downloaded from PDB as .pdb files and were visualized with Protein Workshop, an RCSB/PDB protein structure visualization software. The prediction of the K/R-rich NLS was conducted using the NucPred program implemented as a web-based tool (Brameier et al., 2007).

4. **Results**

4.1. **DDX6 is present in the human cell nuclei**



To confirm the nuclear presence of DDX6 in human cell nuclei, we performed subcellular fractionation followed by WB on the human cervical cancer HeLa cell line and the transformed human embryonic kidney HEK 293T cell line. We first conducted nuclear/cytoplasmic fractionation using the NE-PER reagents and the Gagnon-Li method (Gagnon et al., 2014a; Gagnon et al., 2014b), and subsequently detected DDX6 in both the cytoplasmic and the nuclear extracts prepared from the two methods (Fig. 1a). Simultaneously, we specifically detected the nuclear marker LMNA in the nuclear extracts while the cytosolic marker GAPDH and ER membrane marker CALR were mostly restricted to the cytoplasmic extracts (Fig. 1a). These results confirmed the minimal cross-contamination between cytoplasmic and nuclear fractions. Similarly, exogenous myc-tagged DDX6 was also detected in the HeLa nuclear extract (Fig. 2). To the same end, we exploited an in-house HeLa nuclear/cytoplasmic proteomics dataset produced by the quantitative MS (Fig. 1b and 3). In the two biological replicates of this dataset, the detection and the quantification of DDX6 were supported by at least two unique peptides in both the nuclear and the cytoplasmic fractions (Fig. 1b). Similar to the WB results, several known marker proteins for the nucleus, the cytoplasm, and the ER were differentially detected in the cytoplasmic and the nuclear fractions (Fig. 1b). Likewise, these results confirmed minimal cross-contamination of the two fractions. Independently, we detected the nuclear localization of the endogenous DDX6 with IF followed by confocal microscope imaging in the HeLa cells (Fig. 1c and 4). Conjointly, these results confirmed the presence of DDX6 in the human cell nuclei.



4.2. DDX6 binds the nuclear lncRNA MALAT1

Previously, Xp54 has been shown to bind nascent transcripts in the *Xenopus* oocytes (Smillie and Sommerville, 2002). Since DDX6 is also present in the human cell nuclei, we wondered if it can bind to nuclear RNA. We conducted anti-DDX6 IP followed by qPCR to test if DDX6 binds the nuclear long noncoding RNA (lncRNA) MALAT1 and NEAT1. As a result, MALAT1 but not NEAT1 was co-IPed with DDX6 (Fig. 5a). To circumvent the potential pitfall that DDX6 might have bound the dispersed MALAT1 in the mitotic cells or that the binding could have occur after cell lysis, we tracked the native localization of DDX6 and MALAT1 in the HeLa cells using IF and FISH double staining. As a result, we frequently observed co-localization or docking between DDX6 and MALAT1 in the nuclei (Fig. 5b). This finding supports that DDX6 can bind MALAT1 in vivo. The same finding also supports that the nuclear DDX6 is capable of RNA binding. We then tested if DDX6 regulates the expression level and the subnuclear localization of MALAT1. However, perturbing DDX6 level did not significantly alter the MALAT1 level (Fig. 6) or its subnuclear localization (Fig. 7).

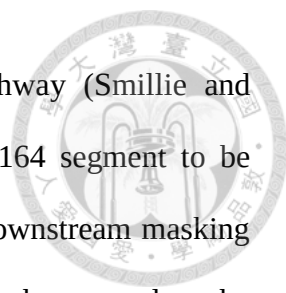
4.3. DDX6 is insensitive to the LMB treatment

Since DDX6 is present in the human cell nuclei, we further questioned the mechanism governing its nucleocytoplasmic localization. It was previously proposed that the vertebrate DDX6 homologs are shuttled by the importin- α/β and the CRM1 pathways (Smillie and Sommerville, 2002; Weston and Sommerville, 2006). However, this shuttling model was challenged by the conflicting data concerning the CRM1-dependent export (Kirli et al., 2015; Marnef et al., 2012; Nishi et al., 2013;

Smillie and Sommerville, 2002). In our survey of the shuttling PB components, only PATL1 and 4E-T, but not DDX6, were sensitive to the treatment of LMB (Fig. 8a), a potent and irreversible CRM1 inhibitor. Even under high dose of LMB, DDX6 did not accumulate in the cell nuclei (Fig. 8b). These results aligned with the opposing evidence to the shuttling model (Kirli et al., 2015; Marnef et al., 2012; Nishi et al., 2013; Smillie and Sommerville, 2002).

4.4. The putative NES of DDX6 is structurally inaccessible to CRM1

To further resolve the conflict, we resorted to the structural analysis for the biochemical insight into the feasibility of the CRM1-dependent export of DDX6. The DDX6 NTD adopts almost identical globular folding in all the three published X-ray crystallographic structures (Fig. 9b and 10) (Mathys et al., 2014; Matsui et al., 2006). The putative NES indeed folds into an L-rich NES-like structure that is characterized by α -helical folding with leucine residues protruding towards the same direction (Fig. 9b and 10). However, this putative NES is buried inside the core region of DDX6 NTD, with L151 and L155 directly blocked by the 174-199 segment (Fig. 9b, inset 1 and 2). In addition, L158 and L160 protrude downward and inward to the core of DDX6 NTD (Fig. 9b, inset 3). Furthermore, the putative NES is one of the most stably folded segments within the DDX6 NTD (Fig. 11a and b). Together, these results indicate that the putative NES is unlikely to contact other proteins at the molecular surfaces. Conversely, the crystallographic assessment of CRM1 revealed that the NES binding is a thin groove and can only accommodate NES at the molecular surface (Dong et al., 2009; Guttler et al., 2010). However, these findings conflict with that the N-terminal 1-164 segment of Xp54, which is homologous to the 1-166 segment of the human



DDX6 (Fig. 9a), can be exported by the CRM1-dependent pathway (Smillie and Sommerville, 2002). One possibility is the capability of Xp54 1-164 segment to be recognized and shuttled by CRM1 arises from the removal of the downstream masking peptide segments, which exposes the putative NES. In other words, the CRM1-dependent export of the Xp54 1-164 segment could be a gain-of-function mutation resulting from the protein truncation. Therefore, we predicted that while the 1-166 segment of human DDX6 can be exported through the CRM1-dependent pathway, the full length DDX6, the DDX6 NTD, and the DDX6 141-483 segment cannot be exported through the CRM1-dependent pathway (Fig. 9c). As expected, the export of YFP-DDX6 1-166 was sensitive to the LMB treatment and the mutations in the putative NES (Fig. 9d and e). The localization of YFP-DDX6 FL, CTD, and 141-483, on the other hand, were not altered by the LMB treatment or the mutations (Fig. 9d and e). These results conformed to our prediction, thereby confirming that the putative NES is structurally inaccessible to CRM1 in the intact DDX6 NTD. Besides, the comparative analysis of the homologous region of the putative NES in other DBPs in the eIF4A-like clade revealed that this segment is highly enriched with the hydrophobic amino acids (AA) arranged in NES-like sequences (Fig. 11c-e). Yet no NES-like function has been reported for these homologous regions. These findings suggest that the putative NES of DDX6 may instead provide the essential hydrophobic interactions for the domain folding common to other DBPs.

4.5. The putative NLS, miRISC, and MALAT1 are dispensable for the nuclear entry of DDX6

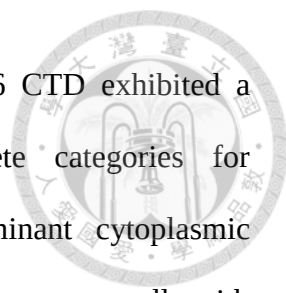


According to the shuttling model, the vertebrate DDX6 homologs are imported to the cell nuclei by the importin- α/β complex, which recognizes the K/R-rich NLS (Smillie and Sommerville, 2002; Weston and Sommerville, 2006). However, this has not yet received any empirical support. To test this model, we assessed if the nuclear import of YFP-DDX6 1-166 is sensitive to mutations in the putative NLS (Fig. 12a). As a result, the mNES YFP-DDX6 1-166 accumulated in the nuclei as efficiently as its WT counterpart upon the LMB treatment (Fig. 12a), the nuclear import of YFP-DDX6 1-166 is independent of the putative NLS. In conjunction, de novo prediction of the K/R-rich NLS by the NucPred program (Brameier et al., 2007) did not hit the putative NLS or any other sequences in DDX6 (Fig. 12c). Furthermore, the putative NLS differs from the typical bipartite K/R-rich NLS in two ways. First, the major K/R-rich cluster in the typical bipartite K/R-rich NLS comprises of 4 consecutive K/R (Marfori et al., 2012), yet the proposed major K/R-rich cluster in the putative NLS is composed of interspersed K/R. Second, while the inter-cluster distance in the typical bipartite K/R-rich NLS is at least 10 ~ 12 AA (Marfori et al., 2012), the inter-cluster distance in the putative NLS is only 9. This observation undermines the structural compatibility of the putative NLS with importin- α . These results strongly oppose the use of the K/R-rich NLS in the nuclear entry of DDX6. Recently, it was discovered that the core RNAi machinery can be shuttled nucleocytoplasmically by TNRC6A (Nishi et al., 2013). Thus, we wondered if DDX6 can be transported as a part of the miRNA-induced silencing complex (miRISC). To this end, we first disrupted the RISC formation by AGO2 KD

(Fig. 13a) and then assessed the nucleocytoplasmic distribution of DDX6 (Fig. 13b). However, the nucleocytoplasmic distribution of DDX6 was unaffected (Fig. 13b). In addition, we tested if the nuclear localization of DDX6 is related to MALAT1. As a result, MALAT1 KD had no effect on the distribution of DDX6 (Fig. 13c and d).

4.6. The CTD of DDX6 drives nuclear localization.

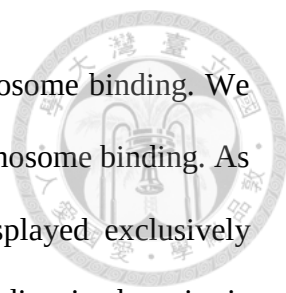
It is possible that DDX6 harbors some yet identified non-canonical NLS. To assess this probability, we first reduced the complexity by tracking the subcellular localization of the two domains of DDX6 separately (Fig. 14a and 15a). Surprisingly, we found that the DDX6 CTD explicitly localized in the cell nuclei of the HeLa and the HEK 293T cells (Fig. 4b), and was detected in the nuclear extract of the HeLa cells (Fig. 15b). Conversely, the DDX6 NTD was barely detected in the cell nuclei (Fig. 14b and 15b). To exclude the potential effect of the YFP-tag, which is capable of entering the cell nuclei by itself (Seibel et al., 2007), we also assessed the subcellular localization of the myc-tagged DDX6 CTD (myc-DDX6 CTD). As a result, the myc-DDX6 CTD also localized predominantly in the cell nuclei (Fig. 15c). Notably, the nucleolar localization of DDX6 CTD was an artifact resulting from 4% PFA fixation (Fig. 15d). Nonetheless, fixation does not change the general nuclear localization of DDX6 CTD (Fig. 15d). Together, these results mark the role of the DDX6 CTD in nuclear entry. To test if the DDX6 CTD can drive nuclear entry, we adopted the chimeric protein approach (Fan and Steitz, 1998; Kataoka et al., 2011; Kataoka et al., 2013; Michael et al., 1997; Siomi and Dreyfuss, 1995). We fused the DDX6 CTD or the DBR1 CTD to the C-terminus of YFP-PKM (Fig. 14c). As compared to YFP-PKM, YFP-PKM-DBR1 CTD was almost exclusively nuclear (Fig. 14d), confirming YFP-PKM could be imported by



C-terminally fused protein. On the other hand, YFP-PKM-DDX6 CTD exhibited a phenotypic continuum (Fig. 14d). We defined three discrete categories for YFP-PKM-DDX6 CTD: Type 1 represents cells with predominant cytoplasmic distribution; type 2 represents cells with even distribution; type 3 represents cells with predominant nuclear distribution. Cells with apparent nuclear signal (type 2 and 3) accounted for about 60% of the population two days after transfection (Fig. 14d), indicating that DDX6 CTD can facilitate nuclear entry in spite of the incomplete genetic penetrance. Since DDX6 CTD provides binding sites for its interacting partners, including CNOT1, PATL1, 4E-T and EDC3 (Kuzuoglu-Ozturk et al., 2016; Ozgur et al., 2015; Sharif et al., 2013; Tritschler et al., 2009; Tritschler et al., 2008), we mutated these sites to test if nuclear localization depends on interaction with these proteins (Fig. 16a). As results, none of the mutants showed abrogated nuclear localization (Fig. 16b). Similarly, mutations in the conserved helicase motifs or in the consensus phosphorylation sites did not abolish nuclear localization (Fig. 16c and d). Alternatively, we tested if the DDX6 CTD harbors one or more non-canonical NLS, which can be diverse in length and sequence. We constructed the DDX6 CTD truncation mutants by further deleting the first 60 and the last 63 amino acids in the DDX6 CTD (Fig. 14e). As a result, both YFP-DDX6 CTD-Δ1 and YFP-DDX6 CTD-Δ2 were majorly cytoplasmic (Fig. 14f). This finding indicated both the DDX6 301-360 and the 421-483 segments alone were insufficient for nuclear entry, suggesting these two segments do not function as stand-alone NLS. The same finding also suggested that the integral DDX6 CTD may be necessary for nuclear localization.

4.7. The CTD of DDX6 binds to the condensed chromosomes during mitosis.

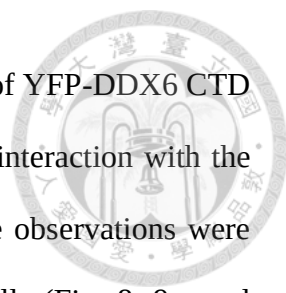
Two alternative mechanisms could potentially underlie the nuclear entry of DDX6. First, the diffuse DDX6 can be engulfed into the newly formed nuclei as the nuclear membrane reassembles at late telophase. However, this model is insufficient to explain the predominant nuclear localization of the DDX6 CTD. We then considered the second possibility that the DDX6 CTD may enter nuclei by “hitchhiking” on the mitotic chromosomes during mitosis. Supporting this notion, the chicken DDX6 has been detected on the mitotic chromosomes in the DT40 cell line (Ohta et al., 2010). Also, DDX6 has been detected in the chromatin extracts prepared from the HeLa-S3 cell line (Torrente et al., 2011). Thus, we tracked the localization of the DDX6 constructs in the mitotic HeLa cells. While YFP and YFP-DDX6 NTD exhibited exclusively diffuse pattern, YFP-DDX6 FL and CTD can bind to the mitotic chromosomes (Fig. 17a, b, and e). Under confocal microscope, YFP-DDX6 FL and CTD displayed indistinguishable chromosome binding pattern (Fig. 17b). These findings indicated the DDX6 CTD can facilitate the binding of the mitotic chromosomes. To test if the DDX6 CTD is responsible for the chromosome binding, we adopted a “domain swapping” approach. By fusing the DDX6 NTD and CTD (Fig. 18), we reconstituted a chimeric version of DDX6 (YFP-DDX6 Rec) which exhibited an elevated chromosome binding potential (Fig. 17c and e). YFP-PKM-DDX6 CTD, the chimeric protein that is the equivalent of swapping the DDX6 NTD for PKM in the YFP-DDX6 Rec, exhibited even higher chromosome binding potential (Fig. 17c and e). This suggested the DDX6 NTD could be inhibitory to the chromosome binding. Conversely, when the DDX6 CTD was removed, the resulting YFP-PKM was exclusively diffuse during in the mitotic cells



(Fig. 17c and e), suggesting PKM does not contribute to the chromosome binding. We further tested if the truncation in the DDX6 CTD abolishes the chromosome binding. As a result, both YFP-DDX6 CTD- $\Delta 1$ and YFP-DDX6 CTD- $\Delta 2$ displayed exclusively diffuse pattern (Fig. 17d and e). Collectively, the chromosome binding in the mitotic cells showed a striking positive correlation with the nuclear localization in the interphase cells, implying a causal link between these two phenomena.

4.8. The CTD of DDX6 facilitates nuclear entry by hitchhiking on the condensed chromosomes during mitosis.

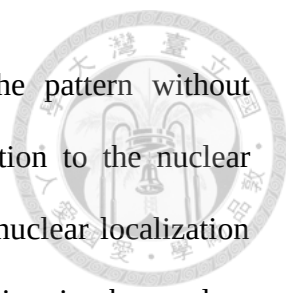
To verify the causal link between mitotic chromosome binding and nuclear localization, we monitor the chromosome-binding dynamics of the DDX6 constructs during mitosis. First, we observed that cytoplasmic YFP-PKM-DDX6 CTD quickly bound the condensed chromosomes after the nuclear membrane broke down (Fig. 19a). This finding demonstrated that the binding of the mitotic chromosome occurs at early prometaphase, and the subcellular localization of YFP-PKM-DDX6 CTD in the previous cell cycle does not limit this process. Furthermore, in some cells, the interaction between YFP-PKM-DDX6 CTD and the mitotic chromosomes persisted even after cytokinesis (Fig. 19b). Consequently, the chromatin bound YFP-PKM-DDX6 CTD became trapped in the newly formed nuclei (Fig. 19b). In other cells, the interaction between YFP-PKM-DDX6 CTD and the mitotic chromosomes declined during the mitosis (Fig. 19c), leaving a conceivably small amount of YFP-PKM-DDX6 CTD to be enclosed in the newly formed nuclei. This divergent dynamics during the mitosis was consistent with the divergent phenotypic continuum in the interphase cells (Fig. 14d). Conversely, we also observed the persistent interaction between YFP-DDX6



CTD and the mitotic chromosomes, and the subsequent entrapment of YFP-DDX6 CTD in the newly formed nuclei (Fig. 19d). As for YFP-DDX6 FL, the interaction with the mitotic chromosomes declined during the mitosis (Fig. 19e). These observations were also consistent with our previous observations in the interphase cells (Fig. 8, 9e, and 14b). Collectively, our findings point to a novel mechanism that DDX6 can enter the newly formed cell nuclei through hitchhiking on the mitotic chromosomes via its CTD.

4.9. 4E-T can mediate piggyback shuttling of DDX6.

Recently, it has been proposed that DDX6 may be shuttled nucleocytoplasmically in a piggyback style (Kirli et al., 2015). Indeed, among the known direct protein-protein interaction (PPI) partners of DDX6, 4E-T, PATL1, and LSM14B are directly transported by the CRM1 pathway (Fig. 8a) (Dostie et al., 2000; Kirli et al., 2015; Marnef et al., 2012). However, DDX6 outnumbers these three proteins in the average copy number per cell (Fig. 20), especially in the HeLa cells (Fig. 20d) as shown by several proteomic profiling efforts (Kirli et al., 2015; Kolker et al., 2012; Nagaraj et al., 2011; Wuhr et al., 2014). This is consistent with the observation that the DDX6 localization was unaffected under the LMB treatment (Fig. 8, 9d, and 9e). To overcome this limitation imposed by the protein stoichiometry, we assessed the localization of the co-expressed CFP-DDX6 following YFP-4E-T, YFP-PATL1, and YFP-LSM14A overexpression (OE) under the LMB treatment. Since LSM14A (Fig. 8), unlike its paralog LSM14B (Kirli et al., 2015), is not a shuttling protein, we observed no change in the subcellular localization of YFP-LSM14A and CFP-DDX6 under the LMB treatment (Fig. 21). Conversely, YFP-PATL1 accumulated in the nuclei under the LMB treatment (Fig. 21). Yet YFP-PATL1 displayed a diffuse nuclear pattern in nearly all the



observed cells (Fig. 21). This was drastically different from the pattern without co-expressing CFP-DDX6, which is characterized by the localization to the nuclear granules (Fig. 8). Although we suspected this alteration in the subnuclear localization was caused by the co-expressed CFP-DDX6, no apparent elevation in the nuclear CFP-DDX6 level was observed under the LMB treatment (Fig. 21). In contrast, YFP-4E-T accumulated and localized to the nuclear granules under the LMB treatment regardless of the co-expression of CFP-DDX6 (Fig. 21). Strikingly, CFP-DDX6 co-localized with YFP-4E-T within the nuclear granules under the LMB treatment (Fig. 21). This finding indicates that DDX6 can be co-transported with 4E-T in a piggyback manner.

5. **Discussion**

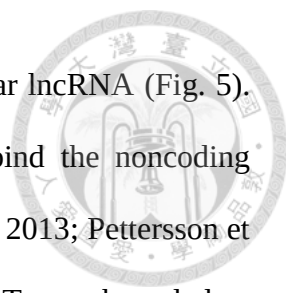
5.1. **The widespread nuclear localization of DDX6**

In this chapter, we confirmed the nuclear localization of DDX6 in two different cell lines using three different methods (Fig. 1-4). Notably, the HeLa cells and HEK 293T cells were derived from very different origins and have very distinct cytological properties. Our results aligned with the previous reports that DDX6 is present in the nuclei of the MKN45 gastric cancer cells and the DM1-affected fibroblasts (Iio et al., 2013; Pettersson et al., 2014). Taken together, DDX6 is now known to localize in the nuclei of four distinct cell types derived from human. Since DDX6 is ubiquitously expressed in the human cells, we expected the nuclear localization of DDX6 to be widespread in the different human cell types. This notion and our empirical data also conform to the widespread nuclear localization of the DDX6 subfamily members in the



eukaryotes (Haimovich et al., 2013; Iio et al., 2013; Kawahara et al., 2014; Kruk et al., 2011; Lodomery et al., 1997; Minshall et al., 2001; Pettersson et al., 2014; Smillie and Sommerville, 2002; Torrente et al., 2011; Yoshida-Kashikawa et al., 2007). The functional implications of the nuclear DDX6, as summarized in the previous chapter, therefore remain an interesting field to be further explored.

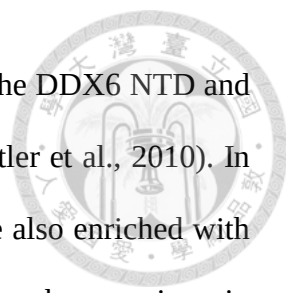
In our WB and MS experiments, we frequently observed comparable signals in the same amount of the cytoplasmic and the nuclear extracts (Fig. 1). Since the protein yields in the cytoplasmic and the nuclear extracts were approximately 9:1 ~ 4:1, we estimated that the molar ratio between the cytoplasmic and the nuclear DDX6 in HeLa does not exceed 4:1. That is, the nuclear DDX6 is unlikely to account for more than 20% of the total DDX6. This estimation is conservative, but it is also consistent with many of the previous estimations. For instance, in the stage III *Xenopus* oocytes, the nuclear Xp54 was estimated to account for about 16.7% of the total Xp54 (Smillie and Sommerville, 2002). In the rat testis, the nuclear Ddx6 was estimated to account for approximately 7% of the total Ddx6 (Kawahara et al., 2014). Notably, in these two instances, the nuclear level of the DDX6 homologs was known to be developmentally regulated (Kawahara et al., 2014; Smillie and Sommerville, 2002). In addition, in the clam *Spisula*, the DDX6 homolog p47 was reported to undergo drastic nuclear translocation (Minshall et al., 2001). Thus, the exact proportion of the nuclear DDX6 in fact depends on the organisms, cell types, and the developmental stages, and may also vary with different means of detection. Collectively, these results indicated that while the cellular DDX6 is majorly cytoplasmic, a minor proportion of it is consistent present in the nuclei of many, if not all, cell types.



Furthermore, we also demonstrated that DDX6 can bind nuclear lncRNA (Fig. 5). This finding aligned with the previous reports that DDX6 can bind the noncoding regions of the mRNA or the lncRNA (de Vries et al., 2013; Iio et al., 2013; Pettersson et al., 2014; Saito et al., 2011; Ward et al., 2011; Yoon et al., 2012). To our knowledge, this finding is also the first description of the interaction between DDX6 and a lncRNA that is highly enriched in the cell nuclei. One similar report to our finding is that DDX6 binds the CUG-expanded DMPK mRNA (Pettersson et al., 2014), which is also retained in the nuclei. Parallel to this, we discovered DDX6 also binds other RNA enriched in the nucleus, including other lncRNA (Chap. 3). We speculated DDX6 can potential regulate the lncRNA as well as the mRNA. Yet DDX6 does not regulate the expression level or the subnuclear localization of MALAT1 (Fig. 6 and 7). Notwithstanding, we demonstrated that DDX6 can regulate the expression noise of both the mRNA and the lncRNA (Chap. 3). In addition, we cannot fully dismiss the possibility that DDX6 may participate in the structural regulation of the lncRNA without further investigations. As such, the role and the mechanism of DDX6 in the regulation of the lncRNA regulation represent yet another field for future research.

5.2. On the nucleocytoplasmic shuttling model of DDX6

In this chapter, we detailedly scrutinized of the shuttling model. First, we showed that both the endogenous and the exogenous DDX6 are insensitive to the LMB treatment (Fig. 8). This finding agreed with several precedent studies (Kirli et al., 2015; Marnef et al., 2012; Nishi et al., 2013; Smillie and Sommerville, 2002). Notably, the localization of DDX6 was still unchanged even when we treated the cells with 5 ~ 10 times of the commonly used dose of LMB (Fig. 8b). Second, our structural analysis

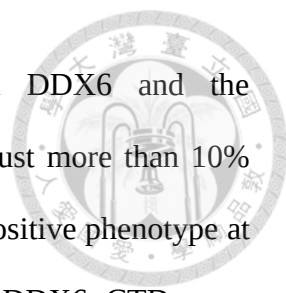


revealed that the putative NES is buried in the hydrophobic core of the DDX6 NTD and was inaccessible to CRM1 (Fig. 9b and 10) (Dong et al., 2009; Guttler et al., 2010). In parallel, the homologous regions in the other eIF4A-like DBPs were also enriched with the hydrophobic residues (Fig. 11), indicating a conserved role for these regions in protein folding. Furthermore, the NES-like linear motifs occur quite frequently in the protein hydrophobic core. As a consequence, they are also frequently mistaken as functional NES because indeed they could serve as NES when put into suitable protein context (Hantschel et al., 2005; Kadlec et al., 2004). Consistent with this notion, our experimental results strongly suggested the capability of the putative NES of DDX6 to facilitate the export of the 1-166 segment arose as a gain-of-function mutation that resulted from the truncation of the DDX6 NTD (Fig. 9). On the other hand, we provided the first assessment of the putative NLS of DDX6 which also turned out against its validity (Fig. 12). Our results conform with the structural features of the K/R-rich NLS and the mode of recognition by importin- α (Marfori et al., 2012). More importantly, we demonstrated that the nuclear localization of DDX6 instead depends on its CTD (Fig. 14, 17, and 19) or 4E-T (Fig. 21). Collectively, these results refuted the shuttling hypothesis.

5.3. The alternative mechanisms for the nucleocytoplasmic distribution of DDX6:

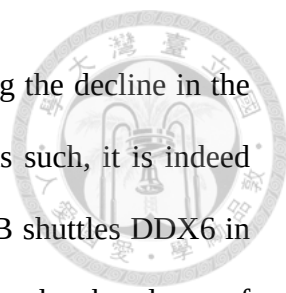
A new synthesis

Our results pointed to a novel mechanism that DDX6 enters the newly formed nuclei through hitchhiking on the mitotic chromosome with its CTD (Fig. 22, the upper panel). This model is consistent with the previous reports that the DDX6 homologs can be detected on the mitotic chromosomes or in the chromatin extract. With regards to the



nuclear entry, our results suggested the interaction between DDX6 and the chromosomes/chromatin is indeed functional. Interestingly, while just more than 10% of the cells expressing YFP-DDX6 FL displayed the chromosome-positive phenotype at prometaphase, approximately 80% of the cells expressing YFP-DDX6 CTD were chromosome-positive. Similarly, approximately 80% of the cells expressing YFP-PKM-DDX6 CTD displayed the chromosome-positive phenotype at prometaphase. This indicated that the DDX6 NTD, but not the YFP tag or PKM, interfered with the chromosomes binding. Nonetheless, about 40% of the cells expressing YFP-DDX6 Rec were chromosome-positive. This increase in the chromosome-positive cell percentage in YFP-DDX6 Rec as compared to YFP-DDX6 FL was accompanied by the loss of the PB localization. This indicated the normal DDX6 function was disrupted by the alteration in the hinge region (Fig. 16), which most probably affected the inter-domain interactions with DDX6. As such, we suspected the chromatin binding during mitosis can be auto-regulated by the inter-domain interactions within the DDX6.

Alternatively, we demonstrated that DDX6 can be transported in a piggyback manner through its direct PPI partner 4E-T (Fig. 22, the lower panel). This finding enlisted DDX6 as another transport cargo of 4E-T in addition to EIF4E (Dostie et al., 2000). On the other hand, even though we had no explicit evidence for the piggyback style transportation by PATL1, we observed the co-expressed CFP-DDX6 affected the subnuclear localization of YFP-PATL1 (Fig. 8 and 21). Thus, we cannot fully dismiss the possibility that PATL1 shuttles DDX6. Moreover, it was originally proposed that the *Xenopus* LSM14B (xRAP55B, xRAPB) may serve as a transporter of Xp54 (Kirli et al., 2015). Interestingly, in the developing *Xenopus* oocytes, the nuclear Xp54 level



decreases after stage III (Smillie and Sommerville, 2002), coinciding the decline in the total xRAPB after stage III (Ladomery and Sommerville, 2015). As such, it is indeed possible xRAPB shuttles Xp54 in the *Xenopus* oocytes and LSM14B shuttles DDX6 in the human cells. However, in the biological context of the HeLa cells, the abundance of DDX6 greatly exceeds that of PATL1, 4E-T, and LSM14B (Fig. 18). Thus, the shuttling rate can be very slow. This is consistent with our observation that the nucleocytoplasmic localization of DDX6 is insensitive to the LMB treatment in the HeLa cells (Fig. 8). Conversely, when we elevated the 4E-T expression through YFP-4E-T OE, we could observe the co-transportation of YFP-4E-T and CFP-DDX6 (Fig. 21). Together, these results suggested that this mode of transportation is regulated by the protein stoichiometry.

To put into the historic context, these two mechanisms were discovered along different lines. While the hitchhiking mechanism stemmed from the rejection of the shuttling model, the piggyback mechanism was rooted in the evidence that supported the CRM1-dependent export. Their parallel existence therefore reconciled the conflicting ideas in the past research. Physiologically, the hitchhiking mechanism is restricted to the proliferating cells, whereas the piggyback mechanism is constrained by the protein stoichiometry. But both mechanisms distribute DDX6 across the nucleus and the cytoplasm. As such, they may dominate under different biological context, and complementing/buffering the effect of one another.

6. Figures

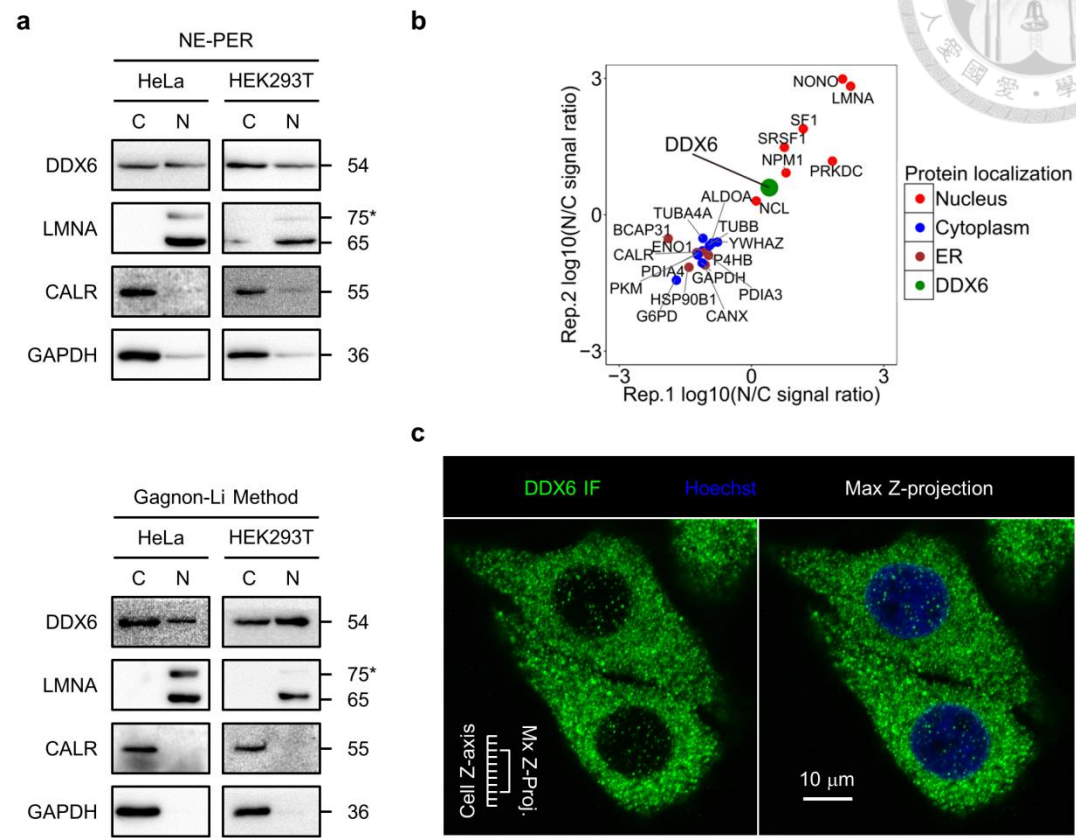


Figure 1. DDX6 is present in the human cell nuclei.

a, The representative WB of the DDX6 in the nuclear and the cytoplasmic extract prepared from different cell lines with different methods. Equal amounts (μg) of nuclear (N) and cytoplasmic (C) extracts were analyzed by WB. Asterisk denotes the 75 kDa LMNA isoform, which is expressed at lower level in the HEK 293T cells than in the HeLa cells. LMNA: lamin A/C, a nuclear marker; GAPDH: a cytosolic marker; CALR: an ER membrane marker. **b**, Scatter plot showing the comparison of the protein nucleocytoplasmic signal ratios between the two replicates of the quantitative MS data. The original samples were fractionated by NE-PER to yield cytoplasmic and nuclear

extracts as in **a**. The N/C signal ratios were calculated using the summed-up extracted ion chromatogram (XIC) intensities, an estimate of the protein abundance. Several known cytosolic (blue), ER (brown), and nuclear (red) proteins were shown side-by-side with DDX6 (green). **c**, The representative IF image of DDX6. DDX6 (shown in green) was detected by IF in the fixed HeLa cells. 11 optical slices were captured with confocal microscope along the Z-axis across a 10 μm range with 1 μm spacing to cover the thickness of the cell. To exclude cytoplasmic staining at top and bottom of the cells, only the middle 7 optical slices were Z-projected (maximum intensity). Nuclei were stained with Hoechst (shown in blue). Scale bar = 10 μm .

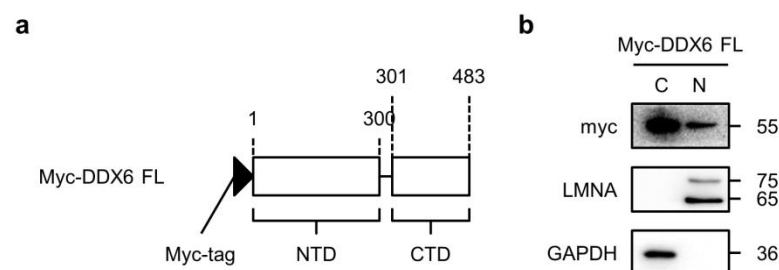


Figure 2. The WB detection of the exogenous DDX6 in the human cell nuclear extract

a, The diagram depicting the myc-tagged DDX6. **b**, The WB showing the detection of myc-DDX6 in both the nuclear and the cytoplasmic extracts. LMNA: nuclear marker; GAPDH: cytoplasmic marker.



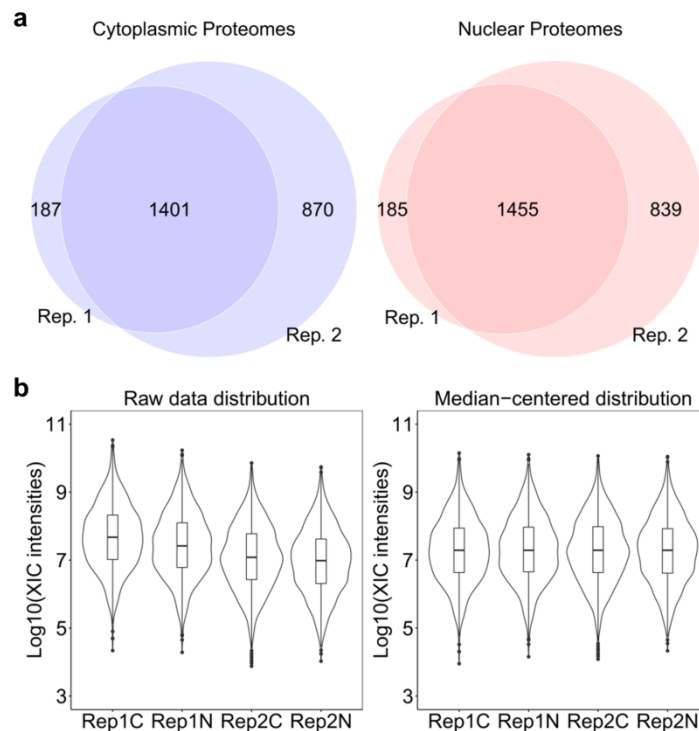


Figure 3. The analysis of the quantitative MS data of the control siRNA-treated HeLa cells

a, The Venn diagrams displaying the overlap of the detected proteins in the two biological replicates. The nuclear and the cytoplasmic fractions were plotted separately. The gene symbols of the major protein in each protein group were used for plotting the Venn diagrams. The quantified protein groups that do not map to official gene symbols were removed from this comparison. No more than 10 protein groups were removed from each fraction. **b**, The violin plots showing the normalization of the log₁₀-transformed XIC intensity distribution. The distributions were median-centered to give comparable distributions.

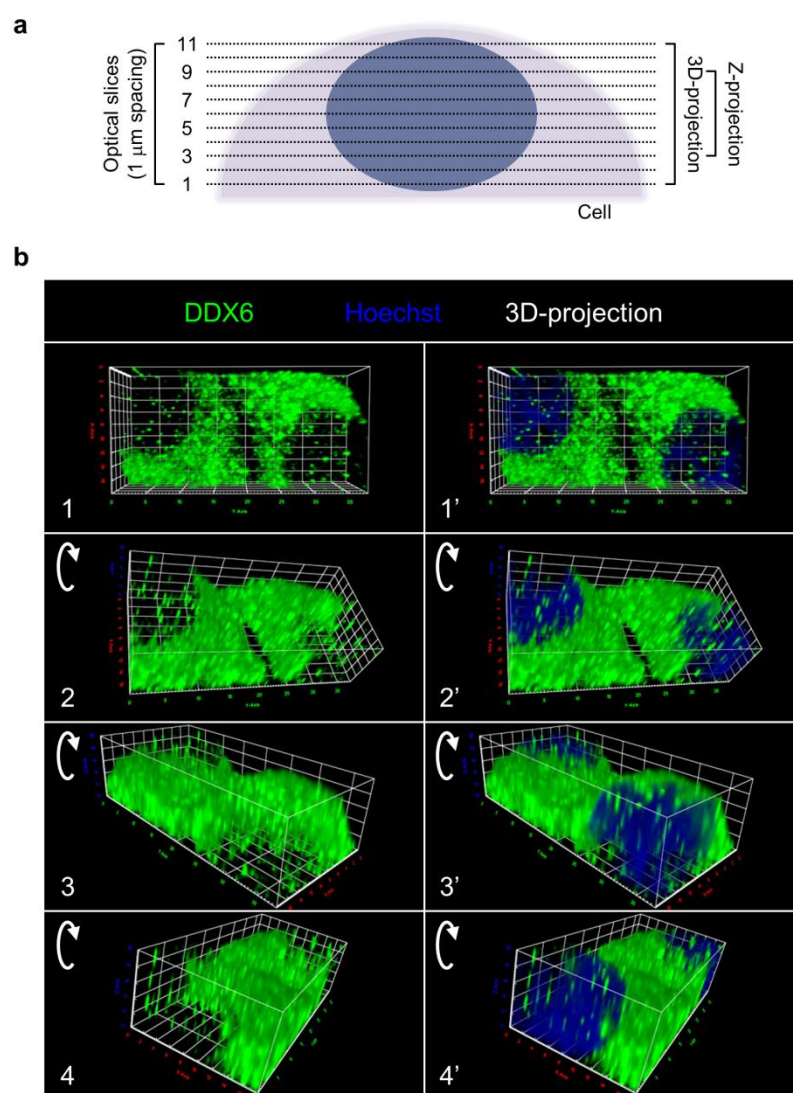


Figure 4. The IF detection of the nuclear DDX6

a, The diagram showing the important spatial parameters of the confocal imaging for Fig. 1c. A total of 11 optical slices were captured across a 10 µm range along the Z axis (covering the cell height) with 1 µm spacing. The resulting image stack was projected 3-dimensionally here, or along the Z-axis (maximum intensity projection; Fig. 1c). For 3D-projection, all 11 optical slices were used. For Z-projection, the middle 7 optical slices were used (Fig. 1c). **b**, The 3D reconstruction of the nuclear staining of DDX6.

The two representative cells shown in Fig. 1c were shown. Grid lines in the X and the Z axes represent 2 μm intervals; and in the Y axis 5 μm intervals.



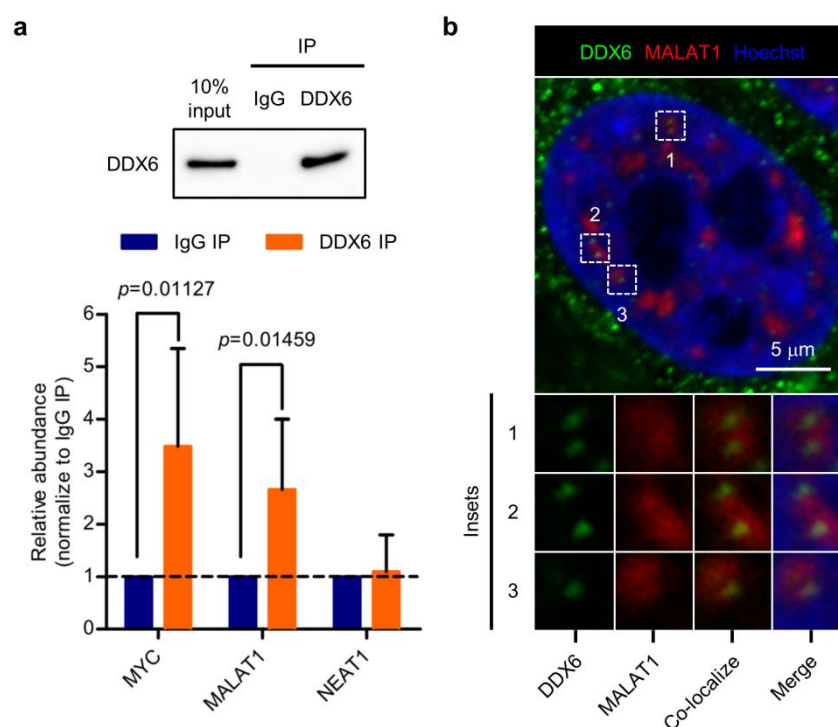


Figure 5. DDX6 binds the nuclear lncRNA MALAT1.

a, The bar graph showing the enrichment of MALAT1 by the anti-DDX6 IP. The representative WB confirmed anti-DDX6 IP specificity. Relative abundance of RNA was assessed with RT-qPCR. MYC mRNA served as a positive control. The bar heights and the error bars represent means and standard deviations of fold enrichments (anti-DDX6/IgG) calculated from 6 independent experiments. The dashed line represents the hypothetical value under null hypothesis. The P values were calculated using right-tailed one-sample t test. **b**, The representative image of DDX6 IF (green) and MALAT1 FISH (red) double staining obtained from confocal imaging. A single optical slice of the representative nucleus was shown. The insets displayed the detailed views of DDX6 co-localizing with and docking to MALAT1. The nucleus was stained by Hoechst (shown in blue). Scale bar = 5 μ m.

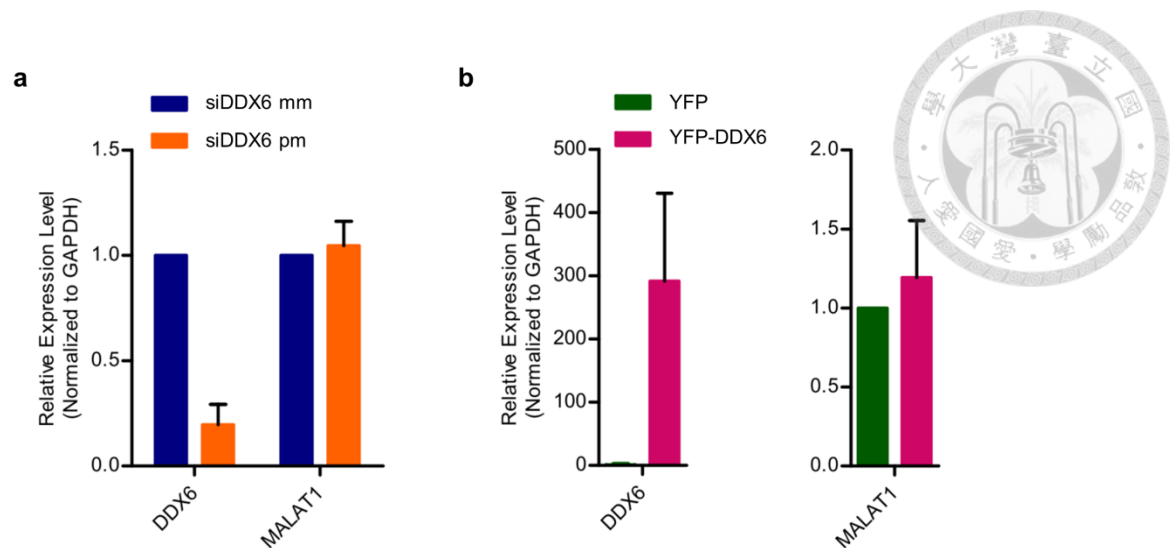


Figure 6. DDX6 has limited effect on the expression level of MALAT1.

a and b, The bar graphs showing the expression level of MALAT1 following DDX6 KD (**a**) or OE (**b**). The RNA expression levels of DDX6 and MALAT1 were assessed with qPCR. Relative expression levels were normalized to GAPDH mRNA. The bar heights and the error bars represent means and standard deviations of relative expression levels calculated from 5 (**a**) and 6 (**b**) biological replicates. **a**, siDDX6-mm: mismatch control siRNA; siDDX6-pm: DDX6-targeting siRNA. **b**, YFP: YFP-tag only control; YFP-DDX6: OE of YFP-DDX6 FL.

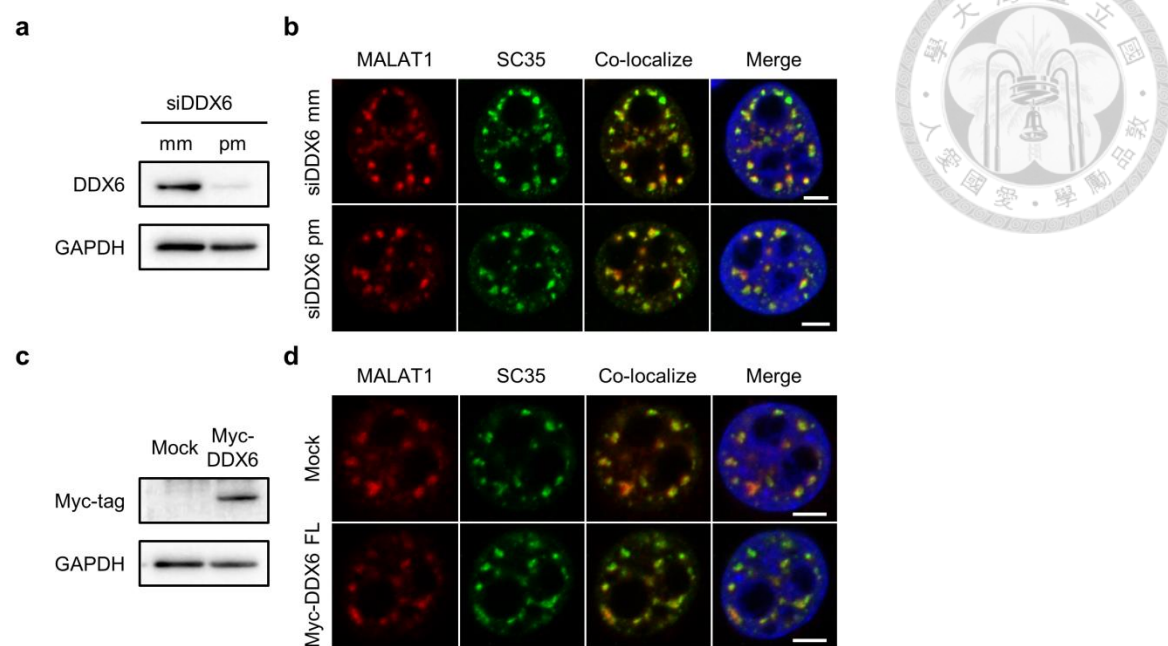


Figure 7. DDX6 has limited effect on the localization of MALAT1.

a and c, The WB confirming DDX6 KD (**a**) and myc-DDX6 OE (**c**) efficacy. GAPDH: loading control. **b and d**, The representative images of the MALAT1 subnuclear localization following DDX6 KD (**b**) and myc-DDX6 OE (**d**). The MALAT1 localization to nuclear speckles was assessed with MALAT1 FISH (shown in red) and phosphorylated SC35 IF (nuclear speckle marker; shown in green) double staining. Nuclei were stained with Hoechst (shown in blue). Scale bar = 5 μ m.

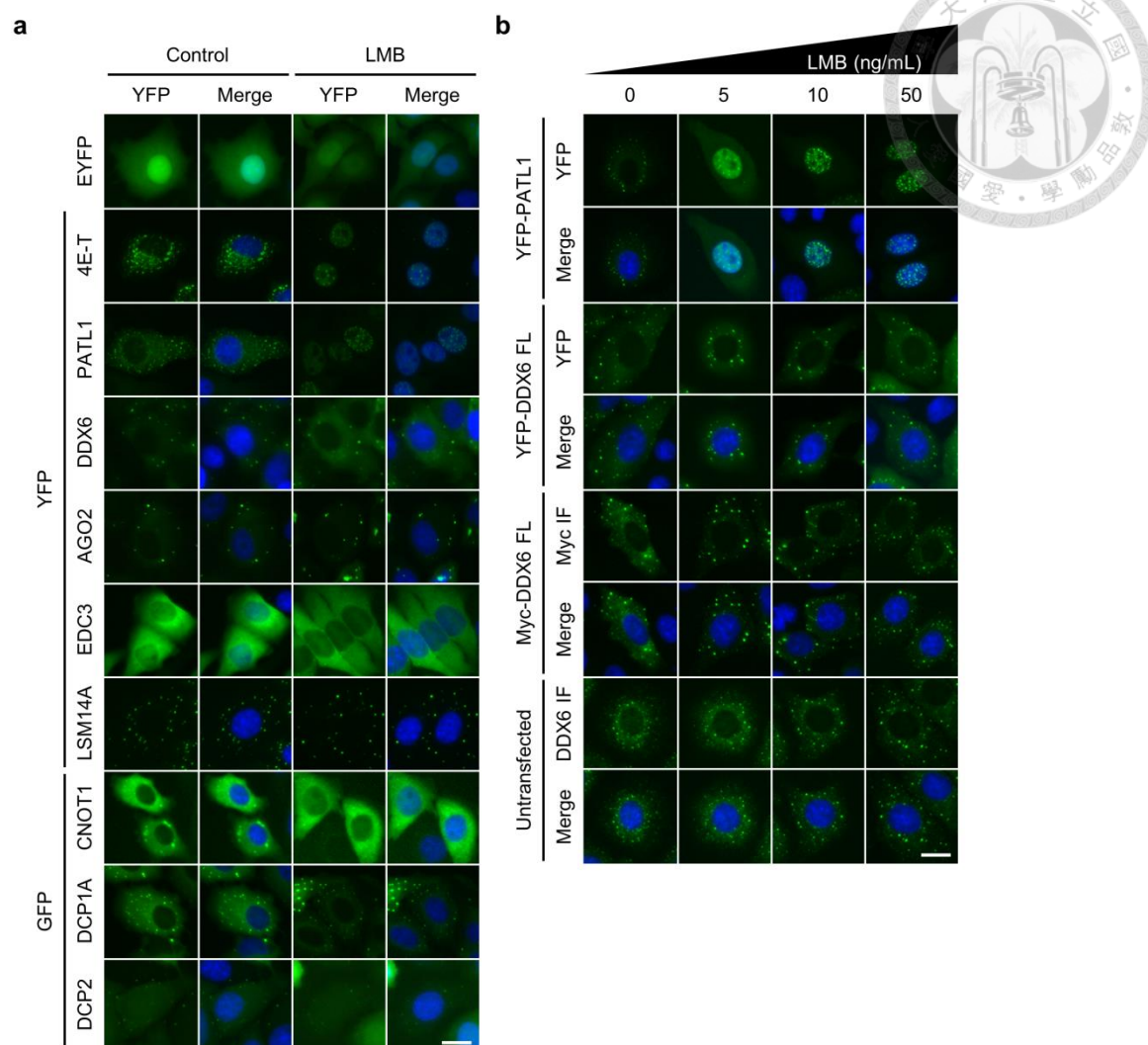


Figure 8. The nucleocytoplasmic localization of DDX6 is insensitive to the LMB treatment even under high dosage.

a, The representative images of the YFP- or GFP-tagged PB components under the control or the LMB treatment. The HeLa cells expressing the PB components were treated with LMB (5 ng/mL) or the solvent control (70% [v/v] MeOH in ddH₂O) of the same volume for 5 hr prior to imaging. **b**, The representative images of the endogenous and the exogenous DDX6 under different doses of the LMB treatment. The HeLa cells

expressing YFP-PATL1 (positive control), YFP-, myc-DDX6, and the untransfected HeLa cells were treated with different doses of LMB for 5 hr prior to imaging. For the zero dose treatment, the cells were treated with the solvent control of the same volume to the highest dose treatment. **a and b**, YFP, GFP, and IF signals were shown in green. Nuclei were stained with Hoechst and their signals were shown in blue. Scale bars = 10 μm .

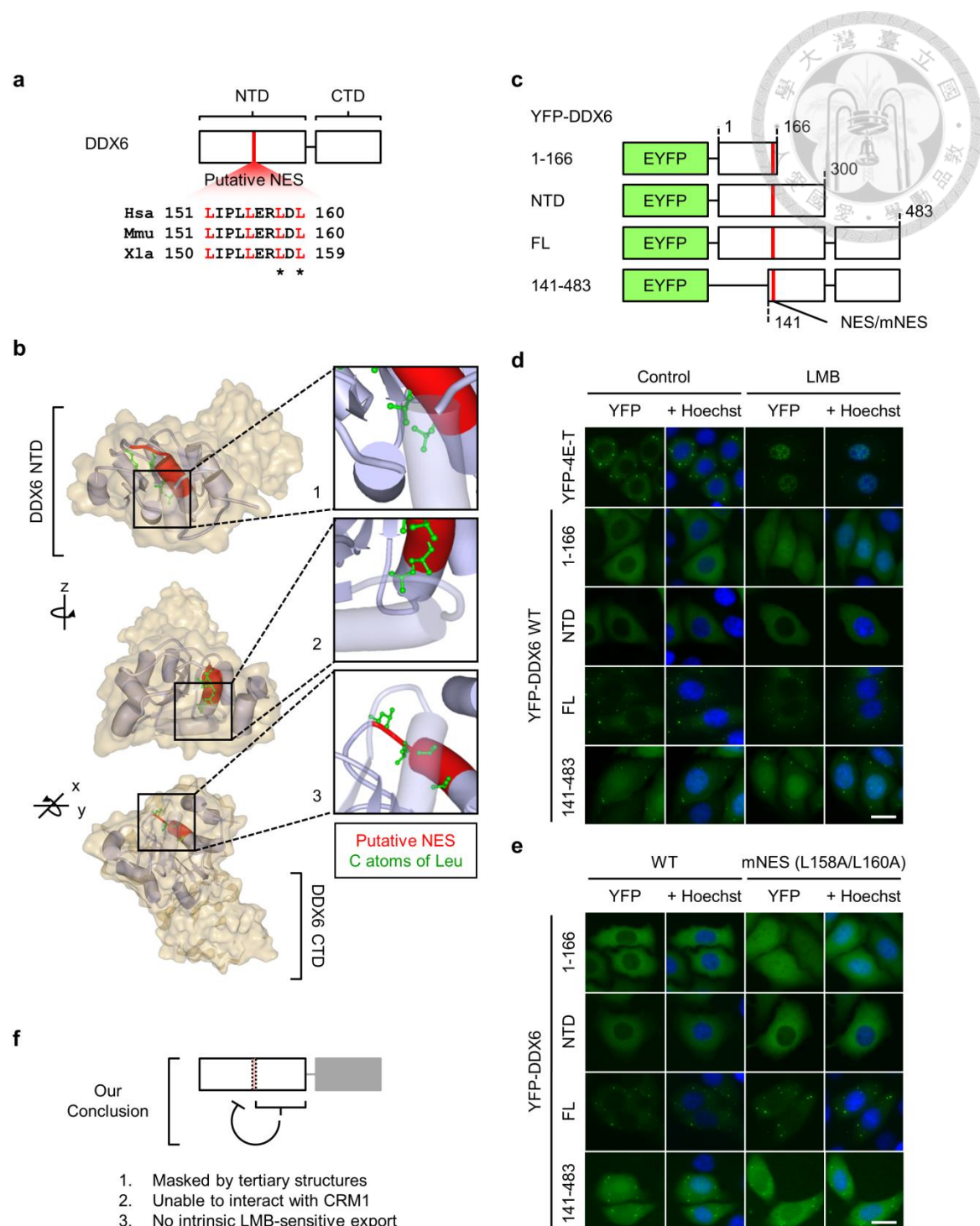
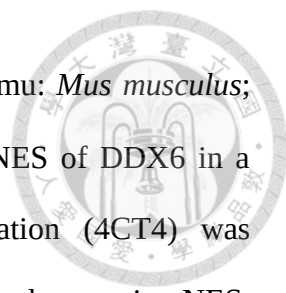


Figure 9. The putative NES of DDX6 is structurally inaccessible to CRM1.

a, The diagram showing the location and the sequence conservation of the putative NES of DDX6. The conserved Leu residues were colored in red. The asterisks marked the



two Leu residues targeted for mutagenesis. Hsa: *Homo sapiens*; Mmu: *Mus musculus*; Xla: *Xenopus laevis*. **b**, The representative views of the putative NES of DDX6 in a crystal structure. DDX6 in the ATPase-incompatible conformation (4CT4) was visualized, colored, rotated to highlight the structural landscape near the putative NES. The cylinders represent α -helices; the arrows represent β -sheets. The putative NES segment was colored in red. The carbon atoms of the Leu residues in the putative NES were colored in green. The insets: the magnified views of the boxed areas near the putative NES. **c**, The diagram depicting the YFP-DDX6 constructs used for the functional analysis of the CRM1-dependent export. All constructs were N-terminally tagged with YFP, containing different segments of the DDX6, and either WT or mutated putative NES. The 1-166 segment spanned the 1-166 AA of the WT DDX6; NTD: the 1-300 AA; FL: the 1-483; the 246-483 segment: the 246-483 AA. **d**, The representative images of the YFP-DDX6 constructs under the control or the LMB treatment. The HeLa cells expressing YFP-4E-T (positive control) or the YFP-DDX6 constructs with the WT putative NES were treated with LMB (5 ng/mL) or the solvent control for 5 hr prior to imaging. **e**, The representative images of the YFP-DDX6 constructs carrying either WT or mutant NES (mNES). The mNES contained L158A and L160A substitutions. **d and e**, The signals of YFP were shown in green and were overlay with the Hoechst signals (blue). Scale bars = 20 μ m. **f**, The diagram summarizing our structural and function analyses of the accessibility of the putative NES.

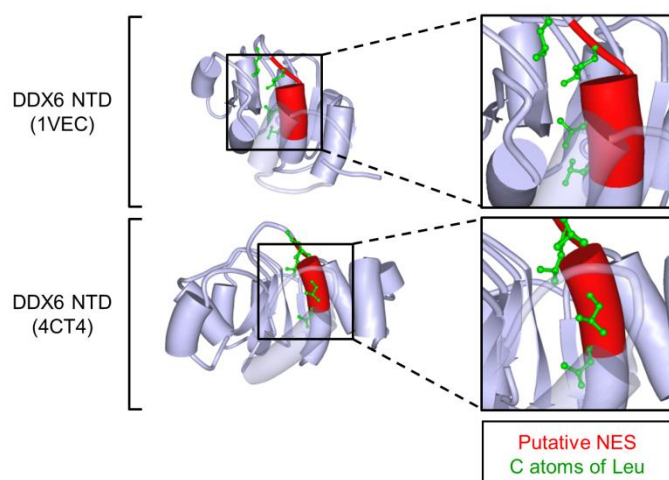


Figure 10. The visualization of the putative NES in the other two published DDX6 structures

The NTD of DDX6 in the 1VEC and the 4CT4 structures were visualized. The cylinders represent α -helices and the arrows represent β -sheets. The putative NES segment was colored in red. The carbon atoms of the Leu residues in the putative NES were colored in green. The insets: the magnified view of the nearby area around the putative NES. 1VEC: The DDX6 NTD; 4CT4: DDX6 complexed with the CNOT1 MIF4G domain.

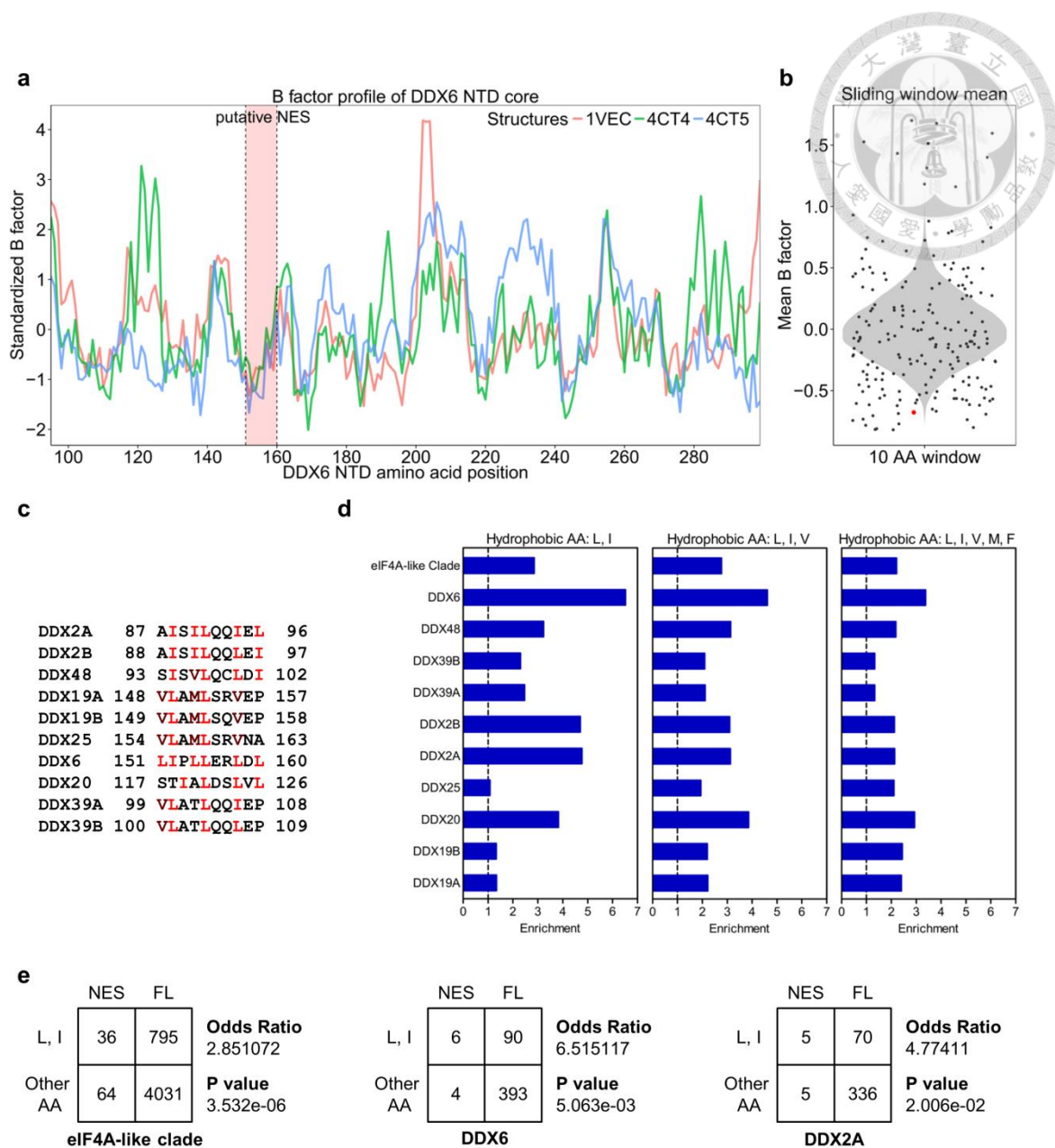
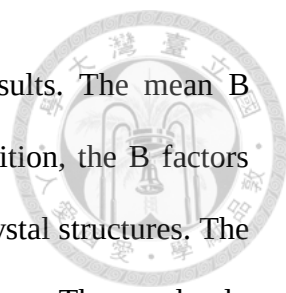


Figure 11. The putative NES is stably folded in the NTD of DDX6 and may serve structural roles

a, The B factor spectrum of the NTD of DDX6 in three different published structures. The B factors from the three structures were standardized to give comparable values. The colored area represents the putative NES segment. **b**, The jitter plot showing the



comparison between real mean B factors and the permutation results. The mean B factors were taken over a sliding window of 10 AA. For each position, the B factors were given by the mean of the standardized values from the three crystal structures. The real data were plot as points. Each point represents one window. The randomly permuted data (100 permutations, 19600 windows) were summarized was the grey distribution, representing the null distribution. P value was estimated from this null distribution. The putative NES has $p = 7.15 \times 10^{-4}$. **c**, The sequence alignment of the homologous segments of the putative NES in other eIF4A-like DBPs. Leu and Ile residues were colored in red. Met and Val residues were colored in dark red. **d**, The bar graphs showing the enrichment of the hydrophobic AA in the homologous segments of the putative NES in other eIF4A-like DBPs. The enrichment ratios were calculated using FL protein sequences as backgrounds. The dashed lines represent no enrichment. **e**, The contingency tables for the hydrophobic AA in the homologous segments of the putative NES in other eIF4A-like DBPs. Three representative tables were shown, including the eIF4A-like DBPs as a whole, DDX6, and DDX2A. P values were calculated using the Fisher exact test.

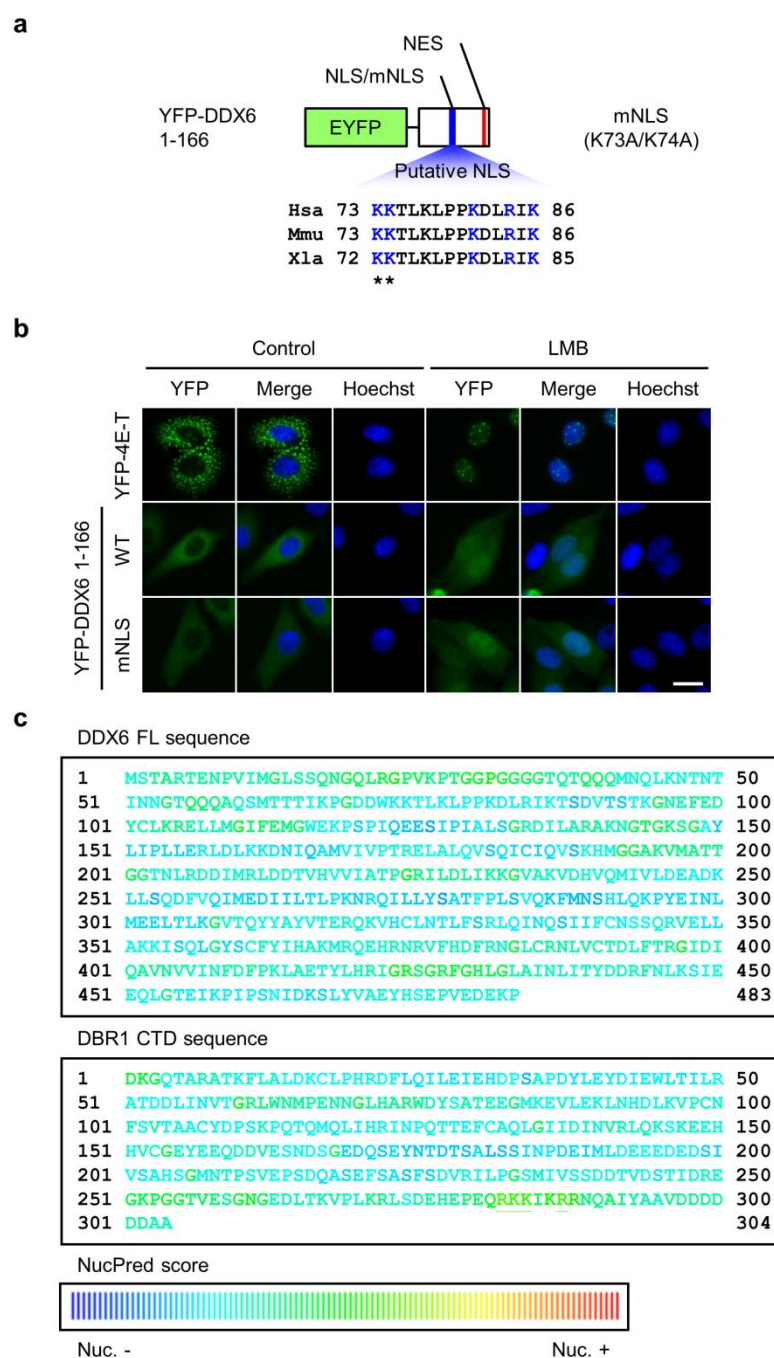


Figure 12. The nuclear entry of DDX6 is independent of the putative K/R-rich NLS.

a, The diagram depicting the use of YFP-DDX6 1-166 as a reporter construct for the

putative NLS. The sequence alignment of the putative NLS is shown. The asterisks mark the Lys residues targeted for mutagenesis. **b**, The representative images of the WT and the putative NLS mutant (mNLS) YFP-DDX6 1-166 under the control or the LMB treatment. YFP-4E-T was used as a positive control. The HeLa cells were treated with LMB (5 ng/mL) for 5 hr prior to imaging. The YFP signals were shown in green. The nuclei were stained with Hoechst (shown in blue). Scale bar = 20 μ m. **c**, The de novo prediction of the K/R-rich NLS in DDX6 sequence. The prediction was carried out using the NucPred program. NucPred successfully called the K/R-rich NLS in the DBR1 CTD.

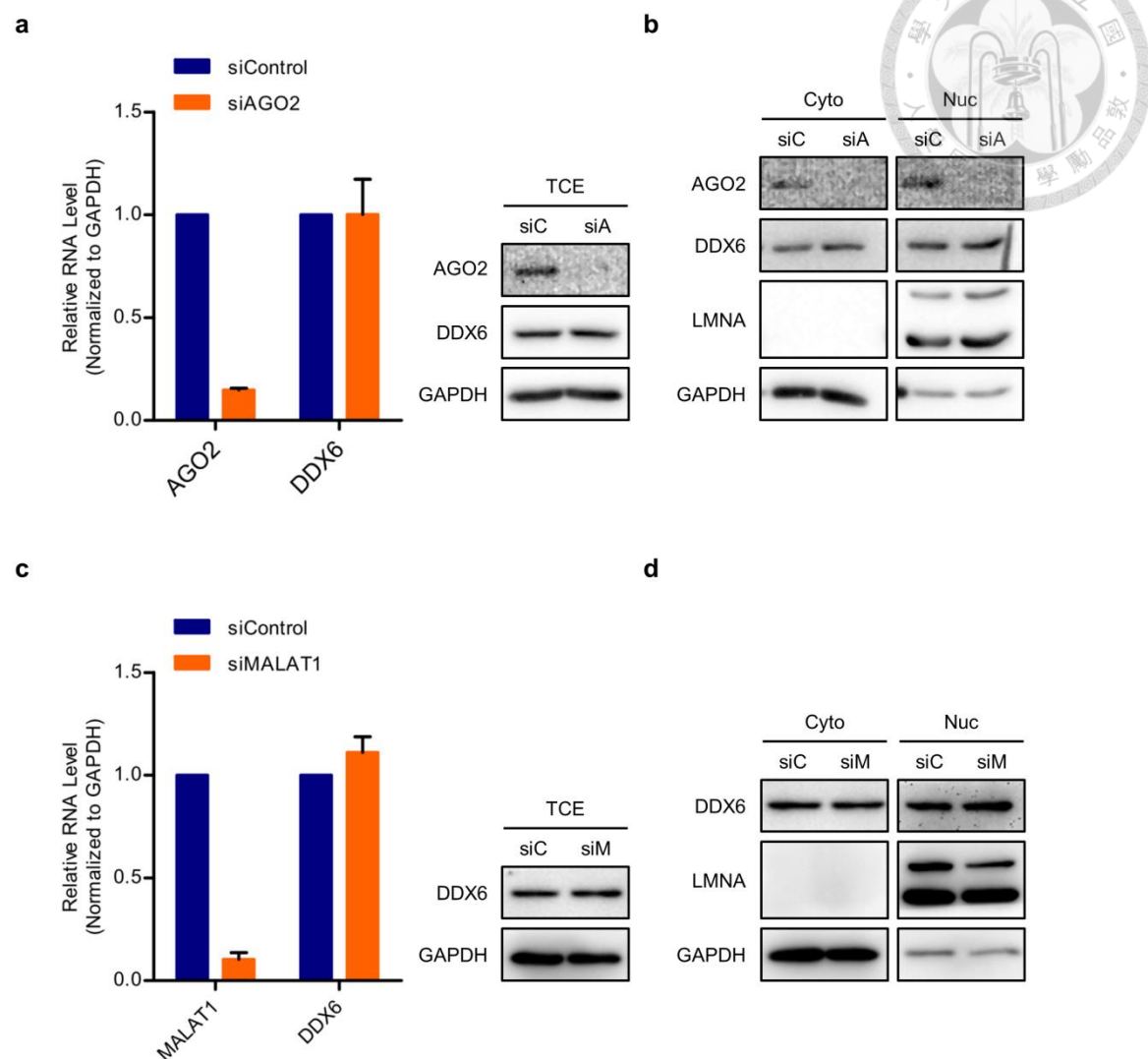


Figure 13. RISC and MALAT1 are not required for the nuclear entry of DDX6.

a and c, The bar graphs of the qPCR results and the representative WB confirming AGO2 (**a**) and MALAT1 (**c**) KD. DDX6 expression level was also assessed. GAPDH was used as control for qPCR and WB. **b and d**, The representative WB showing the DDX6 level in the nuclear and the cytoplasmic extracts following AGO2 (**b**) or MALAT1 (**d**) KD. The HeLa cells were fractionated using the NE-PER reagents two days after siRNA transfection. LMNA: nuclear marker; GAPDH: cytoplasmic marker.

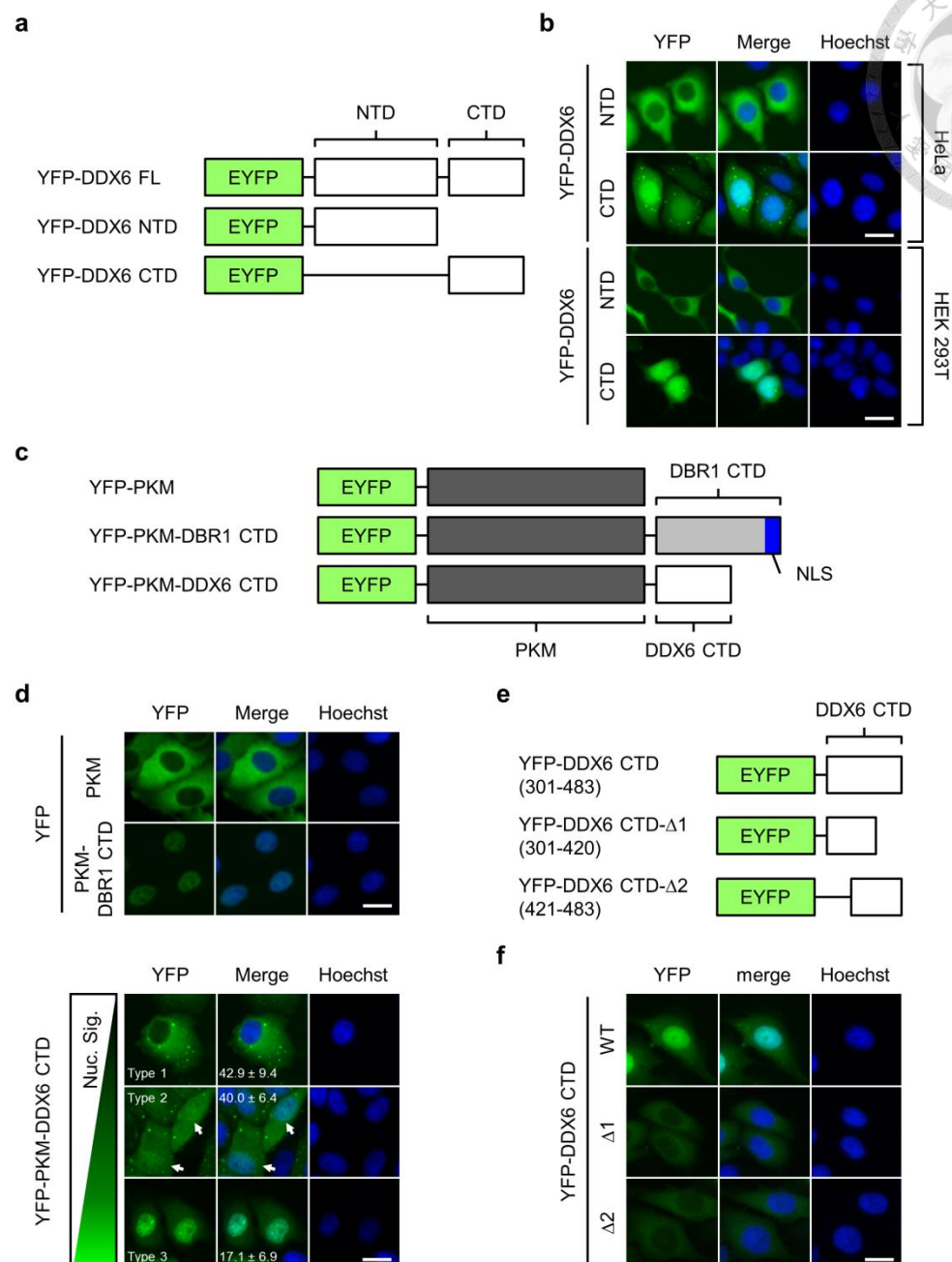
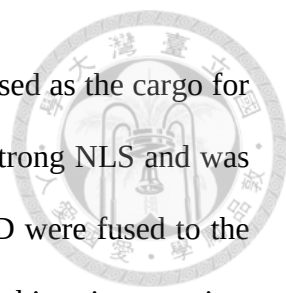


Figure 14. The CTD of DDX6 drives nuclear entry.

a, The diagram showing the YFP-DDX6 domain constructs. NTD: the 1-300 AA. CTD: 301-483 AA. **b**, The representative images of the YFP-DDX6 NTD and CTD in two different model cell lines. **c**, The diagram showing the YFP-tagged chimeric proteins



used for nuclear localization assay. PKM, a cytosolic enzyme, was used as the cargo for active transport. The CTD of DBR1, a splicing factor, contained a strong NLS and was used as a positive control. Both the DBR1 CTD and the DDX6 CTD were fused to the C-terminus of PKM. **d**, The representative images of the chimeric proteins. YFP-PKM-DDX6 CTD exhibited a phenotypic continuum, which was categorized into three types. Type 1: majorly cytoplasmic; type 2: evenly distributed (the two cells pointed by the white arrow); type 3: majorly nuclear. Numbers represents mean $\pm$ standard deviation of the three phenotypes within the population. The means and the standard deviations were calculated from three biological replicates. **e**, The diagram showing the truncation mutants of the YFP-DDX6 CTD. Δ 1: the 301-420 AA; Δ 2: the 361-483 AA. **f**, The representative images of the DDX6 CTD truncation mutants. **b**, **d**, **and f**, The YFP signals were shown in green. The nuclei were stained with Hoechst (shown in blue). Scale bars = 20 μ m.

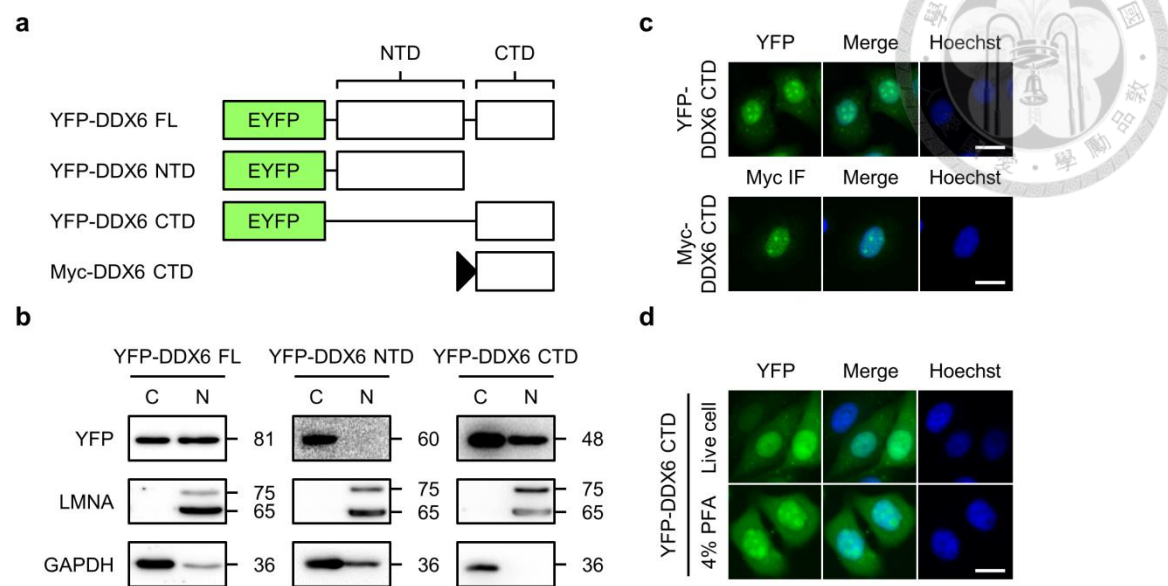


Figure 15. The DDX6 CTD localizes to the cell nuclei.

a, The diagram depicting the DDX6 constructs used in this figure. **b**, The representative WB showing detection of YFP-DDX6 constructs in the HeLa cytoplasmic and the nuclear extracts. LMNA: nuclear marker; GAPDH: cytoplasmic marker. **c**, The representative images of YFP- and myc-DDX6 CTD. **d**, The representative images of the live and the fixed HeLa cells expressing YFP-DDX6 CTD. **c and d**, YFP and anti-myc IF signals were shown in green. The nuclei were stained with Hoechst (shown in blue). Scale bars = 20 μ m.

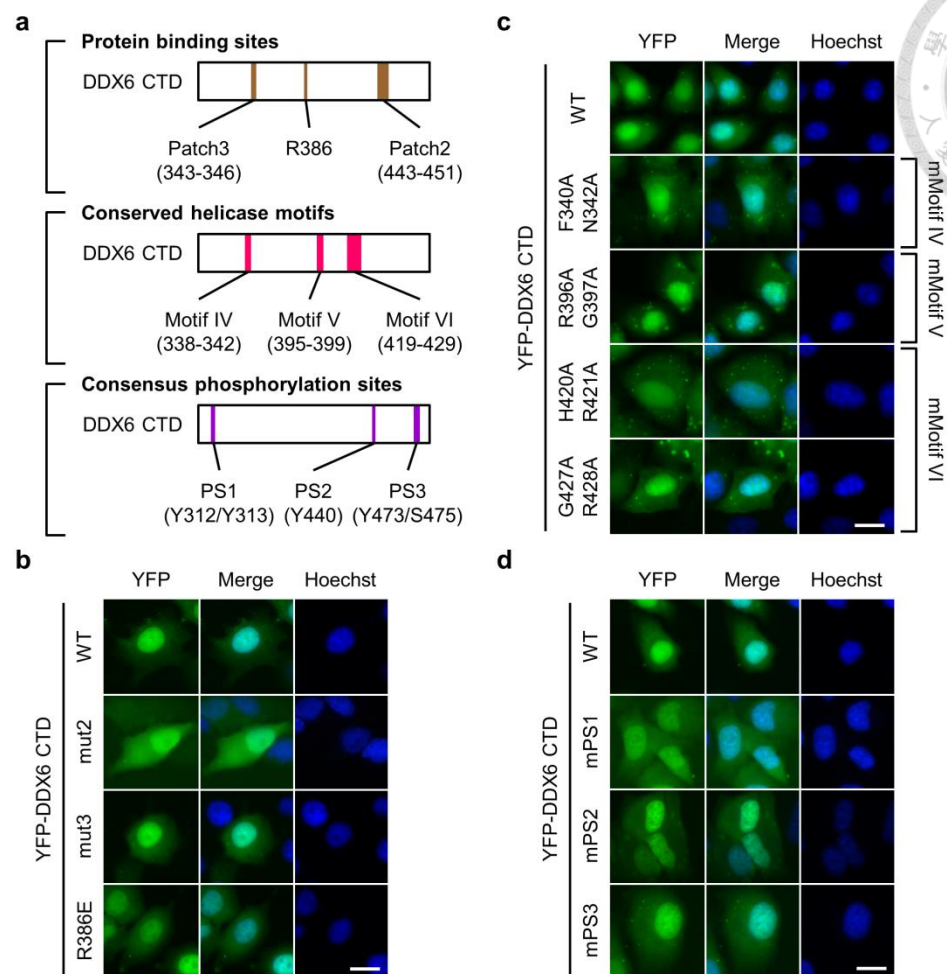


Figure 16. The nuclear localization of YFP-DDX6 CTD is independent of the conserved functional motifs and the phosphorylation sites.

a, The diagram depicting the mutations in the helicase motifs, protein-binding sites, and consensus phosphorylation sites. **b-d**, The representative images of the YFP-DDX6 CTD mutants carrying the mutations in the helicase motifs (**b**), the protein binding sites (**c**), and the phosphorylation sites (**d**). The YFP signals were shown in green. The nuclei were stained with Hoechst and were shown in blue. Scale bars = 20 μm.

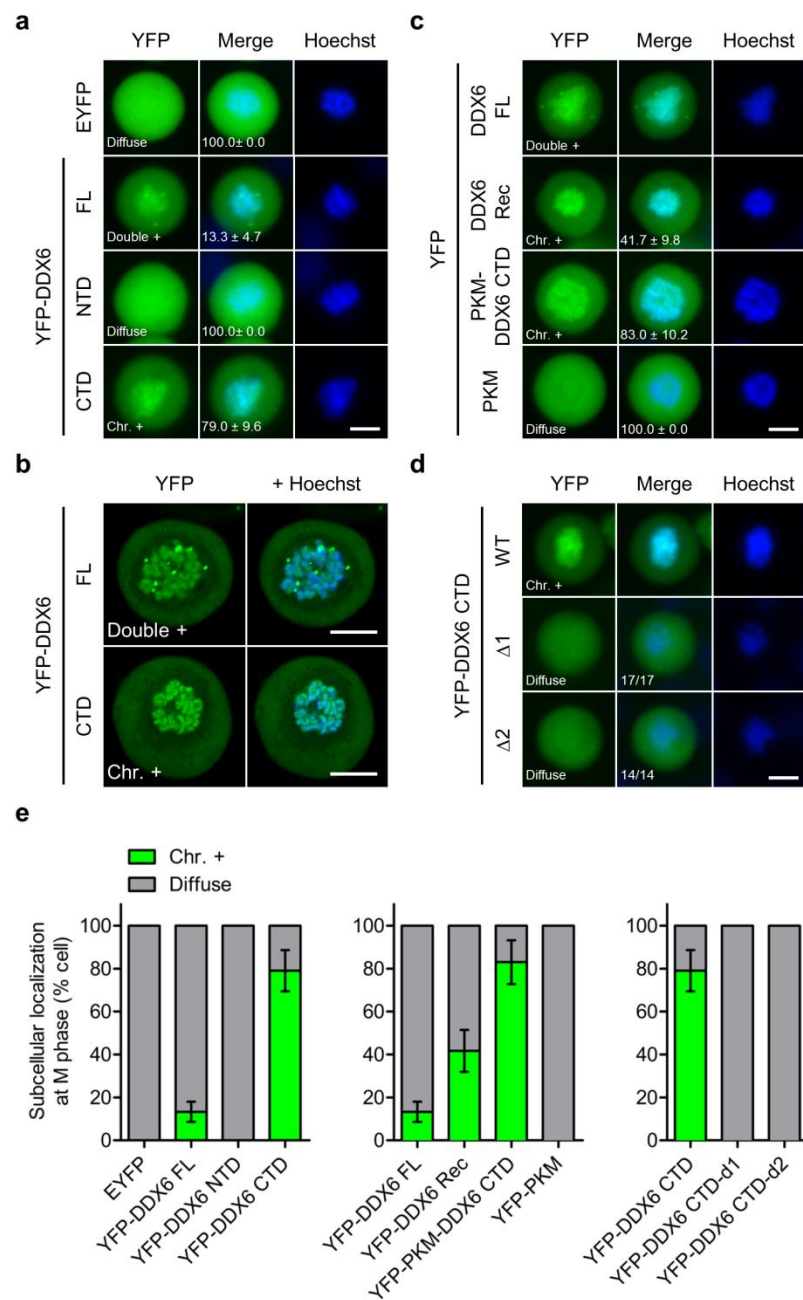


Figure 17. The CTD of DDX6 can bind to the condensed chromosomes during M phase.

a-d, The representative images of the HeLa cells expressing different constructs during

M phase. The HeLa cells expressing DDX6 domains (**a and b**), chimeric proteins (**c**), and CTD truncation mutants (**d**). The HeLa cells were synchronized at prometaphase using nocodazole (50 $\mu\text{g/mL}$). The images in **b** were taken under confocal microscope. Chr. +: chromosome positive; double +: chromosome and granule double-positive; diffuse: diffusive pattern. The numbers represent mean $\pm$ standard deviation of the shown phenotype calculated from the three biological replicates (**a and c**). The fractions represent the frequencies of the observed phenotypes over the total counted cells from the three biological replicates (**d**). The YFP signals were shown in green. The chromatin was stained with Hoechst and was shown in blue. Scale bars = 10 μm . **e**, Bar graph showing the phenotype statistics. Data was calculated from three biological replicates as shown in **a-d**.

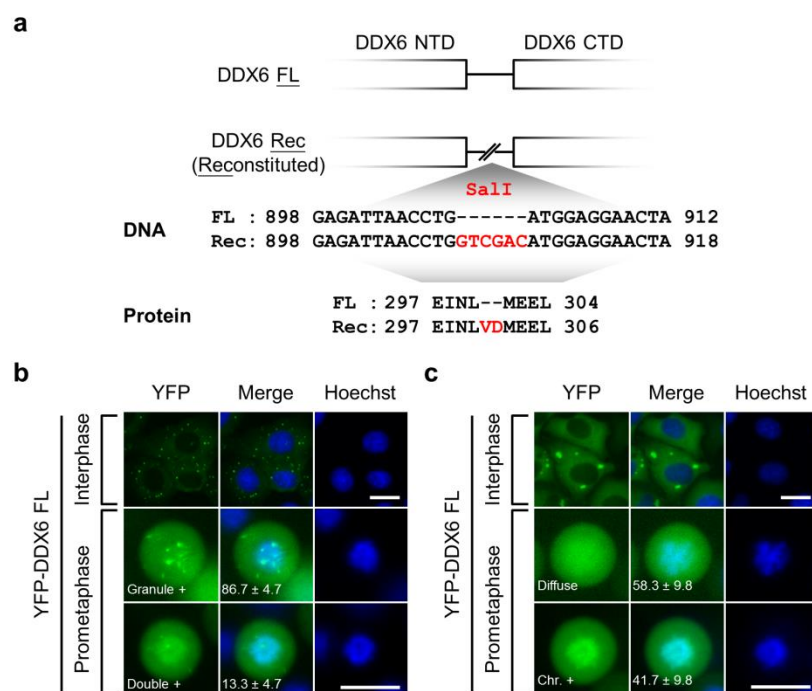


Figure 18. The comparison between YFP-DDX6 FL and YFP-DDX6 Rec

a, The diagram showing the partial DNA and protein sequence alignments between DDX6 FL (WT) and DDX6 Rec. DDX6 Rec contains an additional SalI site in the hinge region. **b and c**, The representative images of YFP-DDX6 FL (**b**) and YFP-DDX6 Rec (**c**) in interphase and mitotic cells. Granule +: granule-positive; Chr. +: chromosome-positive; double +: granule and chromosome double positive. The YFP signals were shown in green, whereas the Hoechst signals, representing nuclei or chromosomes, were shown in blue. Scale bars = 20 μ m.

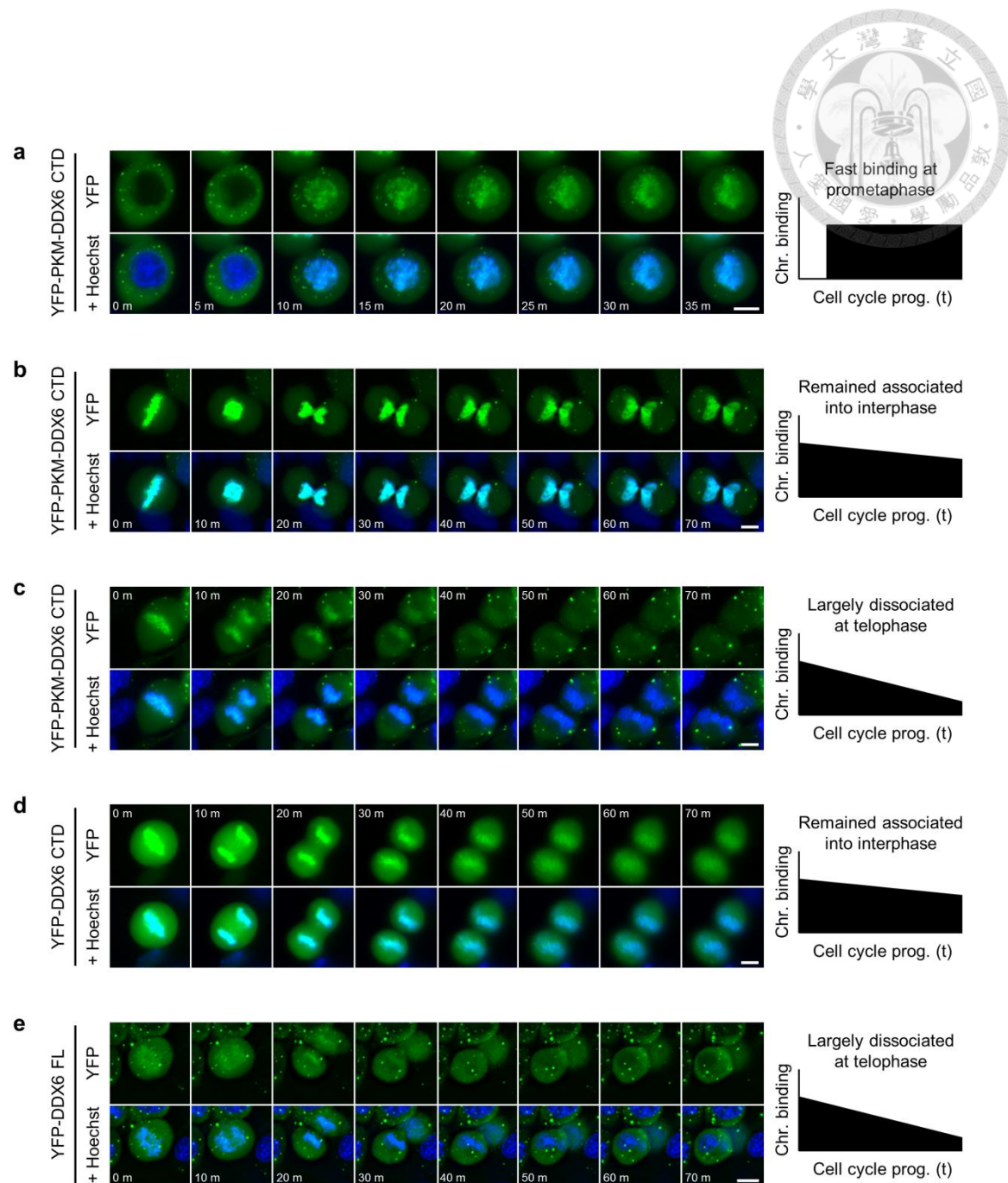
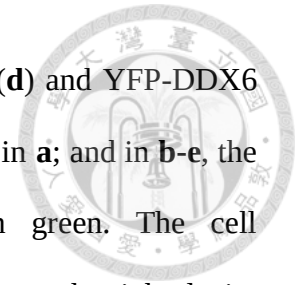


Figure 19. The CTD of DDX6 facilitates nuclear entry via hitchhiking on the mitotic chromosomes.

a-e, The representative images of the time-lapse imaging of the YFP-DDX6 constructs. The HeLa cells expressing YFP-PKM-DDX6 CTD entering (**a**) or exiting M phase (**b**)

and c) were shown. The HeLa cells expressing YFP-DDX6 CTD (**d**) and YFP-DDX6 FL (**e**) exiting M phase were also shown. The time interval = 5 min in **a**; and in **b-e**, the time interval = 10 min. The YFP signals were shown in green. The cell nuclei/chromosomes were stained with Hoechst (blue). The diagrams on the right depict the trend in chromosome binding during the cell cycle progression.



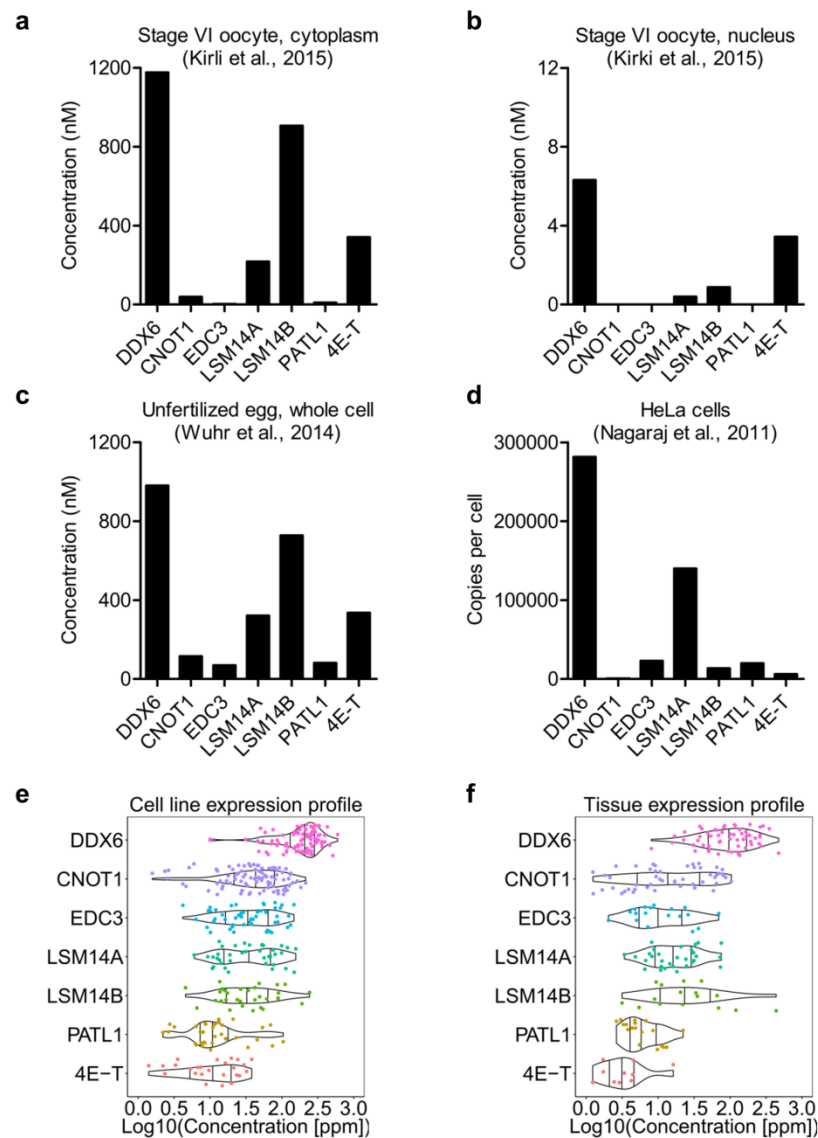


Figure 20. DDX6 generally outnumbers its shuttling protein partners in abundance.

a-f, The bar graphs and the violin plots showing the protein stoichiometry of the DDX6 homologs with their protein partners. The quantitative MS results of the stage VI *Xenopus* oocyte cytoplasm (**a**), the nuclei (**b**), the unfertilized eggs (**c**), the HeLa cells (**d**), the various human cell lines (**e**), and various human tissues (**f**) were plotted.

CNOT1, EDC3, LSM14A are not transported by the CRM1 pathway per se. In contrast, 4E-T, PATL1, and LSM14B are known cargoes for the CRM1 pathway (Fig. 2 and Kirli et al., 2015). The quantitative MS data were collected from Kirli et al., 2015 (**a** and **b**), Wuhr et al., 2014 (**c**), Nagaraj et al., 2011 (**d**), and the MOPED database (**e** and **f**).

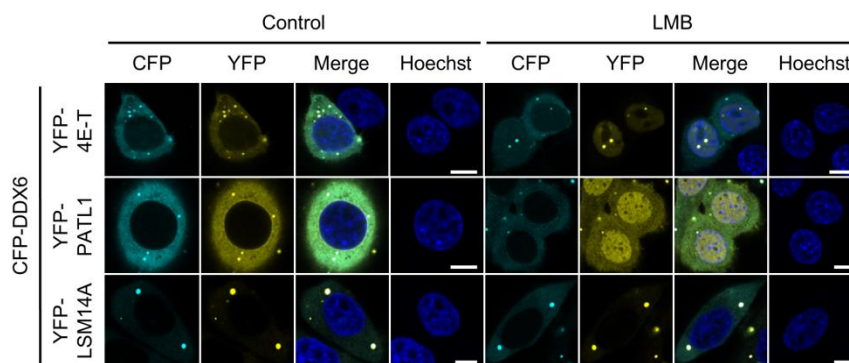


Figure 21. DDX6 can be shuttled by 4E-T in a piggyback manner.

The representative images of CFP-DDX6 co-expressed with YFP-4E-T, YFP-PATL1, or YFP-LSM14A under the LMB treatment. YFP-4E-T, -PATL1, and -LSM14A OE elevated the total abundance of these three proteins closer to the total abundance of DDX6. The transfected HeLa cells were treated with LMB (5 ng/mL) or solvent control for 5 hr prior to imaging. Under the LMB treatment, the strong accumulation in the nuclear YFP-4E-T granules and the mild elevation of the nuclear CFP-DDX6 level were observed. Notably, although YFP-PATL1 accumulated under LMB treatment, granular aggregation within the nuclei was rarely observed as compared to YFP-PATL1 transfection alone (Fig. 2). Nuclei were stained with Hoechst. Scale bar = 20 μ m.

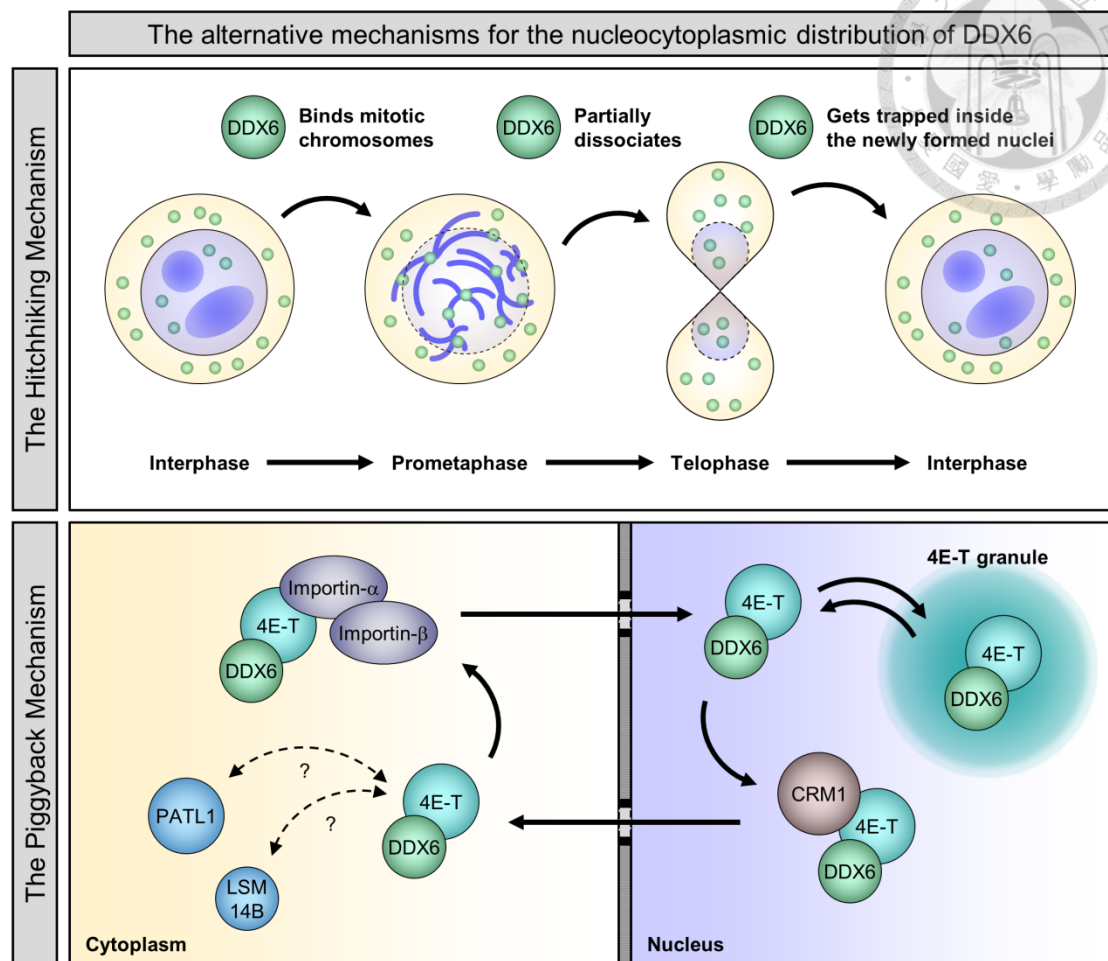
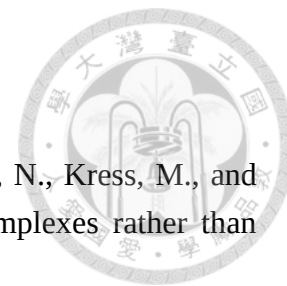


Figure 22. A new synthesis for the nucleocytoplasmic distribution mechanism of DDX6

The diagram summarized the two mechanisms governing the nucleocytoplasmic distribution mechanism of DDX6. The “hitchhiking” model summarized the finding that DDX6 can enter cell nuclei as a passenger on the mitotic chromosomes (the upper panel). The “piggyback” model summarized finding that DDX6 can be co-transported with 4E-T by the importin- α/β and the CRM1 pathways in a piggyback manner (the lower panel).



7. References

Ayache, J., Benard, M., Ernoult-Lange, M., Minshall, N., Standart, N., Kress, M., and Weil, D. (2015). P-body assembly requires DDX6 repression complexes rather than decay or Ataxin2/2L complexes. *Mol. Biol. Cell* 26, 2579-2595.

Brameier, M., Krings, A., and MacCallum, R.M. (2007). NucPred--predicting nuclear localization of proteins. *Bioinformatics* 23, 1159-1160.

Chu, C.Y., and Rana, T.M. (2006). Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. *PLoS Biol.* 4, e210.

Coller, J., and Parker, R. (2005). General translational repression by activators of mRNA decapping. *Cell* 122, 875-886.

Coller, J.M., Tucker, M., Sheth, U., Valencia-Sanchez, M.A., and Parker, R. (2001). The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes. *RNA* 7, 1717-1727.

de Vries, S., Naarmann-de Vries, I.S., Urlaub, H., Lue, H., Bernhagen, J., Ostareck, D.H., and Ostareck-Lederer, A. (2013). Identification of DEAD-box RNA helicase 6 (DDX6) as a cellular modulator of vascular endothelial growth factor expression under hypoxia. *J. Biol. Chem.* 288, 5815-5827.

Dong, X., Biswas, A., Suel, K.E., Jackson, L.K., Martinez, R., Gu, H., and Chook, Y.M. (2009). Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature* 458, 1136-1141.

Dostie, J., Ferraiuolo, M., Pause, A., Adam, S.A., and Sonenberg, N. (2000). A novel shuttling protein, 4E-T, mediates the nuclear import of the mRNA 5' cap-binding protein, eIF4E. *EMBO J.* 19, 3142-3156.

Fan, X.C., and Steitz, J.A. (1998). HNS, a nuclear-cytoplasmic shuttling sequence in HuR. *Proceedings of the National Academy of Sciences of the United States of America* 95, 15293-15298.

Gagnon, K.T., Li, L., Chu, Y., Janowski, B.A., and Corey, D.R. (2014a). RNAi factors are present and active in human cell nuclei. *Cell Rep.* 6, 211-221.

Gagnon, K.T., Li, L., Janowski, B.A., and Corey, D.R. (2014b). Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. *Nat. Protoc.* 9, 2045-2060.

Guttler, T., Madl, T., Neumann, P., Deichsel, D., Corsini, L., Monecke, T., Ficner, R., Sattler, M., and Gorlich, D. (2010). NES consensus redefined by structures of PKI-type and Rev-type nuclear export signals bound to CRM1. *Nat. Struct. Mol. Biol.* 17, 1367-1376.

Haimovich, G., Medina, D.A., Causse, S.Z., Garber, M., Millan-Zambrano, G., Barkai, O., Chavez, S., Perez-Ortin, J.E., Darzacq, X., and Choder, M. (2013). Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell* 153, 1000-1011.

Hantschel, O., Wiesner, S., Guttler, T., Mackereth, C.D., Rix, L.L., Mikes, Z., Dehne, J., Gorlich, D., Sattler, M., and Superti-Furga, G. (2005). Structural basis for the cytoskeletal association of Bcr-Abl/c-Abl. *Mol. Cell* 19, 461-473.

Iio, A., Takagi, T., Miki, K., Naoe, T., Nakayama, A., and Akao, Y. (2013). DDX6 post-transcriptionally down-regulates miR-143/145 expression through host gene NCR143/145 in cancer cells. *Biochim. Biophys. Acta* 1829, 1102-1110.

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

Kataoka, N., Diem, M.D., Yoshida, M., Hatai, C., Dobashi, I., Dreyfuss, G., Hagiwara, M., and Ohno, M. (2011). Specific Y14 domains mediate its nucleo-cytoplasmic shuttling and association with spliced mRNA. *Sci. Rep.* 1, 92.

Kataoka, N., Dobashi, I., Hagiwara, M., and Ohno, M. (2013). hDbr1 is a nucleocytoplasmic shuttling protein with a protein phosphatase-like motif essential for debranching activity. *Sci. Rep.* 3, 1090.

Kawahara, C., Yokota, S., and Fujita, H. (2014). DDX6 localizes to nuage structures and the annulus of mammalian spermatogenic cells. *Histochem. Cell Biol.* 141, 111-121.

Kirli, K., Karaca, S., Dehne, H.J., Samwer, M., Pan, K.T., Lenz, C., Urlaub, H., and Gorlich, D. (2015). A deep proteomics perspective on CRM1-mediated nuclear export and nucleocytoplasmic partitioning. *eLife* 4.

Kolker, E., Higdon, R., Haynes, W., Welch, D., Broomall, W., Lancet, D., Stanberry, L., and Kolker, N. (2012). MOPED: Model Organism Protein Expression Database. *Nucleic Acids Res.* 40, D1093-1099.

Kruk, J.A., Dutta, A., Fu, J., Gilmour, D.S., and Reese, J.C. (2011). The multifunctional Ccr4-Not complex directly promotes transcription elongation. *Genes Dev.* 25, 581-593.

Kuzuoglu-Ozturk, D., Bhandari, D., Huntzinger, E., Fauser, M., Helms, S., and Izaurralde, E. (2016). miRISC and the CCR4-NOT complex silence mRNA targets independently of 43S ribosomal scanning. *EMBO J.* 35, 1186-1203.

Ladomery, M., and Sommerville, J. (2015). The Scd6/Lsm14 protein xRAPB has properties different from RAP55 in selecting mRNA for early translation or intracellular distribution in *Xenopus* oocytes. *Biochim. Biophys. Acta* 1849, 1363-1373.

Ladomery, M., Wade, E., and Sommerville, J. (1997). Xp54, the *Xenopus* homologue of human RNA helicase p54, is an integral component of stored mRNP particles in oocytes. *Nucleic Acids Res.* 25, 965-973.

Marfori, M., Lonhienne, T.G., Forwood, J.K., and Kobe, B. (2012). Structural basis of high-affinity nuclear localization signal interactions with importin- α . *Traffic* 13, 532-548.

Marnef, A., Weil, D., and Standart, N. (2012). RNA-related nuclear functions of human Pat1b, the P-body mRNA decay factor. *Mol. Biol. Cell* 23, 213-224.

Mathys, H., Basquin, J., Ozgur, S., Czarnocki-Cieciura, M., Bonneau, F., Aartse, A., Dziembowski, A., Nowotny, M., Conti, E., and Filipowicz, W. (2014). Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. *Mol. Cell* 54, 751-765.

Matsui, T., Hogetsu, K., Usukura, J., Sato, T., Kumasaka, T., Akao, Y., and Tanaka, N. (2006). Structural insight of human DEAD-box protein rck/p54 into its substrate recognition with conformational changes. *Genes Cells* 11, 439-452.

Michael, W.M., Eder, P.S., and Dreyfuss, G. (1997). The K nuclear shuttling domain: a novel signal for nuclear import and nuclear export in the hnRNP K protein. *EMBO J.* 16, 3587-3598.

Minshall, N., Thom, G., and Standart, N. (2001). A conserved role of a DEAD box helicase in mRNA masking. *RNA* 7, 1728-1742.

Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Paabo, S., and Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* 7, 548.

Nishi, K., Nishi, A., Nagasawa, T., and Ui-Tei, K. (2013). Human TNRC6A is an Argonaute-navigator protein for microRNA-mediated gene silencing in the nucleus. *RNA* 19, 17-35.

Ohta, S., Bukowski-Wills, J.C., Sanchez-Pulido, L., Alves Fde, L., Wood, L., Chen, Z.A., Platani, M., Fischer, L., Hudson, D.F., Ponting, C.P., *et al.* (2010). The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. *Cell* 142, 810-821.

Ozgun, S., Basquin, J., Kamenska, A., Filipowicz, W., Standart, N., and Conti, E. (2015). Structure of a Human 4E-T/DDX6/CNOT1 Complex Reveals the Different Interplay of DDX6-Binding Proteins with the CCR4-NOT Complex. *Cell Rep.* 13, 703-711.

Pettersson, O.J., Aagaard, L., Andrejeva, D., Thomsen, R., Jensen, T.G., and Damgaard, C.K. (2014). DDX6 regulates sequestered nuclear CUG-expanded DMPK-mRNA in dystrophin myotonia type 1. *Nucleic Acids Res.* 42, 7186-7200.

Saito, K., Kondo, E., and Matsushita, M. (2011). MicroRNA 130 family regulates the hypoxia response signal through the P-body protein DDX6. *Nucleic Acids Res.* 39, 6086-6099.

Seibel, N.M., Eljouni, J., Nalaskowski, M.M., and Hampe, W. (2007). Nuclear localization of enhanced green fluorescent protein homomultimers. *Anal. Biochem.* 368, 95-99.

Sharif, H., Ozgun, S., Sharma, K., Basquin, C., Urlaub, H., and Conti, E. (2013). Structural analysis of the yeast Dhh1-Pat1 complex reveals how Dhh1 engages Pat1,

Edc3 and RNA in mutually exclusive interactions. *Nucleic Acids Res.* 41, 8377-8390.

Siomi, H., and Dreyfuss, G. (1995). A nuclear localization domain in the hnRNP A1 protein. *J. Cell Biol.* 129, 551-560.

Smillie, D.A., and Sommerville, J. (2002). RNA helicase p54 (DDX6) is a shuttling protein involved in nuclear assembly of stored mRNP particles. *J. Cell Sci.* 115, 395-407.

Torrente, M.P., Zee, B.M., Young, N.L., Baliban, R.C., LeRoy, G., Floudas, C.A., Hake, S.B., and Garcia, B.A. (2011). Proteomic interrogation of human chromatin. *PLoS ONE* 6, e24747.

Tritschler, F., Braun, J.E., Eulalio, A., Truffault, V., Izaurralde, E., and Weichenrieder, O. (2009). Structural basis for the mutually exclusive anchoring of P body components EDC3 and Tral to the DEAD box protein DDX6/Me31B. *Mol. Cell* 33, 661-668.

Tritschler, F., Eulalio, A., Helms, S., Schmidt, S., Coles, M., Weichenrieder, O., Izaurralde, E., and Truffault, V. (2008). Similar modes of interaction enable Trailer Hitch and EDC3 to associate with DCP1 and Me31B in distinct protein complexes. *Mol. Cell Biol.* 28, 6695-6708.

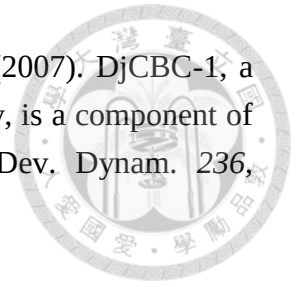
Ward, A.M., Bidet, K., Yinglin, A., Ler, S.G., Hogue, K., Blackstock, W., Gunaratne, J., and Garcia-Blanco, M.A. (2011). Quantitative mass spectrometry of DENV-2 RNA-interacting proteins reveals that the DEAD-box RNA helicase DDX6 binds the DB1 and DB2 3' UTR structures. *RNA Biol.* 8, 1173-1186.

Weston, A., and Sommerville, J. (2006). Xp54 and related (DDX6-like) RNA helicases: roles in messenger RNP assembly, translation regulation and RNA degradation. *Nucleic Acids Res.* 34, 3082-3094.

Wuhr, M., Freeman, R.M., Jr., Presler, M., Horb, M.E., Peshkin, L., Gygi, S.P., and Kirschner, M.W. (2014). Deep proteomics of the *Xenopus laevis* egg using an mRNA-derived reference database. *Curr. Biol.* 24, 1467-1475.

Yoon, J.H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., Huarte, M., Zhan, M., Becker, K.G., and Gorospe, M. (2012). LincRNA-p21 suppresses target mRNA translation. *Mol. Cell* 47, 648-655.

Yoshida-Kashikawa, M., Shibata, N., Takechi, K., and Agata, K. (2007). DjCBC-1, a conserved DEAD box RNA helicase of the RCK/p54/Me31B family, is a component of RNA-protein complexes in planarian stem cells and neurons. *Dev. Dynam.* 236, 3436-3450.





Chapter 3

On the Roles of DEAD-box Protein 6
in Post-transcriptional Regulation

An Integrative, Functional Genomics Study



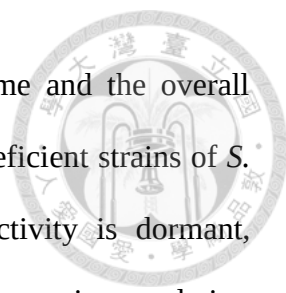


1. **Abstract**

DDX6 is an evolutionarily and functionally conserved protein. Early studies have revealed the functional roles of DDX6 in facilitating RNA turnover and translational repression. Furthermore, DDX6 has been implicated in a wide range of physiological processes. One fundamental commonality between these physiological processes is the demand of precise gene expression. As such, DDX6 is hypothesized to contribute to these physiological processes through controlling gene expression. In this chapter, we integrated datasets from different high-throughput profiling experiments for a comprehensive survey of the DDX6's RNA targets and its impact on the overall gene expression. We discovered that, instead of strongly destabilizing or inhibiting the translation of certain targets, DDX6 acts to reduce the gene expression noise on both the RNA and the protein levels.

2. **Introduction**

The DDX6 subfamily proteins are evolutionarily and functionally conserved among the eukaryotes. The most well-established functions of the DDX6 homologs are promoting RNA decay and repressing translation (Chu and Rana, 2006; Collier and Parker, 2005; Collier et al., 2001; Minshall et al., 2001). Early studies in *S. cerevisiae* have displayed that Dhh1p is required for efficient RNA turnover through decapping (Collier et al., 2001; Fischer and Weis, 2002). For this reason, the DDX6 homologs are sometimes referred to as the “activator of decapping”. More recently, it has been demonstrated that artificially tethering Dhh1p to reporter mRNA destabilize the reporter mRNA (Carroll et al., 2011; Sweet et al., 2012). These results indicate that the DDX6 homologs facilitate the decapping of its direct binding targets. On the other hand, the



overexpression (OE) of Dhh1p can drastically reduce the polysome and the overall protein synthesis in yeast (Coller and Parker, 2005). In decapping-deficient strains of *S. cerevisiae* and in the *Xenopus* oocytes, where the decapping activity is dormant, artificially tethering Dhh1p and Xp54 to the reporter mRNA can repress its translation regardless of the absence of the active decapping machinery (Carroll et al., 2011; Minshall et al., 2001). Furthermore, Dhh1p, Xp54, and DDX6 have been shown to be capable of repressing translation in vitro (Coller and Parker, 2005; Tanaka et al., 2006). These results indicate the translational repression mediated by the DDX6 homologs represents an independent mechanism from the facilitation of RNA decapping. Altogether, these established DDX6 as a potent post-transcriptional regulator. Indeed, the loss of DDX6's activity has widespread physiological consequences. For example, the lethality of *S. cerevisiae* in response to stresses (Bergkessel and Reese, 2004; Carroll et al., 2011; Cheng et al., 2005; Dutta et al., 2011; Westmoreland et al., 2003), larval lethality or aberrant development in *Drosophila* (Barbee et al., 2006; Nakamura et al., 2001), aberrant embryonic development and gametogenesis in *C. elegans* (Audhya et al., 2005; Navarro et al., 2001), disrupted regeneration in *D. japonica* (Rouhana et al., 2010), and mis-specification of cell fates in mammalian progenitor cells (Nicklas et al., 2015; Wang et al., 2015). Since these biological processes all command the changes in the gene expression, it is natural to hypothesize the DDX6 homologs play important roles in these processes through regulating gene expression. Furthermore, since the mammalian genomes harbor tens of thousands of the long noncoding RNA (lncRNA) genes (Mercer and Mattick, 2013; Rinn and Chang, 2012) that have widespread implications in the development and physiology (Sauvageau et al., 2013), it is possible that DDX6 also participates in the physiological and developmental regulation through

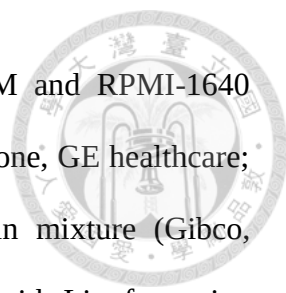
the lncRNA. Supporting this notion, we (Chap. 2) and others (Iio et al., 2013; Yoon et al., 2012) have provided evidence that DDX6 can bind the lncRNA.

Entering the era of the genomics, we now command powerful high-throughput technologies to assess molecular interaction and gene expression on the genome-wide scale. So far, the protein-RNA interaction of Dhh1p has been assessed with cross-linking followed by immunoprecipitation and deep sequencing (CLIP-seq) (Mitchell et al., 2013), and the impact of DDX6 on gene expression during epithelial differentiation has been assessed with microarray (Nicklas et al., 2015). However, DDX6 act through as least two distinct mechanisms, namely, translational repression and RNA degradation, to silence protein production. Consequently, a more comprehensive analysis is required to distinguish between the effects of the different mechanisms. To this end, we conducted an integrative, genome-wide assessment of the DDX6's regulatory effect in human cell model by integrating RNA co-immunoprecipitation followed by deep sequencing (RIP-seq), deep sequencing of cDNA (RNA-seq), and quantitative mass spectrometry (MS) datasets. Furthermore, we drew from the conceptual analogy between DDX6 and the miRNA to explore the commonality between the two promiscuous RNA regulators.

3. Materials and methods

3.1. Cell culture and siRNA transfection

Cell culture and siRNA transfection were performed as described in Chap. 2. Briefly, both the cervical cancer cell line HeLa and the lung cancer cell line CL1-5 were cultivated under controlled condition (37°C, 5% CO₂, with moistened air). The HeLa cells were supplied with DMEM medium (GE healthcare) whereas the CL1-5 cells were



supplied with RPMI-1640 medium (GE healthcare). Both DMEM and RPMI-1640 media were supplemented with 10% (v/v) fetal bovine serum (HyClone, GE healthcare; Gibco, Thermo Fisher Scientific), and 1% penicillin-streptomycin mixture (Gibco, Thermo Fisher Scientific). Transfection of siRNA was carried out with Lipofectamine 2000 reagent (Thermo Fisher Scientific), following the supplier's protocol, except for reducing the Lipofectamine 2000 reagent dosage to one fourth to decrease cell toxicity.

3.2. RNA co-immunoprecipitation (RIP)

Anti-DDX6 RIP was performed as described in Chap. 2. Briefly, cell lysate was incubated with Dynabeads M-280 sheep anti-rabbit IgG (Thermo Fisher Scientific) and with either control normal rabbit IgG (sc-2027, Santa Cruz) or anti-DDX6 antibody (A300-460A, Bethyl Laboratories) at 4°C for 3 hours. RNA and protein co-precipitated with the Dynabeads were extracted with TRIzol reagent (Thermo Fisher Scientific) or eluted with heating and 4× SDS sample buffer (200 mM Tris-HCl pH 6.8, 400 mM dithiothreitol [DTT], 8% [w/v] sodium dodecyl sulfate [SDS], 0.4% [w/v] bromophenol blue, 40% [v/v] glycerol), respectively.

3.3. RNA extraction, reverse transcription, and quantitative real-time PCR (qPCR)

The workflow for analyzing RNA expression level or abundance were the same as described in Chap. 2. Briefly, RNA samples were extracted from whole cells and Dynabeads with TRIzol reagent, whereas RNA samples were extracted from polysome fractions with TRIzol LS reagent (Thermo Fisher Scientific). RNA extraction was performed following the supplier's protocol, except for the addition of glycogen (molecular biology grade, Roche) prior to adding isopropanol to facilitate RNA

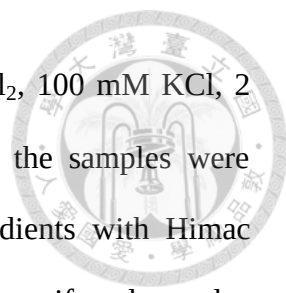
precipitation for the RIP and the polysome fraction samples. Reverse transcription was performed with the SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific) using random hexamer priming. The resulting cDNA were analyzed with qPCR using the iQ SYBR Green Supermix (Bio-Rad) and gene-specific primer sets on the CFX Connect Real-Time PCR Detection system (BioRad).

3.4. Protein extraction and Western blot (WB)

The workflow for analyzing protein expression level or abundance were the same as described in Chap. 2. Briefly, cell lysate prepared from RIPA lysis buffer or NE-PER reagent and protein samples eluted from Dynabeads were all separated by SDS-PAGE, transferred to PVDF membranes, and probed with antibodies. The antibodies used for WB experiments in this chapter are listed as followed: anti-DDX6 (A300-460A, Bethyl Laboratories), anti-GAPDH (GTX100118, GeneTex), goat anti-rabbit IgG-HRP (sc-2004, Santa Cruz), and Rabbit Trueblot anti-rabbit IgG-HRP (18-8816-33, Rockland).

3.5. Polysome fractionation and profiling

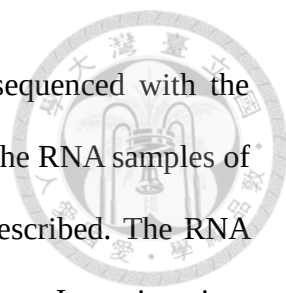
Both the control and the DDX6 knockdown (KD) HeLa cells ($\approx 10^7$ cells) were treated with cycloheximide (CHX, 100 $\mu\text{g}/\text{mL}$; Amresco) at 37°C for 15 min prior to cell lysis. From this step on, all steps were performed in the presence of CHX at 100 $\mu\text{g}/\text{mL}$ concentration. The cells were then collected by scraping and lysed by gentle shearing with 26.5 gauge needles in the polysome extraction buffer (10 mM Tris-HCl pH 7.4, 5 mM MgCl_2 , 100 mM KCl, 1% Triton X-100, 2 mM DTT, 1 $\times$ protease inhibitor, 100 $\mu\text{g}/\text{mL}$ CHX). Cell debris was then removed by centrifugation (at 1,300 $\times g$, 4°C for 10 min). We used equal amount of each sample (rRNA = 8 OD). We used



10% ~ 50% sucrose gradients (20 mM HEPES-KOH, 5 mM MgCl₂, 100 mM KCl, 2 mM DTT, 100 µg/mL CHX, 10% or 50% [v/v] sucrose). Next, the samples were centrifuged at 36,000 rpm, 4°C for 120 min on the sucrose gradients with Himac CP80WX ultracentrifuge (Hitachi) and P40ST rotor (Hitachi). The centrifuged samples were then slowly pumped out for UV (260 nm) absorbance recording and fraction collection. For each sample, we collected 12 fractions. The resulting fractions were mixed with TRIzol LS reagent for the subsequent analysis of the RNA content in each fraction.

3.6. RIP-seq and RNA-seq sample preparation and data acquisition

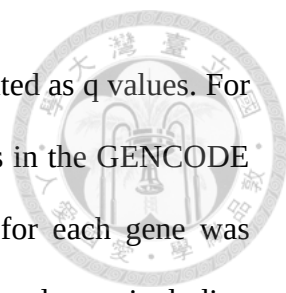
The HeLa anti-DDX6 RIP-seq dataset and the HeLa poly(A)⁺ RNA-seq dataset were previously generated by C. Y. Chu and Y. C. Chen, as described in Y. C. Chen's master thesis. In brief, RNA samples isolated from the anti-DDX6 RIP experiment was sequenced by the SOLiD 5500xl sequencer (Applied Biosystems, Thermo Fisher Scientific), which generated 75 nt single-end reads. On the other hand, poly(A)⁺ RNA extracted from the control and the DDX6 KD cells were sequenced with the MiSeq sequencer (Illumina), which generated 2×151 nt paired-end reads. Both the HeLa anti-DDX6 RIP-seq samples and the HeLa poly(A)⁺ RNA-seq samples were cleared of rRNA by poly(A)-selection. The CL1-5 anti-DDX6 RIP-seq samples and the HeLa total RNA-seq samples were generated by J. H. Huang and were described as followed. The anti-DDX6 RIP in CL1-5 cells and the subsequent RNA extraction were conducted as described in the previous section. The RIP efficacy in both replicates was verified with the Agilent 2100 bioanalyzer (Agilent technologies). The RNA isolated from the two biological replicates were pooled together to match up the input requirement for the



sequencing library construction. Subsequently, the samples were sequenced with the MiSeq sequencer, resulting in 2×151 nt paired-end reads. Similarly, the RNA samples of the control and the DDX6 KD HeLa cells were also prepared as described. The RNA integrity of these samples was assessed by the Agilent 2100 bioanalyzer. In conjunction, the DDX6 KD efficacy was assessed with qPCR and WB. The samples were then sequenced with the MiSeq sequencer, resulting in 2×76 nt paired-end reads. Both the CL1-5 RIP-seq samples and the HeLa total RNA-seq samples were cleared of rRNA using the Ribo-Zero removal kit (Epicentre). The HeLa-S3 nuclear and cytoplasmic RNA-seq dataset, containing two poly(A)-selected replicates for each fraction, was retrieved from the ENCODE RNA dashboard (Encode Project Consortium, 2012). All sequence data in this dataset were 2×76 nt paired-end reads. The miR-122 KD expression dataset were retrieved from the supplementary table 1 of the original study. We analyzed the expression profile as it is in the retrieved data table without handling the raw sequence data.

3.7. RIP-seq and RNA-seq data analysis

We used the cutadapt program (Martin, 2011) to remove the adaptor sequences in the datasets generated by our group. All RIP-seq and RNA-seq reads were mapped to reference human genome (hg19) with the TopHat2 software (Kim et al., 2013). For the HeLa-S3 nuclear and cytoplasmic RNA-seq dataset, we were not confident about the adaptor sequences, therefore the reads were mapped to the reference genome as it is. For CL1-5 RIP-seq dataset, only the forward reads (R1) were mapped to the reference genome. The RNA abundance and differential testing were carried out with the Cuffdiff2 software (Trapnell et al., 2013). The P values resulting from multiple testing

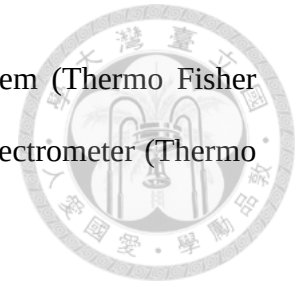


were corrected with Benjamini-Hochberg procedure and were presented as q values. For all datasets, the human genome was annotated with the gene models in the GENCODE annotation (version 15) (Harrow et al., 2012). The gene biotype for each gene was retrieved from the same annotation file. Downstream statistical analyses, including correlation analysis, principal component analysis (PCA), and integrative dataset comparisons, were performed with the expression table generated by the Cuffdiff2 software using R custom scripts. The miR-122 KD dataset was likewise analyzed using R custom scripts. The decomposition of statistical distributions based on Gaussian mixture model and posterior probabilities were carried out using the “mixtools” package (Benaglia et al., 2009) as previously described (Lu et al., 2014).

3.8. Quantitative MS sample preparation and data acquisition

The HeLa cells were transfected with control siRNA (siDDX6-mm) or targeting siRNA (siDDX6-pm). At 48 hours post-transfection, the transfected HeLa cells were fractionated with NE-PER reagents as described in Chap. 2. Among the five biological replicates, two replicates were prepared by Y. C. Chen and were described in her master thesis, whereas the other three were prepared by J. H. Huang as described herein. The DDX6 KD and the fractionation efficacies were assessed by WB. Both nuclear and cytoplasmic fractions were subjected to quantitative proteomics using dimethyl labeling as previously described (Boersema et al., 2009). Briefly, the proteins were reduced by DTT, alkylated by iodoacetamide, digested by Lys-C and trypsin, and dimethyl-labeled by formaldehyde. The labeled peptides were then fractionated with the strong cation exchange (SCX) StageTip and desalted (Rappsilber et al., 2007). The fractionated peptides were analyzed using the nano-liquid chromatography followed by tandem MS

(nanoLC-MS/MS) on the Dionex Ultimate 3000 RSLC nano system (Thermo Fisher Scientific) that was online-coupled to an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific) as previously described (Hu et al., 2012).



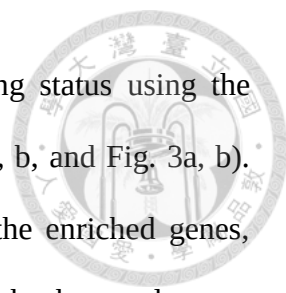
3.9. MS data analysis

Protein identification and quantification was performed by the Andromeda algorithm (Cox et al., 2011) in the MaxQuant software (version 1.5.2.8) (Cox and Mann, 2008) using the Swissprot database (version 2014_11 with 20,226 human sequence entries). The differentially expressed proteins were detected based on the statistical significance B value given by the MaxQuant software. To estimate the relative abundance of each protein, the summed-up extracted ion chromatogram (XIC) of all quantified peptides was used. We performed the downstream statistical analyses were with this expression data using R custom scripts. We first normalized the protein abundance. The XIC intensities were log10-transformed and median-centered to give comparable intensity distributions. The changes in protein expression following DDX6 KD were estimated by the ratio of DDX6 KD over control. The expression ratios were log2-transformed and median-centered to remove the systematic errors. Similar to the RNA-seq data, the statistical analyses, including correlation analysis, PCA, and integrative dataset comparisons, were carried out using R.

4. Results

4.1. DDX6 binds both the mRNA and the lncRNA in human cell lines

To explore the physiological targets of DDX6, we performed anti-DDX6 RIP-seq (RNA co-immunoprecipitation followed by deep sequencing) in two distinct human cell lines, the cervical cancer cell line HeLa (Y. C. Chen's master thesis) and the lung cancer

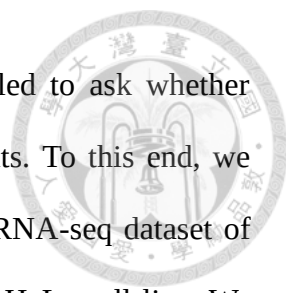


cell line CL1-5 (Fig. 1). We designated genes with binary binding status using the enrichment (anti-DDX6 IP/control IgG) threshold at 2 fold (Fig. 2a, b, and Fig. 3a, b). The genes above the enrichment threshold will be referred to as the enriched genes, whereas the genes below the threshold will be referred to as the background genes hereafter. In the HeLa cells, we obtained 3156 enriched genes, while in the CL1-5 cells we obtained 2879 enriched genes. Similar, yet stricter, results were obtained when we utilized an approach based on the Gaussian mixture model and the posterior probabilities (Fig. 4). In the following sections, we will present the results using the enrichment threshold ≥ 2 fold. We perceived this cutoff to be sufficiently stringent for a promiscuous RNA-binding protein such as DDX6, as in the case of FMRP, the average enrichment (determined by RIP-microarray) of the best FMRP targets (as determined by CLIP-seq and binding motifs) is less than 1.68 fold (Ascano et al., 2012).

In both the HeLa and the CL1-5 cells, the enriched genes included all eight major gene biotypes (more numerous in the human genome), including protein-coding gene biotype and various lncRNA gene types (Fig. 2c and 3c). In the HeLa cells, the protein-coding genes constituted a major fraction ($\approx 51.3\%$) of the enriched genes (Fig. 2d), while in the CL1-5 cells it was the predominant fraction ($\approx 71.6\%$; Fig. 3d). Conversely, the same results also proved that the lncRNA gene types contributed substantially to the enriched genes in both cell lines (Fig. 2d and 3d). Notably, in the HeLa cells, the lncRNA gene biotypes were overrepresented in the enriched genes as compared to the expressed genes in the HeLa cells (Fig. 5).

4.2. DDX6 binds the RNA localized in different subcellular compartments

In the previous chapter, we demonstrated that DDX6 is present in the cell nuclei

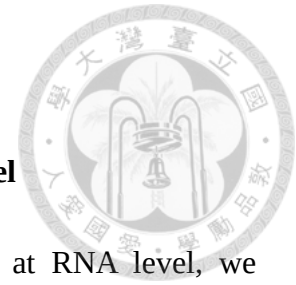


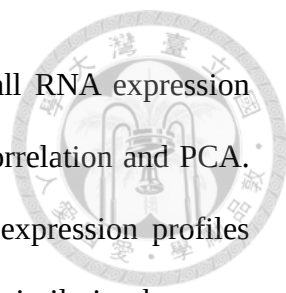
and binds the nuclear lncRNA MALAT1. Thus, we were compelled to ask whether DDX6 can bind RNA enriched in specific subcellular compartments. To this end, we compared the HeLa RIP-seq dataset to a cytoplasmic and nuclear RNA-seq dataset of the HeLa-S3 cell line (Fig. 6a), a subline derived from the parental HeLa cell line. We first utilized the HeLa-S3 cytoplasmic and nuclear RNA-seq dataset to define the RNA subcellular localization of each gene (Fig. 7a). More specifically, the genes were classified into four categories (Fig. 7b and c). Category 1: nuclear (≥ 4 fold enrichment in the nucleus, q value ≤ 0.05); category 2: cytoplasmic (≥ 4 fold enrichment in the cytoplasm, q value ≤ 0.05); category 3: not significantly enriched in either compartment (q values > 0.05 , or < 4 fold enrichment in both compartments); category 4: low abundance/no detection (FPKM < 1 in both compartments). We then used this localization information to annotate the enriched genes in the HeLa RIP-seq dataset. As a result, we observed an minor fraction of the enriched genes were biased to the nuclear or the cytoplasmic fraction (Fig. 6b), despite most of the enriched genes were not significantly enriched in either compartment or could not be determined due to low abundance or no detection (Fig. 6b). In the category 4 (low abundance/no detection), the lncRNA gene biotypes were overrepresented (Fig. 6c). This cautioned the strikingly high percentage of the lncRNA gene biotypes in the enriched genes could be inflated to some degree. Notwithstanding, the lncRNA biotypes were still found in the remaining three categories (Fig. 6c). For the category 1 and the category 3, which together represented the majority of the remaining enriched genes (Fig. 6b), the lncRNA biotypes accounted for $\approx 30\%$ (Fig. 6c), closed to that in the CL1-5 cells (Fig. 3d). In this sense, the lncRNA biotypes still confidently constituted up to 30% of the enriched genes in the HeLa cells. Together, our results supported that DDX6 binds both the

mRNA and the lncRNA in different subcellular compartments.

4.3. DDX6 has limited effect on the overall RNA expression level

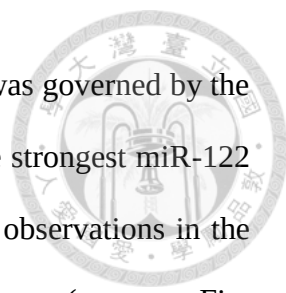
To further assess the impact of DDX6 on gene expression at RNA level, we conducted total RNA-seq experiments on the DDX6 KD HeLa cells. The total RNA-seq quantification covered more 10,000 genes expressed across more than 6 orders of magnitude (Fig. 9a). In this study, we only considered the genes that were expressed at $\text{FPKM} \geq 1$ in both conditions, as we presumed this cutoff represents the lower detectable range for other methods for RNA expression such as qPCR. In our result, very few genes displayed statistically significant elevation in the RNA expression after DDX6 KD (Fig. 9a-c). This contradicted our expectation that a subset of RNA would be upregulated through stabilization upon the loss of DDX6, for DDX6 is known to promote RNA decay. Moreover, among the 14 statistically significant genes, only 4 displayed consistent elevation in RNA abundance following DDX6 KD (Fig. 9b and c). Conversely, very few genes displayed statistically significant decrease in their RNA expression level (Fig. 9a-c), which could result from a CGH-1-like RNA stabilization or from some other secondary effects (Boag et al., 2008). In addition, DDX6 KD did not specifically affect the expression of specific gene biotype (Fig. 10). Very similar results were obtained from an independently conducted poly(A)⁺ RNA-seq experiments (Y. C. Chen master thesis; Fig. 11a, b, and d). The scarce number of the statistically significant genes was a consequence of the conservative P value distribution of the differential expression testing (Fig. 9d and 11c). The conservative P value distribution could result from many potential causes, ranging from the lack of biological difference to the inappropriate statistical assumptions for the differential expression testing (Burden et al.,





2014). To assess the biological impact of DDX6 KD on the overall RNA expression pattern, we explored the total RNA-seq dataset with the pairwise correlation and PCA. The pairwise correlation revealed that the similarity in the RNA expression profiles between different experimental treatments was greater than the similarity between different biological replicates (Fig. 9e). PCA further revealed that $\approx 50\%$ of the total variance was attributed to the PC1 and the PC3, which mostly corresponded to the biological replicates, while only 20% of the total variance was attributed to the PC2, which mostly corresponded to the experimental treatment (Fig. 9f). Together, these results evidenced that the variance within the DDX6 KD total RNA-seq datasets was dominated by the between-replicate variation instead of the between-treatment variation. The between-replicate variation was unlikely to be caused by low DDX6 KD efficacy or RNA degradation since these factors were assessed during sample preparation (Fig. 8).

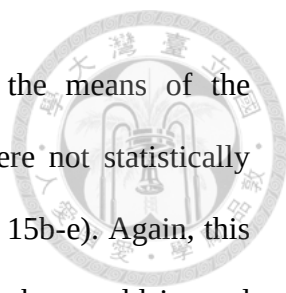
Since DDX6 is in many ways similar to the miRNA, both binds RNA promiscuously and negatively regulates RNA (Fig. 12), we wondered if there are any commonalities in the change of the overall RNA expression upon the loss of DDX6 and miRNA. Previously, the miRNA has been shown to repress hundreds of genes, each to a modest degree, but the effect of miRNA KD on overall RNA expression has not been demonstrated explicitly. To this end, we revisited the miR-122 KD RNA-seq dataset from Denzler et al., *Mol. Cell* (2014), because the miR-122 KD conditions, efficacy, and effects were detailedly characterized and were presented explicitly in this study (Denzler et al., 2014). As described by Denzler et al., the most confident miR-122 targets (33 out of 8909 genes) were mildly stabilized upon miR-122 KD (Fig. 13a and b), averaging a 1.4 fold (geometric mean) increase in the RNA level, while the less confident targets (554 out of 8909 genes) averaged less than 10% upregulation (Denzler



et al., 2014). Moreover, the overall RNA expression pattern, which was governed by the vastly outnumbering non-targets, was barely affected even under the strongest miR-122 KD condition (Fig. 13a, MOI 200 3s/mut). This was similar to our observations in the DDX6 KD total RNA and the DDX6 KD poly(A)+ RNA-seq datasets (compare Fig. 13a with b and c). Furthermore, the pairwise correlation and PCA of the miR-122 KD dataset also yielded similar results to that of the DDX6 KD total RNA-seq dataset (Fig. 14). More specifically, the variance within the miR-122 datasets was dominated by the multiplicity of infection (MOI) of the viral vectors (Fig. 14a and b), which was attributed to the PC1 (Fig. 14c). The variance corresponded to miR-122 KD was majorly attributed the PC5 (Fig. 14c), which only accounted for $\approx 4\%$ of the total variance (Fig. 13b). Together, these results suggested that DDX6 and miRNA alike have limited impact on the overall RNA expression. This finding also supported the conceptual similarities between DDX6 and miRNA.

4.4. DDX6 regulates the expression noise of its target RNA

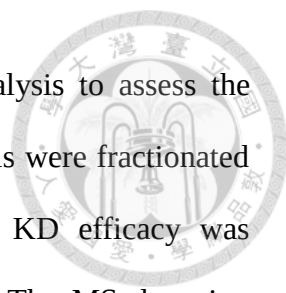
To elucidate the regulatory effect of DDX6 on its target RNA, we integrated the anti-DDX6 RIP-seq dataset to the DDX6 KD total RNA-seq dataset. We compared the log₂-transformed fold enrichment distribution (anti-DDX6 IP/control IgG), abbreviated as log₂-FED, which represented the DDX6 binding strength, and the log₂-transformed RNA expression ratio distribution (DDX6 KD/control), abbreviated as log₂-ERD_{RNA}, which represented the effect size of DDX6 on the RNA expression. As a result, we discovered that DDX6 binding strength and the effect size was uncorrelated (Fig. 15a). This apparent lack of correlation could result from the presence of a large number of irrelevant (not regulated by DDX6) genes. To circumvent this, we stratified genes based



on the binary binding status as described above. As a result, the means of the $\log_2\text{-ERD}_{\text{RNA}}$ in the enriched genes and the background genes were not statistically different, for both protein-coding genes and the lncRNA genes (Fig. 15b-e). Again, this did not rule out the possibility that DDX6 facilitates RNA decay, but could instead indicate the average effect for each target was too weak to be assessed precisely. Nonetheless, the standard deviation of the $\log_2\text{-ERD}_{\text{RNA}}$ were statistically greater in the enriched genes than in the background genes (Fig. 15b-d, and f). Again, we drew analogy from miR-122. The larger variation in the $\log_2\text{-ERD}_{\text{RNA}}$ correlated with the higher efficacy of miR-122 KD and lower the context+ scores, which represented the higher miR-122 binding potential (Fig. 16). Since $\log_2\text{-ERD}_{\text{RNA}}$ was a distribution of ratios and by the laws of the propagation of uncertainty, the increased variation could result from the increased variation in the denominator (control) distribution, or from the numerator (KD) distribution, or from both. In our experiment and the case of miR-122, despite we could not fully dismiss the possibility that they both contributed to some extent, it makes more sense if the increased variation was majorly attributed to DDX6 or miR-122 KD instead of control. As such, we would argue that DDX6 acts to reduce the variations in the RNA expression level of its targets. The variations in the gene expression level are generally considered as “noises”, as they represent the deviation from the presumed values. With this regard, our results indicated that DDX6 regulates the expression noises of its targets.

4.5. DDX6 has limited effect on the overall protein expression level

Since DDX6 KD had limited effect on the overall RNA expression level, we questioned whether DDX6, as a translational repressor, function primarily in blocking



protein synthesis. To this end, we conducted quantitative MS analysis to assess the effect of DDX6 KD in the HeLa cells. Prior to MS analysis, the cells were fractionated into the cytoplasmic and the nuclear fractions and the DDX6 KD efficacy was confirmed (Fig. 17). The two fractions were analyzed separately. The MS detection spanned more than 5 orders of magnitudes, covering more than 3000 proteins (Fig. 18a). Similar to our observation on the RNA level, few proteins were deemed statistically significant (Fig. 18a and b), and this could be attributed to the conservative significance B distribution (Fig. 19a). This contradicts our expectation that certain protein would exhibit elevated protein level due to the translational de-repression that results from DDX6 KD. In addition, the majority of the significant proteins differed less than 2 fold in expression level (Fig. 18b). Furthermore, the replication-wise examination revealed inconsistent fluctuations in the expression ratios in approximately half of the significant proteins (Fig. 18c). A collection of the significant proteins was selected according to the coefficient of variation and was displayed as representatives (Fig. 18d). Noticeably, the expression ratios of the representative proteins were limited (Fig. 18d). To further assess the impact of DDX6 KD on the overall protein expression level, we explored the data with pairwise correlation and PCA. The control and DDX6 KD samples in the same replicates resembled each other more than the same treatment in other biological replicates (Fig. 19b). Again, this is reminiscent of our observation on the RNA level (Fig. 9e). Moreover, only $\approx 1\%$ of the total variance in both the cytoplasmic and the nuclear fractions were attributed to DDX6 KD, which corresponded to the PC5 (Fig. 19c and d). The PC1 to the PC4 corresponded to differences between the replicates and accounted for $> 95\%$ of the total variance (Fig. 19c and d). One thing to be noted, though, is that the quantified proteins (≈ 3000 proteins) in this study cover only $\approx 30\%$

of the human proteome. Therefore, it is possible that we have not quantified all the potential variation caused by DDX6 KD.



4.6. DDX6 regulates the protein expression noise of its bound mRNA

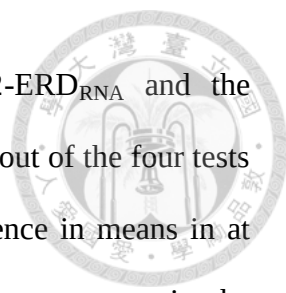
We further examined whether DDX6 binding affects protein production by integrating the anti-DDX6 RIP-seq results with the DDX6 KD quantitative MS dataset. We compared the log₂-FED (anti-DDX6 IP/control IgG), and the log₂-transformed protein expression ratio distribution (DDX6 KD/control), abbreviated as log₂-ERD_{protein}, which represented the effect size of DDX6 on the protein expression. In both the cytoplasmic and the nuclear fractions, we found no correlation between the DDX6 binding strength and the effect size on the protein expression (Fig. 20a and b). To remove the effect of the irrelevant genes, we stratified the genes into the enriched genes and background genes according to the binary binding status. As a result, in the cytoplasmic fraction, there were no statistically significant differences in the means and the standard deviations of the log₂-ERD_{protein} between the enriched genes and the background genes (Fig. 20c and e). On the opposite, the means and the standard deviations of the log₂-ERD_{protein} in the nuclear fraction were both higher in the enriched genes (Fig. 20d and f). However, the (geometric) means of the enriched genes and the background genes differed only about 5% (on the normal scale), a difference presumably too small to be precisely verified by WB. On the other hand, the standard deviations increased $\approx 20\%$ (Fig. 20f). Yet again, the results in the cytoplasmic fraction and the nuclear fraction were inconsistent, indicating a more restricted role for DDX6 in regulating protein expression noise than in regulating RNA expression noise.

4.7. DDX6 has limited effect on the polysome occupancy.

In *S. cerevisiae*, Dhh1p can reduce polysome substantially when overexpressed (Coller and Parker, 2005). However, Dhh1p knockout and DDX6 KD have very limited effect on the polysome (Coller and Parker, 2005; Yoon et al., 2012), despite DDX6 KD can mildly upregulate overall protein synthesis rate (Chu and Rana, 2006; Jangra et al., 2010). We therefore explored the effect of DDX6 KD on polysome integrity and the translation of its target mRNA. We first assessed the impact of DDX6 KD on polysome. Consistently, we observed no explicit, qualitative difference in the polysome profiles between control and DDX6 KD cells (Fig. 21a and b). Furthermore, we examined the polysome occupancy of three different genes whose RNA products are bound by DDX6 (Fig. 21c), including GPRC5A (Y. C. Chen's master thesis), GAPDH, and GAS5, a lncRNA. The expression of GAPDH, GPRC5A, and GAS5 on both RNA and protein levels were not altered by DDX6 KD (Fig. 21a and 22). The polysome occupancy for both GAPDH and GPRC5A were also unaltered by DDX6 KD (Fig. 21d). Consistently, the sedimentation of GAS5, who was associated with the low molecular weight RNP, was also unaffected (Fig. 21d).

4.8. DDX6-regulated RNA and protein expression are weakly coupled

Last, we explored the relationship between changes in the mRNA and the protein expression under DDX6 KD. To this end, we compared the DDX6 KD total RNA-seq dataset with the DDX6 KD quantitative MS dataset. The $\log_2\text{-ERD}_{\text{RNA}}$ and the $\log_2\text{-ERD}_{\text{protein}}$ exhibited low positive correlations in both the cytoplasmic and the nuclear fraction (Fig. 23a and b). We further binned the mRNA and the protein expression level into 5 percentile groups: 0-20, 20-40, 40-60, 60-80, and 80-100

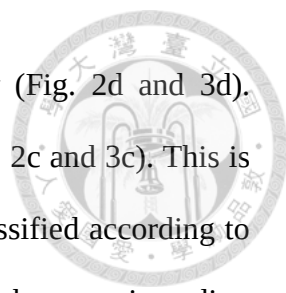


percentiles (Fig. 23c-f), and compared the means of the $\log_2\text{-ERD}_{\text{RNA}}$ and the $\log_2\text{-ERD}_{\text{protein}}$ of the different percentile groups by ANOVA. Three out of the four tests passed the F test (Fig. 24a), and had statistically significant difference in means in at least one pair of percentile groups (Fig. 24b). In general, the larger means in the $\log_2\text{-ERD}_{\text{RNA}}$ correlated with the larger means in the $\log_2\text{-ERD}_{\text{protein}}$ and vice versa (Fig. 23b). Together, this indicates that the change in mRNA expression and the change in protein expression upon DDX6 KD were weakly coupled. Furthermore, as the RNA-seq samples and the quantitative MS samples were prepared separately, the positive correlation between the $\log_2\text{-ERD}_{\text{RNA}}$ and the $\log_2\text{-ERD}_{\text{protein}}$ indicated DDX6 KD produced repeatable effect on the overall gene expression, albeit the effect size was limited.

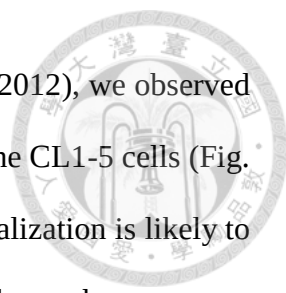
5. Discussion

5.1. A generalization of the concept of the DDX6's RNA binding activity

In the previous chapter, we demonstrated that, in addition to binding cytoplasmic mRNA, DDX6 present in the human cell nuclei can also bind to the nuclear lncRNA MALAT1 (Chap. 2). This observation supports the idea that DDX6 can bind to the lncRNA. In this chapter, we further generalized this idea. First, we demonstrated that lncRNA binding is a prevalent phenomenon for DDX6. In both the HeLa total RNA-seq and the poly(A)⁺ RNA-seq datasets, the lncRNA genes accounted for less than 20% of the expressed genes (Fig. 5b and c); however, the lncRNA accounted for more than 40% of the enriched genes in the HeLa RIP-seq dataset (Fig. 5a). Although this number is at risk of being inflated, it nevertheless indicates DDX6 should at least bind the mRNA and the lncRNA proportionally to their fractions in the expressed genes. In other words,



DDX6 binding is not restricted to few lncRNA genes, but many (Fig. 2d and 3d). Furthermore, DDX6 can bind the lncRNA of different biotypes (Fig. 2c and 3c). This is not very surprising because the lncRNA gene biotypes were not classified according to the biochemical properties but to their genomic position relative to other protein coding gene loci (referred to the GENCODE website for more details). Second, binding to nuclear RNA is a general phenomenon for nuclear DDX6 (Fig. 6). In a narrower sense, the interaction between DDX6 and MALAT1 is not a strange case (Chap. 2), nor is the CUG-expanded DMPK mRNA (Pettersson et al., 2014). In a broader sense, the nucleocytoplasmic localization of DDX6 does not impose apparent constraint on its RNA binding activity. As such, the nuclear DDX6 are expected to form yet uncharacterized RNP complexes in the cell nuclei, since DDX6 binds RNA at high affinity (Dutta et al., 2011; Ernoult-Lange et al., 2012). Resonating with this notion, the DDX6 homologs were proposed to bind nascent transcripts or modulate transcription (Haimovich et al., 2013; Kruk et al., 2011; Smillie and Sommerville, 2002), both of which involve forming functional RNPs. In addition, we also demonstrated that DDX6 can co-localize with 4E-T in the nuclear 4E-T granules, and presumably in the same protein complex (Chap. 2). Furthermore, despite the subcellular localization of the lncRNA bias towards the cell nuclei and many lncRNA possess known nuclear functions (Cabili et al., 2015; Derrien et al., 2012), DDX6 does not interact exclusively with the lncRNA in the cell nuclei (Fig. 6c). Instead, a predominant fraction ($\approx 70\%$) of the nuclear binding targets of DDX6 is mRNA (Fig. 6c). This again resonates with the potential of binding to nascent RNA (Smillie and Sommerville, 2002). Third, lncRNA binding is a general phenomenon for DDX6 across different cell types. In addition to the previously reported case of lncRNA binding by DDX6 in human cancer cells and



mouse embryonic fibroblasts (Chap. 2) (Iio et al., 2013; Yoon et al., 2012), we observed the widespread binding of lncRNA by DDX6 in both the HeLa and the CL1-5 cells (Fig. 2 and 3). Since these cell lines have very different origins, this generalization is likely to apply to other mammalian cell types. The generalization in these three aspects demonstrates lncRNA binding can be as prevalent as mRNA binding for DDX6. This strongly supports our argument that DDX6 perceive no difference between the mRNA and the lncRNA on a local scale (Chap. 1). As such, DDX6 is likely to be involved in the post-transcriptional regulation of the lncRNA. The urgent mission at hand is to provide empirical evidence for such an idea.

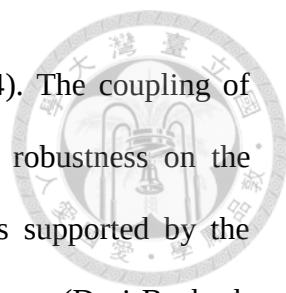
5.2. On the role of DDX6 in regulating RNA and protein expression

Originally, we expected DDX6 KD would result in strong stabilization or the translational de-repression of some genes at RNA level, which would then led to the alteration in the gene expression profile on both the RNA and the protein level. However, our RNA-seq (Fig. 8-10) and quantitative MS results (Fig. 17-19), both of which revealed limited impact of DDX6 KD on the overall gene expression, strongly contradicted such expectation. One rather naïve, but not completely implausible explanation to this discrepancy is that the residual amount of DDX6 is sufficient to ensure normal RNA turnover rate and translational repression (Fig. 9 and 17). However, the similar level of DDX6 KD readily produce observable consequences in human cell models, including the mild elevation in protein synthesis rate, the disruption of the PBs, and the de-repression of the miRNA reporters (Chu and Rana, 2006), all of which are causally related to the general RNA turnover and translational repression. Similarly, strong miR-122 KD also had a limited effect on the overall RNA expression level (Fig.

13 and 14). Thus, instead of interpreting our results as a lack of impact, we sought for alternative explanations that might help uncover the underlying biology.

One possibility is that DDX6 has widespread effects over a large group of genes, but each to a modest degree. This is consistent with the results obtained from various different post-transcriptional regulators, including the miRNA (Agarwal et al., 2015; Baek et al., 2008; Denzler et al., 2014; Eichhorn et al., 2014; Guo et al., 2010; Hafner et al., 2010; Helwak et al., 2013; Wong et al., 2015), HuR (Lebedeva et al., 2011; Mukherjee et al., 2011), QKI (de Bruin et al., 2016; Hafner et al., 2010), HNRNPA1 and HNRNPA2/B1 (Geissler et al., 2016), PUM1 and PUM2 (Hafner et al., 2010; Lee et al., 2016; Tichon et al., 2016), and IGF2BP (Hafner et al., 2010). For these regulators, the average alteration in gene expression for the best targets (by prediction or empirical identification, such as CLIP-seq) is typically under 30% when the concerned regulator was perturbed. Moreover, the effect size decreases with the binding strength. Thus, weak but widespread effect appears to be one commonality among a considerable subset of the post-transcriptional regulators. Unlike Dhh1p (Mitchell et al., 2013), these regulators displayed preferences to short RNA motifs of various sequences and lengths that were readily detected by CLIP-seq, allowing for the better stratification of true targets and non-targets and further allowing for the detection of the weak regulatory effects over the ensemble of the true targets. In other words, the deterministic interaction directed by the target RNA motifs can played a more prominent role for these regulators as compared to DDX6.

Another possibility is that the reduced activity of DDX6 in activating RNA decay also compromises transcription. This possibility stems from the theory that transcription



and RNA decay are dynamically coupled (Braun and Young, 2014). The coupling of these two seemingly antagonizing processes is thought to confer robustness on the overall gene expression (Haimovich et al., 2013). This concept is supported by the coupled evolution of the transcription rate and the RNA degradation rate (Dori-Bachash et al., 2011), and the positive roles of decay factors in facilitating transcription (Braun et al., 2014; Haimovich et al., 2013; Sun et al., 2013), including Dhh1p (Haimovich et al., 2013). The key idea in this concept is disrupting the RNA decay also compromises transcription. Under this framework, the limited change in the overall RNA expression level following DDX6 KD could be interpreted as a consequence of the compromised RNA decay and synthesis. The largely invariant RNA expression level may therefore constrain the protein expression level to some degree. Consistently, it was reported that DDX6 KD in HeLa cells only mildly (≈ 1.33 fold) increases total protein synthesis (Chu and Rana, 2006), if such phenomena was detected at all (Jangra et al., 2010). Consistently, we (Fig. 23) and others (Coller and Parker, 2005; Yoon et al., 2012) all observed that DDX6 loss-of-function had limited effect on the polysome profiles, which also indicates only mild changes, if any at all, in the cell's translational status. Accordingly, our quantitative MS data also showed that the overall protein expression level remained largely invariant following DDX6 KD, suggesting the de-repression effect of DDX6 KD was evenly distributed to most genes. Notably however, our workflow for the quantitative MS was set to detect the differentially expressed proteins between the control and the DDX6 KD conditions rather than quantifying the global changes in the absolute abundance. Thus, the samples were normalized in many steps in our workflow, on both the wet bench and the dry bench, making it impossible for us to infer the difference in protein synthesis rate under the two conditions from the MS data.

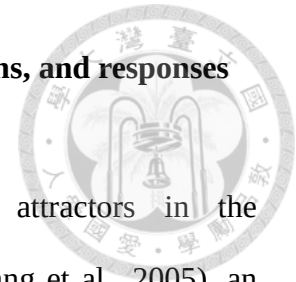
It is worthy to note here that the two possibilities are not mutually exclusive to each other and both remain to be verified by further investigation.



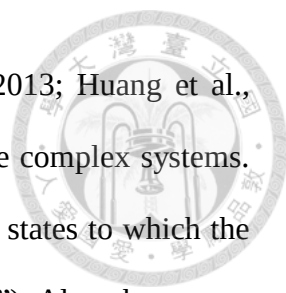
Previously, a gene expression profiling effort has identified ≈ 500 differentially expressed genes in the DDX6 KD human primary keratinocytes (Wang et al., 2015), which is drastically different to our results. Nonetheless, approximately 80% of them belong to the keratinocyte proliferation/differentiation program, and the activation of this program was attributed to the deregulation of several master genes that governs it in the wake of DDX6 KD (Wang et al., 2015). Wang et al.'s observation is consistent with the fact that DDX6 is a promiscuous RNA regulator but not a transcription factor. As such, this case represents a different scenario from our observations in the HeLa cells.

Despite observing only limited changes in the overall gene expression in the HeLa cells, we discovered the increased variation in the expression ratio in the enriched genes on both the RNA (Fig. 15) and the protein level (Fig. 20). The increased variation in the RNA expression ratio can be attributed to the increased expression noise in the DDX6 KD cells (Fig. 25b). Very similar phenomenon was observed in the miR-122 KD datasets (Fig. 16). Furthermore, it has been shown that miRNA decreases the protein expression noise (Schmiedel et al., 2015). And since target RNA destabilization was the dominant effect of miRNA (Eichhorn et al., 2014; Guo et al., 2010), the reduction in the protein expression noise by miRNA can at least partially result from the reduction in the RNA expression noise. This notion is also consistent with the same DDX6/miRNA analogy since DDX6 also reduces both the RNA and the protein expression noises (Fig. 15 and 20). Thus, we conclude that silencing the gene expression noise may be one novel function for the DDX6.

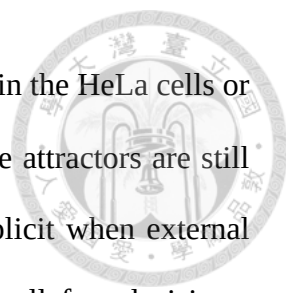
5.3. On the roles of DDX6 in cell fate maintenance, specifications, and responses to stimuli: Under the “cell fate attractor” theory



The cell fate attractor theory describes cell fate as attractors in the high-dimensional gene expression state space (Huang, 2009b; Huang et al., 2005), an idea rooted in the dynamical systems theory. For a genome with N genes, the gene expression state space is an abstract space that has N dimensions, with each axis representing the possible expression value of one gene. Each point in the state space corresponds to a possible state, or configuration, of the system and is represented as an N -dimensional state vector. Here, systems refer to cells. Essentially, the state vector is the mathematical formalization of the expression profile of a cell. In the state vector, the genes are described as time-dependent variables. Consequently, the state vector, representing the cell's state, is also a function of time. Connecting the state vectors at different time points gives a trajectory in the state space, which describes the evolution of the system. Here, the evolution does not refer to the Darwinian evolution but the time-dependent changes in the system. In other words, this theory considers the dynamics of each gene and the system as a whole. Furthermore, the genes are treated as interdependent variables, meaning the value of one variable is causally related to the others. These causal relationships correspond to the regulatory interactions between each gene, which together rear the genome-wide gene regulatory network (Huang, 2009b). In other words, the dynamics of any given state is determined by the architecture of the gene regulatory network (Huang, 2009b). In essence, this theory is the mathematical formalization of the idea in molecular biology that gene expression and the underlying gene regulatory network dictate the cells' behavior. Both theoretical and empirical studies have demonstrated that even simple network motifs can give rise



to multi-stability to the system (Chang et al., 2008; Chen et al., 2013; Huang et al., 2007), suggesting the co-existence of multiple attractors in the more complex systems. In the context of the dynamical system, the attractors are the sets of states to which the system tends to converge from nearby states (“the basin of attraction”). Also, the system trapped by the attractor can remain close to the attractor even if the system is slightly perturbed. This concept therefore resonates with the concept of cell fate in that the cells of certain fates have distinct and rather stable gene expression profiles. The theory of cell fate attractor therefore represents a viable theoretic framework for the study of the cell fates. However, the dynamics of the cell is not completely deterministic as described above. Since the cells are complex molecular systems, gene expression is also subject to the molecular noise, which is presumed to be fluctuating randomly (Huang, 2009a; Raj and van Oudenaarden, 2008). Thus, over a population of cells, the population-wide behavior is subjected to the two opposing force; one attracts, and the other counteracts (Chang et al., 2008; Li et al., 2016). While the attracting dynamics of the system defines the cell fate (Chang et al., 2008; Li et al., 2016), the counter-attracting force of the gene expression noise can push a subset of cells to the edge of the attractor (Chang et al., 2008; Li et al., 2016), where they are prone to escaping the attractor and adopt different fates, such as differentiation (Chang et al., 2008) or apoptosis (Li et al., 2016). Under the attractor theory, adopting a different fate means to move along a trajectory to another attractor. In this chapter, we provided evidence that the loss of DDX6 increases gene expression noise. Therefore, under the framework of the cell fate attractor theory, the loss of DDX6 can increase the escapability of the attractor, which can be idealized as “flattening” the attractor landscape (Fig. 25c). This may not produce explicit phenotypic changes to the cells



under normal conditions on a population scale, as we have observed in the HeLa cells or in the Δ Dhh1p *S. cerevisiae* mutant, because the underlying cell fate attractors are still present. However, the effect of the loss of DDX6 may become explicit when external stimuli (signals or stresses), which normally trigger well-defined cell fate decisions, come into play. More specifically, the cells on the edge of the attractor are prone to adopting diverse routes to exit the attractor depending on their states, which may differ from the intended route that the external stimuli seek to direct. This notion is consistent with that the DDX6 homologs are widely appreciated in the stress responses. Similar mechanism may also underlie the effect of the loss of DDX6 in cell fate specification and development. As such, this theory provide a framework that can circumvent the need to exact specific downstream targets for the DDX6 homologs only to justify their significance, which really is against their nature. More importantly, this theory can potentially unify the diverse roles that have been proposed for the DDX6 subfamily members in physiology and post-transcriptional regulation.



6. Figures

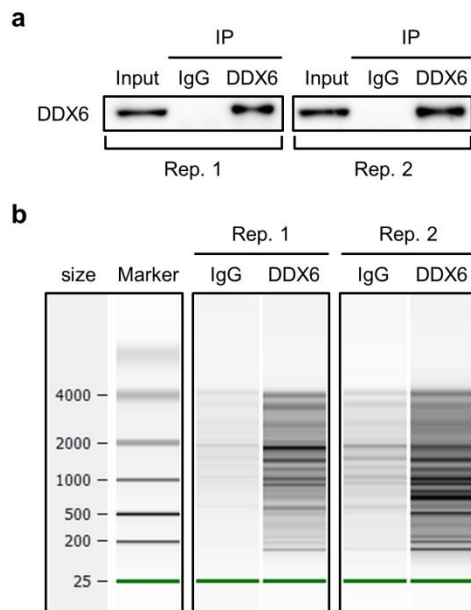


Figure 1. The quality assessment of the CL1-5 RIP-seq sample

a, The WB confirmation of anti-DDX6 IP. **b**, The virtual gels showing co-IPed RNA contents with control IgG and anti-DDX6 IP. The RNA co-purified from the two replicates were pooled together for sequencing.

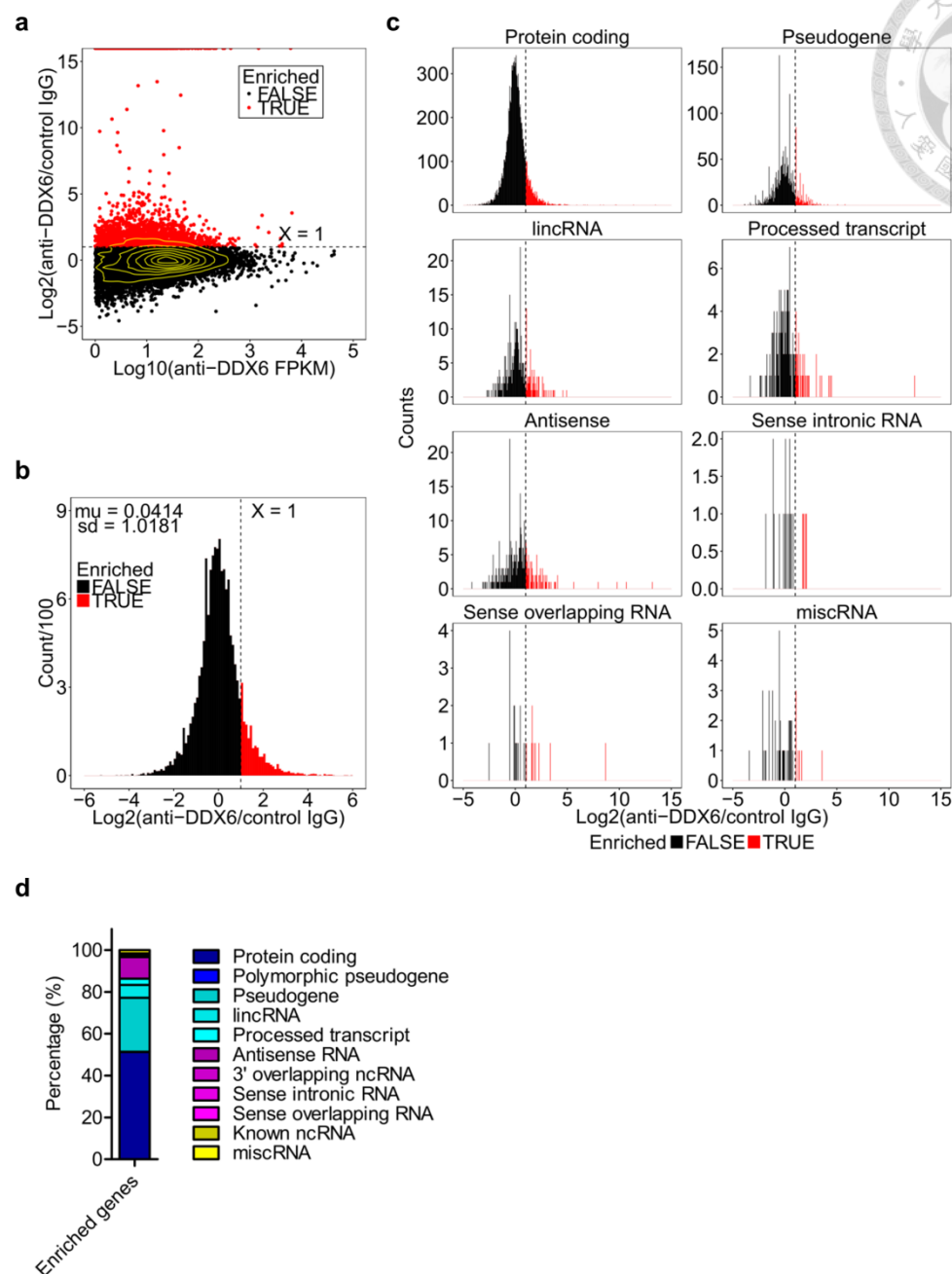
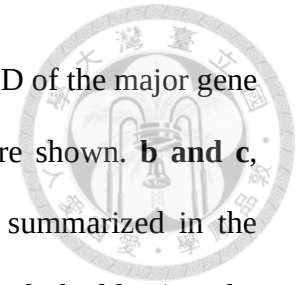


Figure 2. DDX6 binds both the mRNA and the lncRNA in the HeLa cells.

a, The MA plot of the HeLa anti-DDX6 RIP-seq dataset. The point density was depicted by the yellow contour lines. Only genes with FPKM ≥ 1 in the anti-DDX6 sample were plotted; other genes were filtered. **b**, The histogram of showing the log2-transformed fold enrichment distribution (log2-FED). mu: the population mean of the log2-FED; sd:

the standard deviation of the log2-FED. **c**, Histogram of the log2-FED of the major gene biotypes. The eight major biotypes with higher gene numbers were shown. **b and c**, Only the genes with FPKM ≥ 1 in the anti-DDX6 sample were summarized in the histograms. **a-c**, The genes with ≥ 2 fold enrichment (marked by the dashed line) in the anti-DDX6 IP sample relative to the control IgG sample were considered as the enriched (DDX6-bound) and were colored in red. **d**, The bar graph showing the biotype composition of the enriched genes. The protein coding (mRNA) biotype was colored in dark blue and was placed at the bottom of the stack.



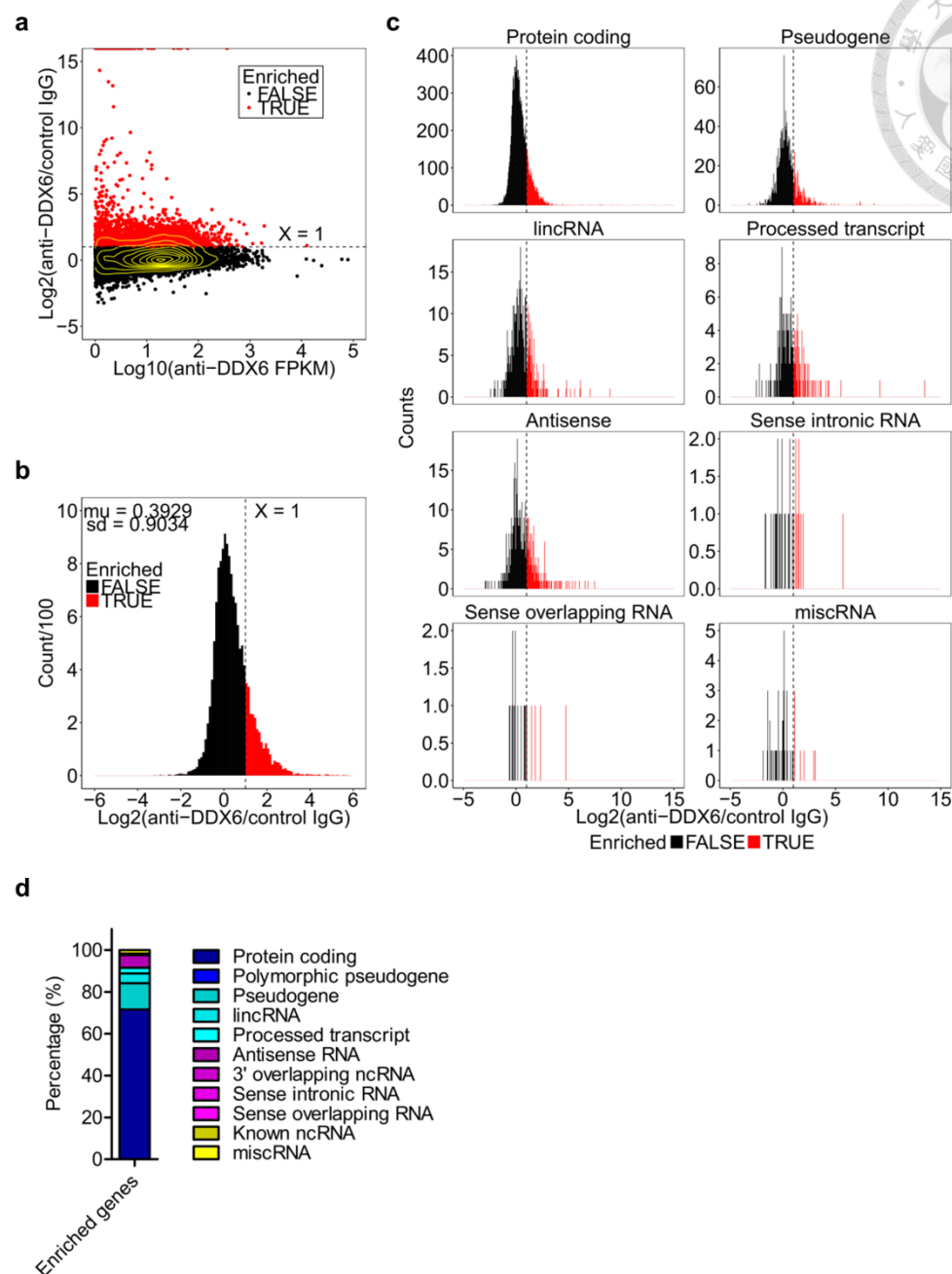
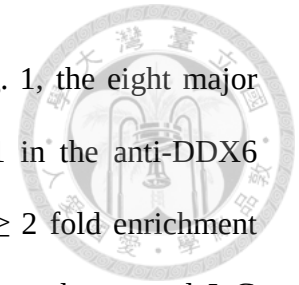


Figure 3. DDX6 binds both the mRNA and the lncRNA in the CL1-5 cells.

a, The MA plot of the CL1-5 anti-DDX6 RIP-seq dataset. The point density was depicted by the yellow contour lines. Only genes with $\text{FPKM} \geq 1$ in the anti-DDX6 sample were plotted; other genes were filtered. **b**, Histogram of the log2-FED. μ : the population mean of the log2-FED; sd : the standard deviation of the log2-FED. **c**,

Histogram of the log₂-FED of the major gene biotypes. As in Fig. 1, the eight major biotypes were shown. **b and c**, Only the genes with FPKM ≥ 1 in the anti-DDX6 sample were summarized in the histograms. **a-c**, The genes with ≥ 2 fold enrichment (marked by the dashed line) in the anti-DDX6 IP sample relative to the control IgG sample were considered as the enriched and were colored in red. **d**, The bar graph showing the biotype composition of the enriched genes. The protein coding biotype was colored in dark blue and was placed at the bottom of the stack.



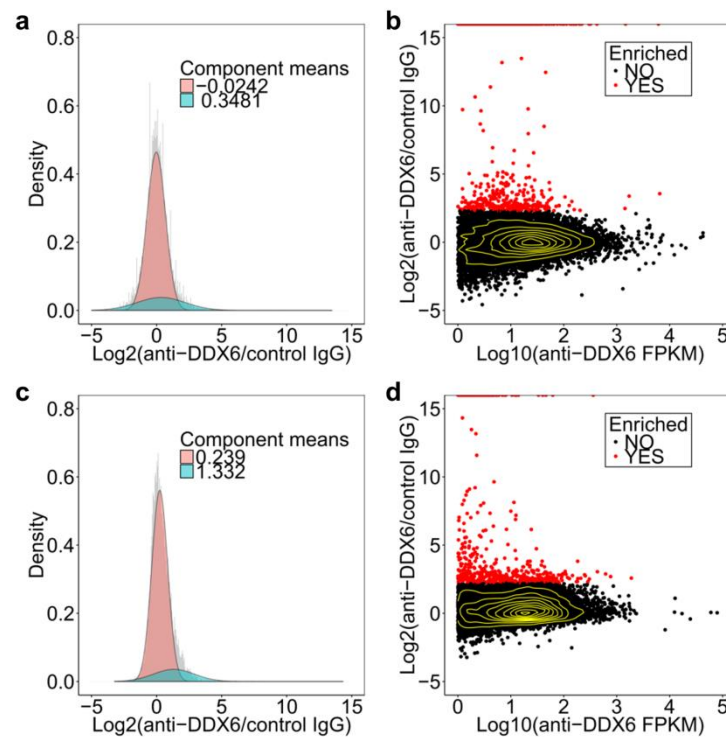
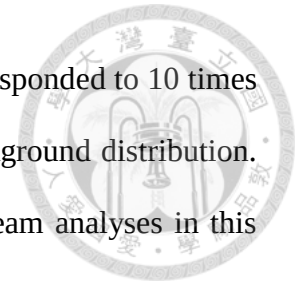


Figure 4. Decomposing the log₂-FED in the RIP-seq datasets by Gaussian mixture modeling and the posterior probabilities

a and c, The density plots showing the two components of the log₂-FED. The log₂-FED in the HeLa (**a**) and the CL1-5 (**c**) RIP-seq datasets were decomposed into two components (distributions). The component with larger mean was interpreted as the enriched gene distribution, whereas the component with smaller mean was interpreted as the background distribution. **b and d**, The MA plots showing the enriched genes filtered by mixture decomposition and posterior probabilities in the HeLa cells (**b**) and the CL1-5 cells (**d**). The enriched genes were colored in red. The enriched genes were determined by the log₁₀-transformed odds ratios (log of odds, LOD), the ratios of the posterior possibilities for a given fold enrichment value to be in one component relative to another. This approach was previously described by Lu et al., *Genome Biol.* (2014).

Here, the threshold was $\text{LOD} \geq 1$, the same as Lu et al., which corresponded to 10 times more probable to be in the enriched gene distribution than the background distribution. This criterion was conservative and was restrictive to the downstream analyses in this study.



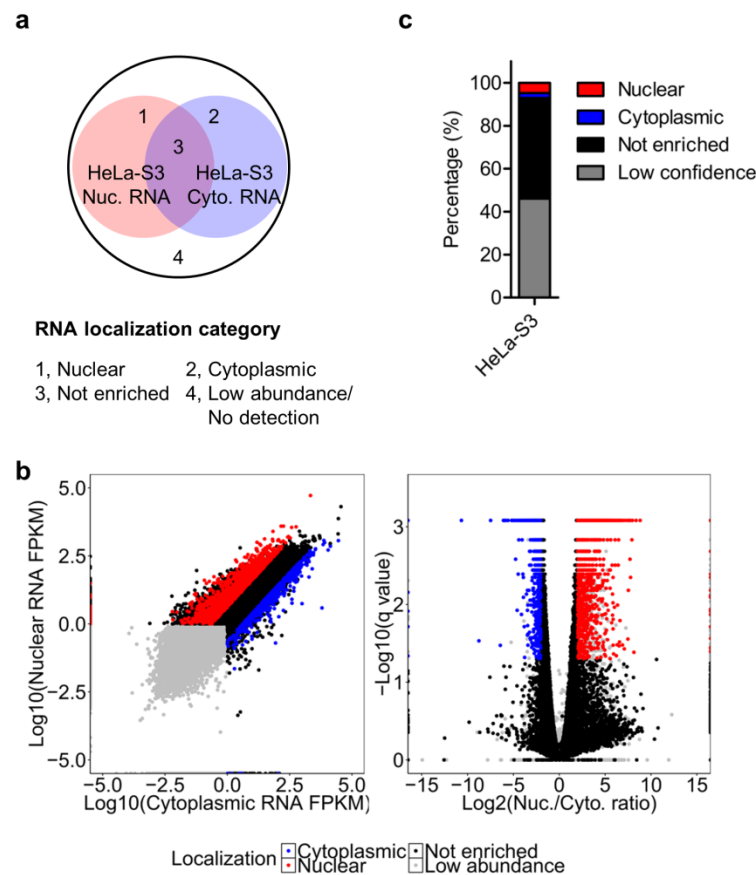


Figure 5. The biotype compositions in the enriched genes, the total RNA transcriptome, and the poly(A)+ RNA transcriptome in the HeLa cells

a-c, The bar graphs showing the biotype compositions in the enriched genes in the HeLa anti-DDX6 RIP-seq dataset (**a**) and the detected genes in the DDX6 KD total RNA-seq dataset (**b**) and in DDX6 KD poly(A)+ RNA-seq dataset (**c**). The protein-coding biotype was filled with dark blue and was placed at the bottom of the stack. All other shown biotypes are noncoding RNA and have lncRNA products.

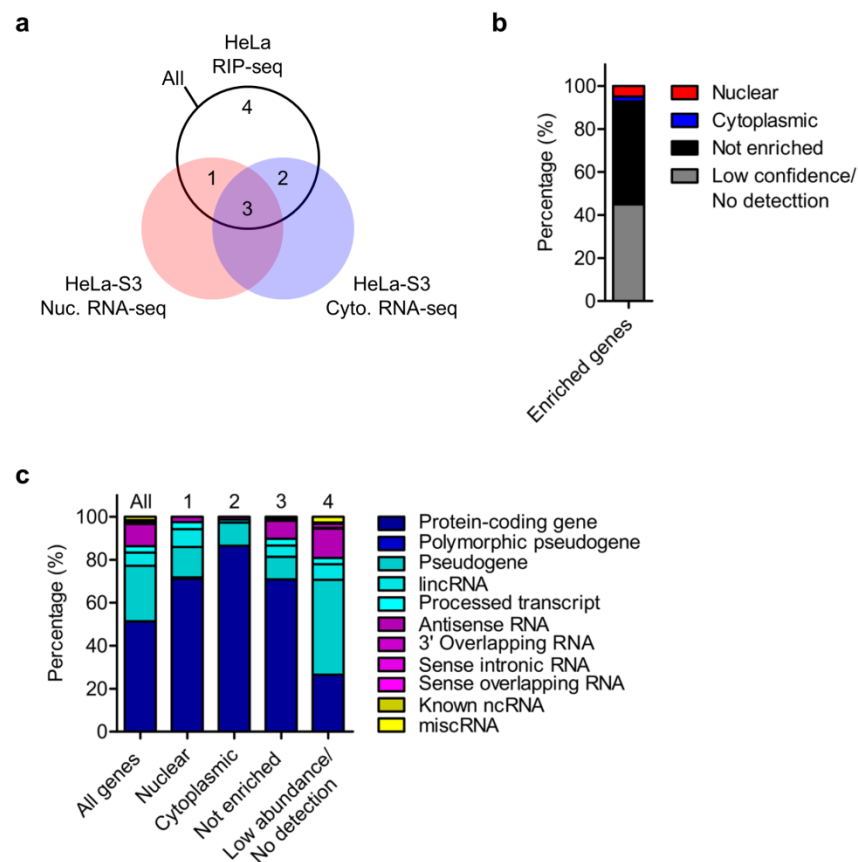
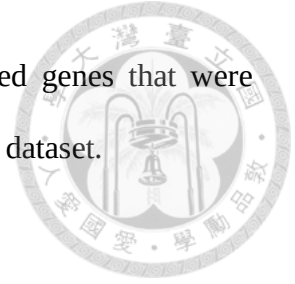


Figure 6. DDX6 binds compartment-enriched RNA, including both the mRNA and the lncRNA.

a, The schema for the integrative analysis of the anti-DDX6 RIP-seq dataset and the HeLa-S3 cytoplasmic and nuclear RNA-seq dataset. **b and c**, The bar graphs showing the subcellular localization (**b**) or the biotype (**c**) compositions of the enriched genes in the anti-DDX6 RIP-seq dataset. The enriched genes were defined as in Fig. 1. **b**, The classification of the RNA subcellular localization is described in Fig. S3. **a and c**, All: all enriched genes in the anti-DDX6 RIP-seq dataset; category 1: The enriched genes with predominant nuclear RNA localization; category 2: The enriched genes with predominant cytoplasmic RNA localization; category 3: The enriched genes without

significant nucleocytoplasmic enrichment; category 4: The enriched genes that were expressed at very low level or not detected in the HeLa-S3 RNA-seq dataset.



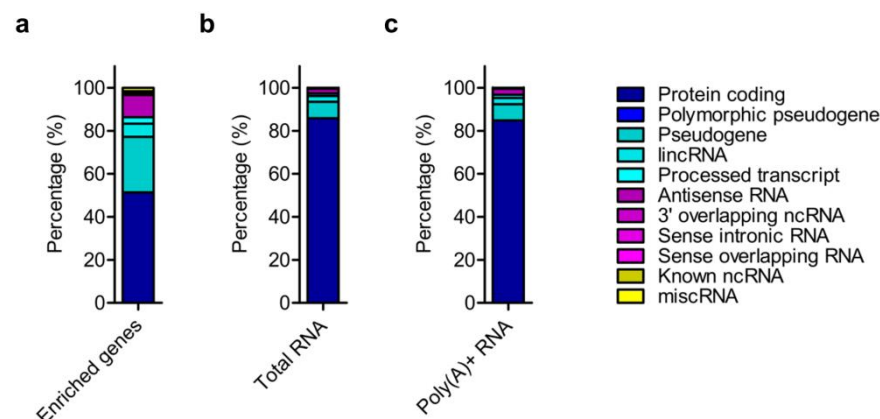


Figure 7. Defining the RNA subcellular localization using the HeLa-S3 cytoplasmic and nuclear RNA-seq dataset

a, The schema for defining RNA subcellular localization using the HeLa-S3 cytoplasmic and nuclear RNA-seq dataset. Category 1: genes with predominant nuclear localization; category 2: genes with predominant cytoplasmic localization; category 3: genes without significant nucleocytoplasmic enrichment; category 4: low abundance or non-detected genes. **b**, The scatter plot and the volcano plot depicting the classification of genes into the four localization categories. The non-detected genes were omitted from plotting. The low abundance genes, defined by FPKM < 1 in both cytoplasmic and nuclear RNA-seq samples, were colored in grey. The nuclear and the cytoplasmic genes were defined by q value threshold at $q < 0.05$ and by enrichment threshold at ≥ 4 fold in either compartment. The rest of the genes were deemed as not significantly enriched in either compartment. **c**, The bar graph showing the localization composition of the HeLa-S3 transcriptome. As in **a**, the non-detected genes were omitted from plotting.

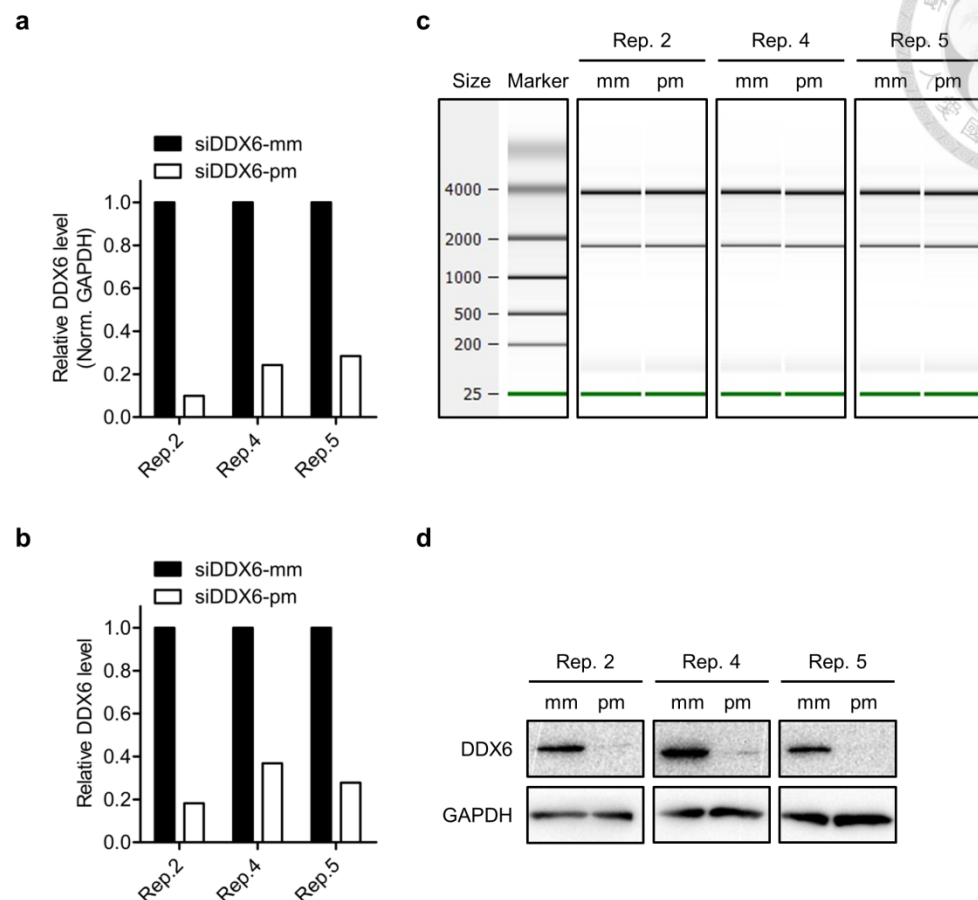


Figure 8. The quality assessment of the HeLa total RNA-seq samples

a, The bar graph showing DDX6 KD confirmation by qPCR. **b**, The virtual gel showing the total RNA content and integrity in the RNA-seq samples. All samples had RIN values ≥ 9.9 . **c**, The bar graph showing DDX6 KD in the RNA-seq result. **d**, The WB confirming DDX6 KD at protein level. GAPDH was used as internal control.

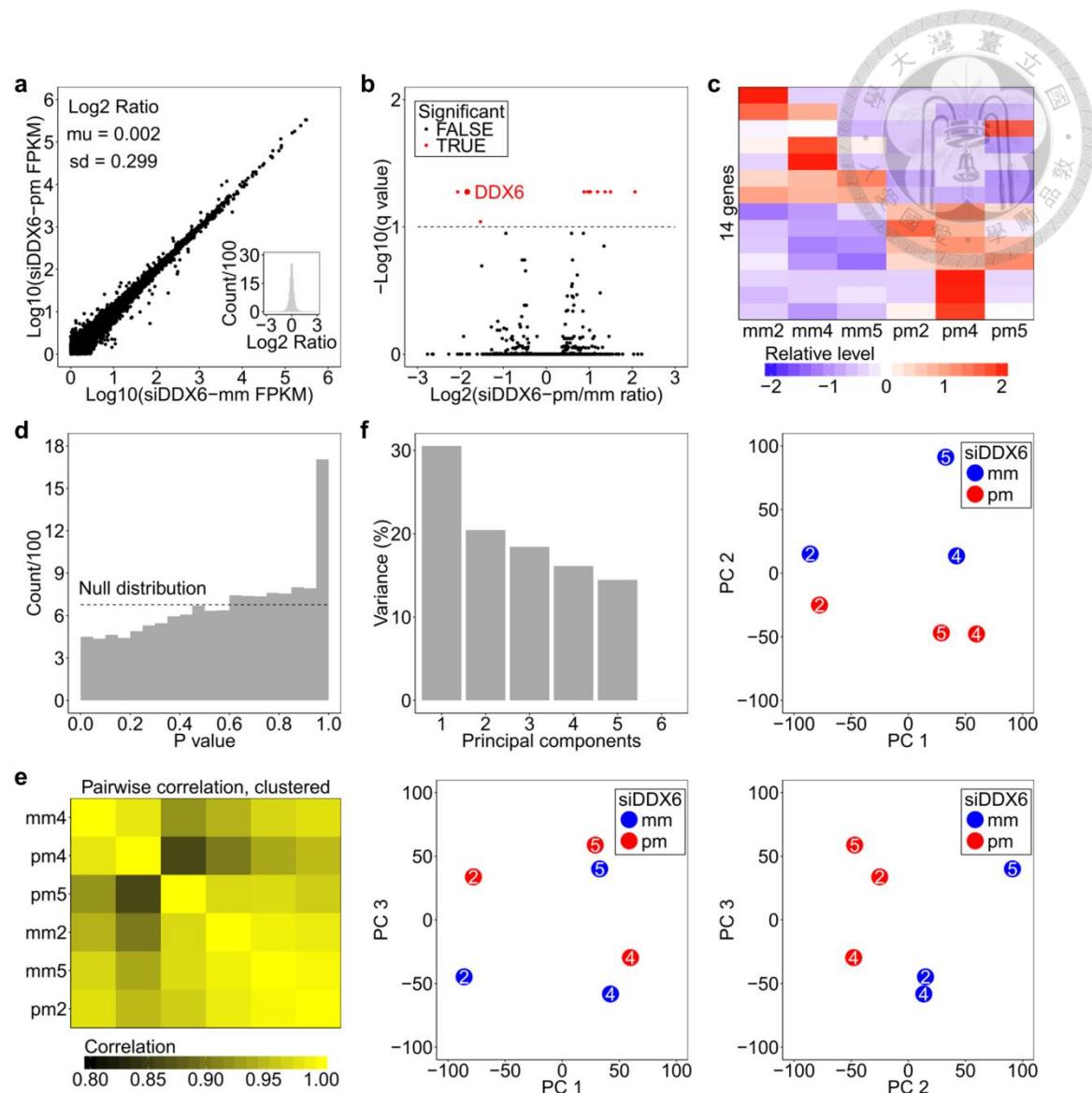
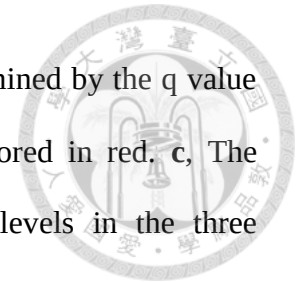


Figure 9. DDX6 has limited effect on the overall RNA expression level.

a, The scatter plot of the DDX6 KD total RNA-seq dataset, showing comparison between the control and the DDX6 KD samples. siDDX6-pm (pm): siRNA with perfect-matching sequence to DDX6; siDDX6-mm (mm): control siRNA with mismatched sequence. The inset shows the histogram of log2-transformed RNA expression ratio distribution ($\log_2\text{-ERD}_{\text{RNA}}$). μ : the population mean of the $\log_2\text{-ERD}_{\text{RNA}}$; sd : the standard deviation of the $\log_2\text{-ERD}_{\text{RNA}}$. **b**, The volcano plot of

the DDX6 KD RNA-seq dataset. Statistical significance was determined by the q value threshold at $q \leq 0.1$. The statistically significant genes were colored in red. **c**, The heatmap of the significant genes, showing relative expression levels in the three biological replicates. mm2, 4, 5, and pm2, 4, 5: the RNA-seq sample ID. **d**, The histogram showing the P value distribution of the differential expression testing. The dashed line indicates the theoretical P value distribution under null hypothesis. **e**, The heatmap showing the pairwise correlations between each RNA-seq samples. The sample order was arranged by unsupervised hierarchical clustering. **f**, PCA of the DDX6 KD RNA-seq dataset. The bar graph depicts the variance explained by each PC. The scatter plots display the mapping of the RNA-seq samples into the PC spaces. In the scatter plots, DDX6 KD samples were colored in red whereas the control samples were colored in blue.



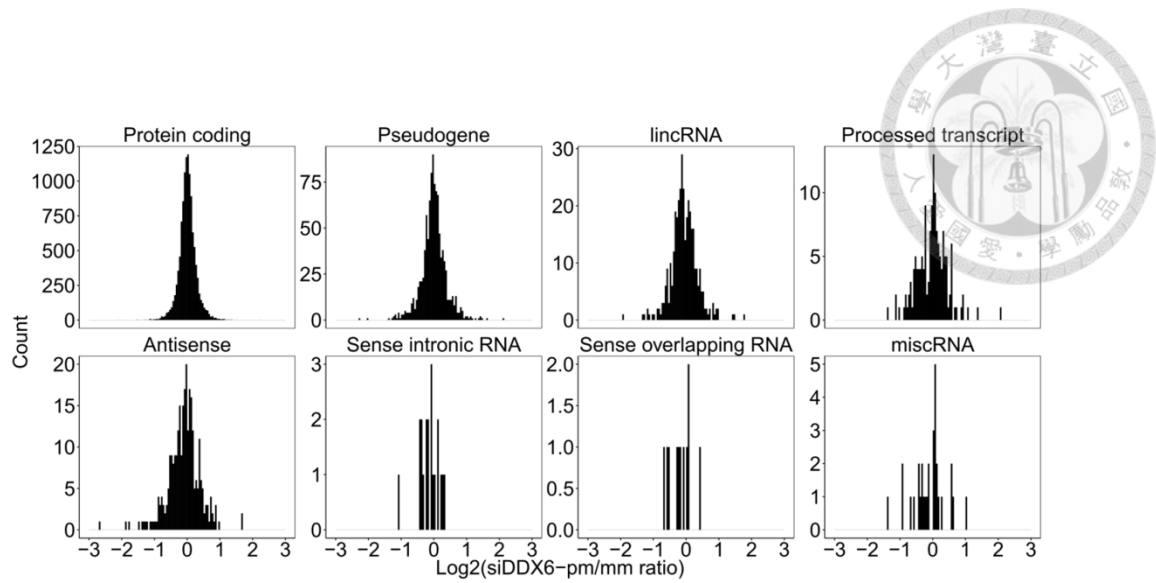


Figure 10. The RNA expression ratio distributions of the major biotypes in the DDX6 KD total RNA-seq dataset

The histograms showing the $\log_2\text{-ERD}_{\text{RNA}}$ in eight major biotypes. The eight major biotypes were selected as previously described.

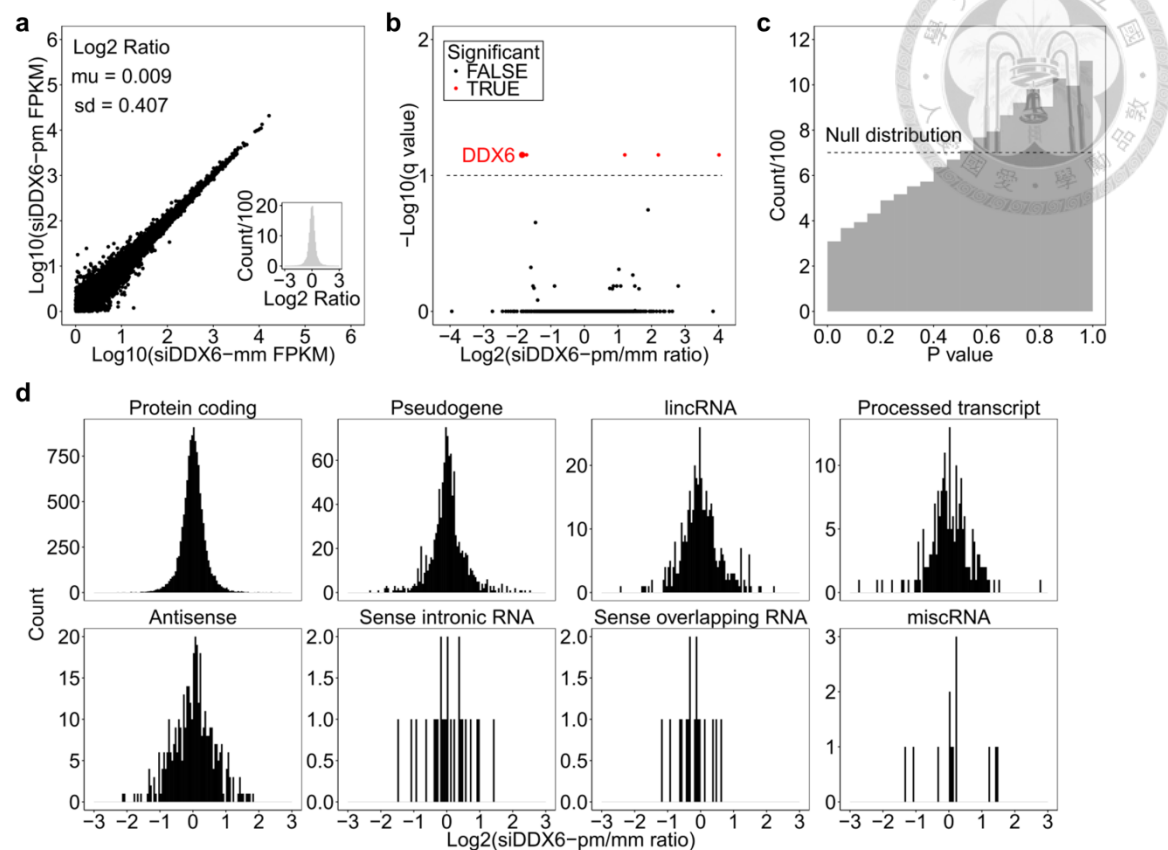


Figure 11. The analysis of the DDX6 KD poly(A)+ RNA-seq dataset

a, The scatter plot of the DDX6 KD poly(A)+ RNA-seq dataset, showing comparison between the control and the DDX6 KD samples. siDDX6-pm (pm): siRNA with perfect-matching sequence to DDX6; siDDX6-mm (mm): control siRNA with mismatched sequence. The inset shows the histogram of the $\log_2\text{-ERD}_{\text{RNA}}$. μ : the population mean of the $\log_2\text{-ERD}_{\text{RNA}}$; sd : the standard deviation of the $\log_2\text{-ERD}_{\text{RNA}}$. **b**, The volcano plot of the DDX6 KD poly(A)+ RNA-seq dataset. The statistically significant genes were determined by the q value threshold at $q \leq 0.1$, and were colored in red. **c**, The histogram showing the P value distribution of the differential expression testing. The dashed line marks the hypothetical distribution under the null hypothesis. **d**, The histogram depicting the distributions of the $\log_2\text{-ERD}_{\text{RNA}}$ in eight major biotypes.

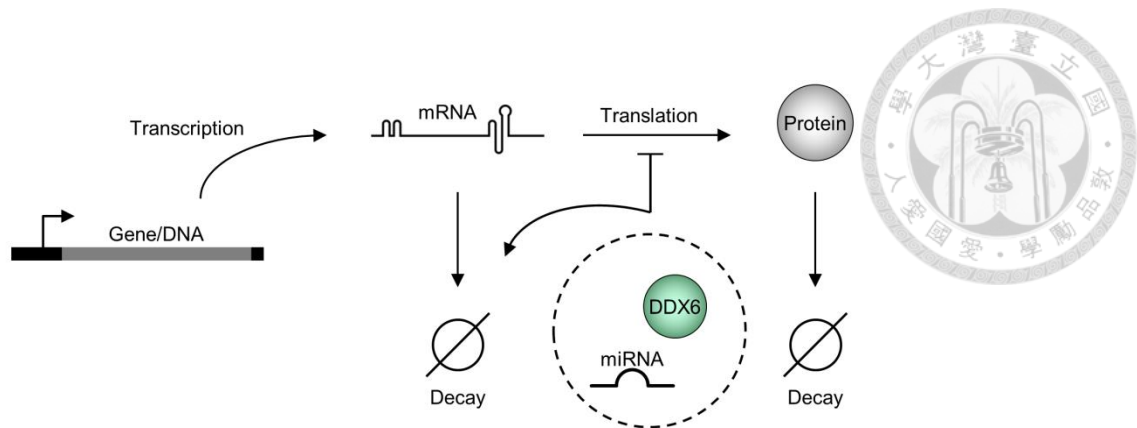


Figure 12. The conceptual analogy between roles of DDX6 and the miRNA in post-transcriptional gene regulation

The diagram depicts the genetic information flow from DNA to protein. Both mRNA and protein are subject to decay. DDX6 and the miRNA both act to facilitate mRNA degradation and inhibit translation. Conceptually, DDX6 and the miRNA share the same molecular “niche” in the post-transcriptional regulation. In reality, DDX6 and the miRNA exert regulatory effects primarily through related, but biochemically distinct mechanisms. Furthermore, both DDX6 and the miRNA bind target RNA promiscuously. Although the miRNA are known to recognize potential targets through seed-region pairing, the miRNA still binds hundreds of targets and exhibit tolerance to the mismatches at the end of the seed regions. On the other hand, DDX6 presumably lack sequence-specificity, but may be differentially recruited to different RNA targets through a diverse set of mechanisms.

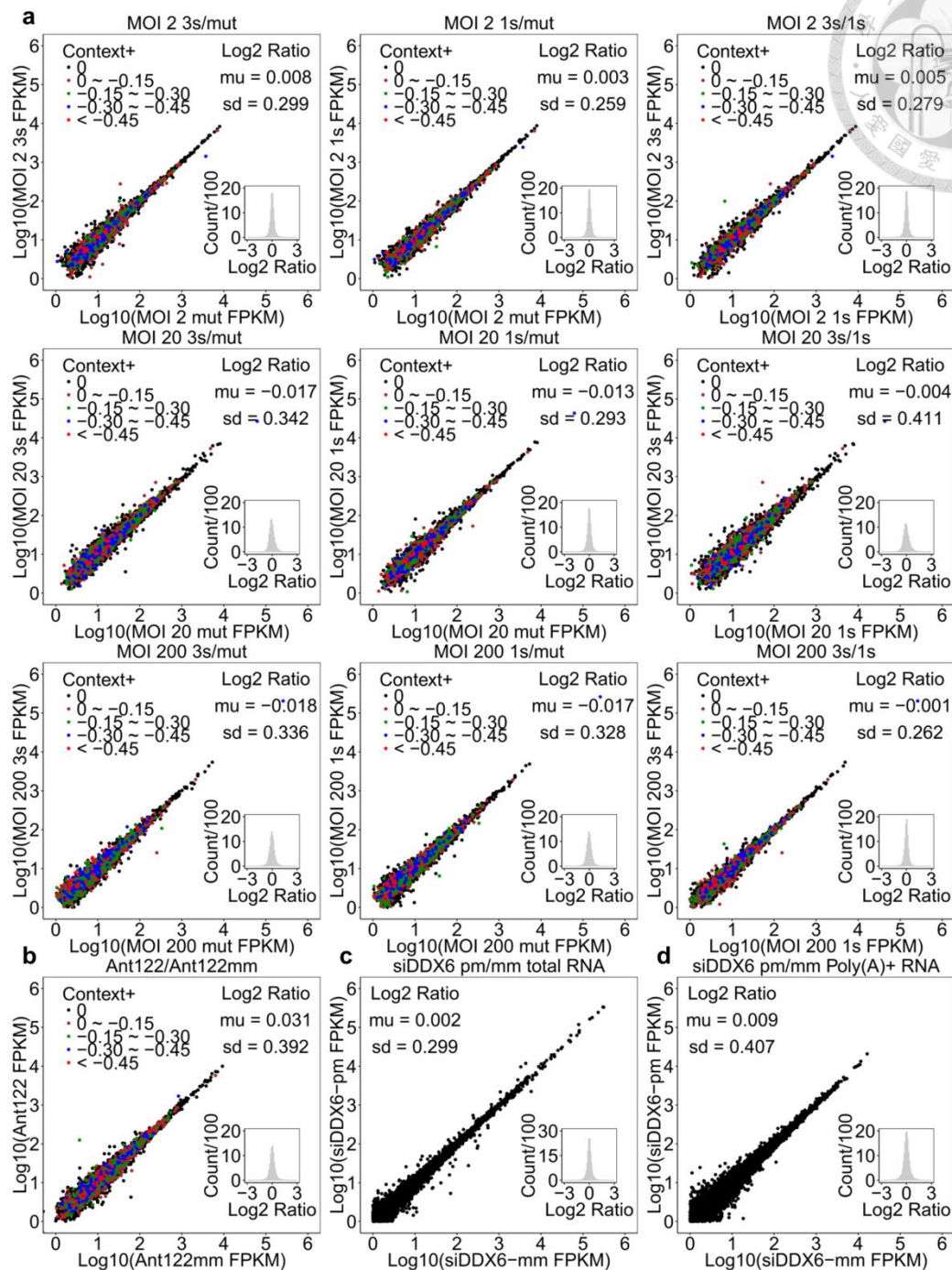
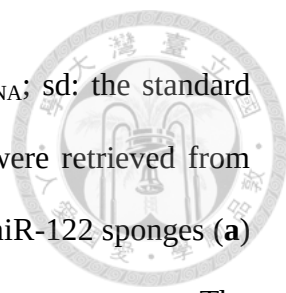


Figure 13. The comparison between the DDX6 KD RNA-seq datasets and the miR-122 KD RNA-seq dataset

a-d, The scatter plots showing comparisons between control and miR-122 KD (**a and b**) or between control and DDX6 KD (**c and d**) RNA-seq datasets. The Insets show the



histograms of the $\log_2\text{-ERD}_{\text{RNA}}$. μ : the means of the $\log_2\text{-ERD}_{\text{RNA}}$; sd : the standard deviations of the $\log_2\text{-ERD}_{\text{RNA}}$. **a and b**, The RNA-seq datasets were retrieved from Denzler et al., Mol. Cell (2014). The mir-122 KD was achieved by miR-122 sponges (**a**) or by the miR-122 antagomir (**b**). The genes were stratified by the context+ scores. The lower scores correspond to the higher confidence in miR-122 binding. Zero score corresponds to non-targets. 3s: three miR-122 site sponge; 1s: one site sponge; mut: control sponge with no miR-122 target site. MOI: multiplicity of infection, the dosage of the mir-122 sponges. Ant122: the antagomir of miR-122; Ant122m: the mismatched control antagomir for Ant122. **c and d**, The scatter plots of the DDX6 KD total (**c**) and poly(A)+ (**d**) RNA-seq datasets as shown in Fig. 5a and Fig. S7a.

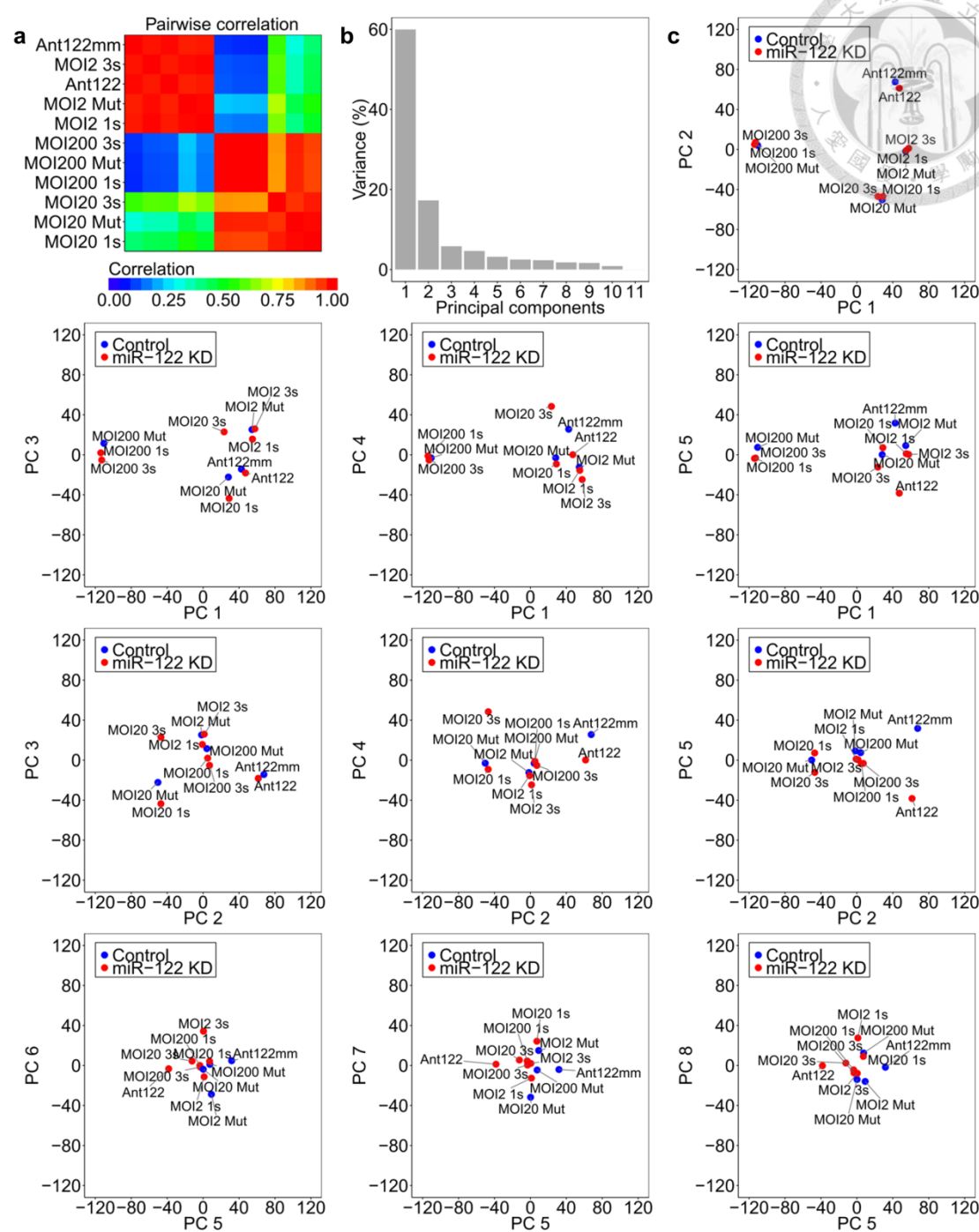


Figure 14. The pairwise correlation and PCA of the miR-122 KD dataset

a, The heatmap showing the pairwise correlation between samples from the miR-122 KD dataset. The sample orders were arranged by unsupervised hierarchical clustering.

MOI: multiplicity of infection, representing the viral vector dosage; Mut: mutant sponge containing no miR-122 sites; 3s: sponge with three miR-122 sites; 1s sponge with one miR-122 sites; Ant122: antagomir of miR-122; Ant122mm: the mismatched control for Ant122. **b**, The bar graph showing the variance explained by each PC. **c**, The scatter plots showing the mapping of the samples into the PC spaces. The control samples and the miR-122 KD samples were colored in blue and red, respectively.

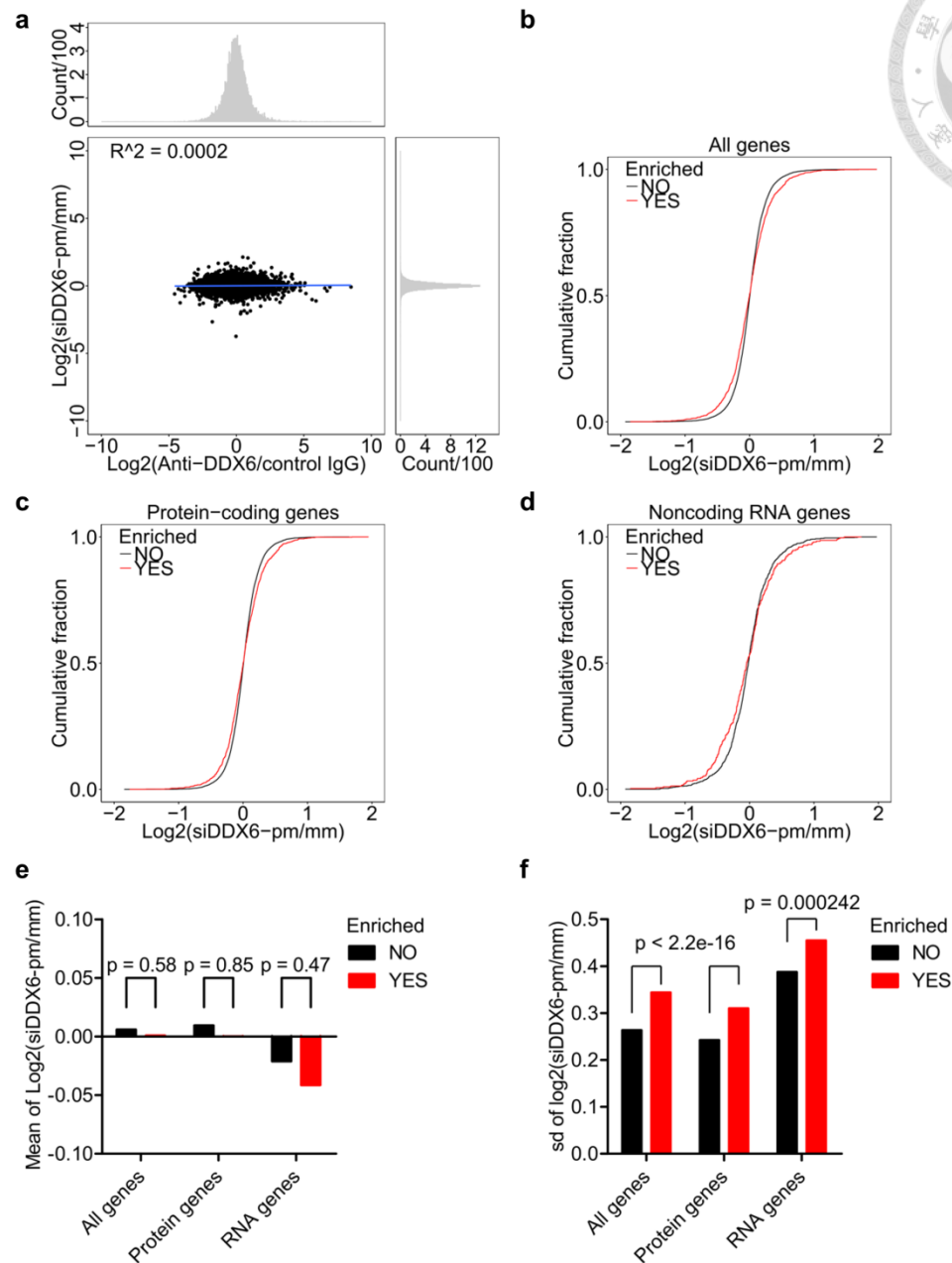
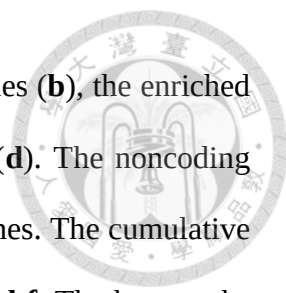


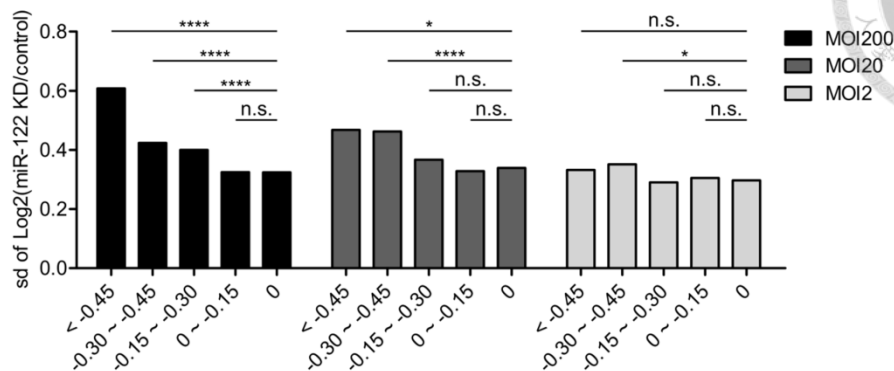
Figure 15. DDX6 regulates the expression noise of its bound RNA.

a, The scatter plot showing comparison between the anti-DDX6 RIP-seq dataset and the DDX6 KD RNA-seq dataset. The regression line was colored in blue. The histograms on the sides depict the log2-FED (top) and the log2-ERD_{RNA} (right). The genes with infinite fold enrichment were omitted from plotting and the regression analysis. **b-d**,

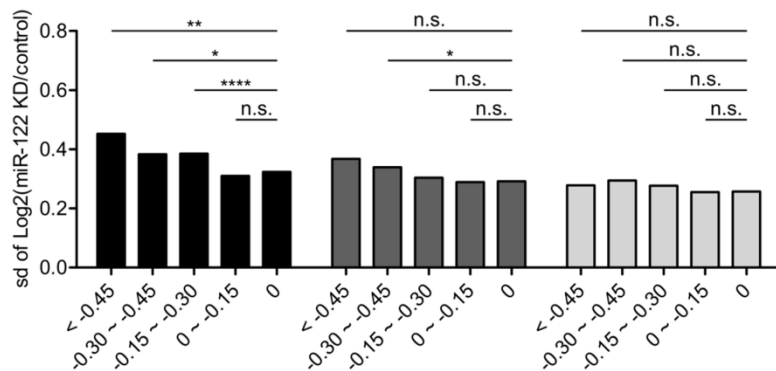


The cumulative fractions of the $\log_2\text{-ERD}_{\text{RNA}}$ of all the enriched genes (**b**), the enriched protein-coding genes (**c**), and the enriched noncoding RNA genes (**d**). The noncoding RNA category contains all biotypes except for the protein-coding genes. The cumulative fractions were stratified according to the binary binding status. **e and f**, The bar graphs depicting the means (**e**) and the standard deviations (**f**) of the $\log_2\text{-ERD}_{\text{RNA}}$. The biotypes of the enriched genes were stratified as in **b-d**. The difference in the means (**e**) and the standard deviations (**f**) were statistically tested with Welch's t test and F test, respectively. sd: standard deviation. **b-f**, The enriched genes were colored in red.

a miR-122 sponge 3s/mut



b miR-122 sponge 1s/mut



c Ant122/Ant122m

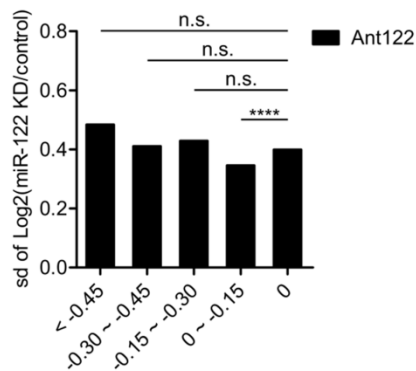


Figure 16. miR-122 KD increases variability in gene expression

a-c, The bar graphs showing the standard deviations of the $\log_2\text{-ERD}_{\text{RNA}}$ of the miR-122 targets and the non-targets under different miR-122 KD conditions. The miR-122 targets and the non-targets were stratified by the context+ score as in Fig. S9.

miR-122 KD was achieved by miR-122 sponge 3s (**a**), or 1s (**b**), or antagomir (**c**). For each condition (MOI and sponge target site number/antagomir), the standard deviations of the miR-122 targets, stratified by the context+ score, were compared to that of the non-targets (0 context+ score). The standard deviations were compared with F tests and the resulting P values were adjusted by Bonferroni correction. Notations for statistical significance: n.s.: not significant, $p > 0.05$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$.

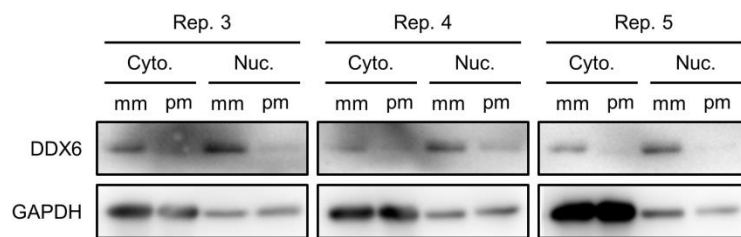


Figure 17. The quality assessment of the DDX6 KD quantitative MS samples

The WB confirming the DDX6 KD in both the cytoplasmic and nuclear fractions.

GAPDH was used as internal control.

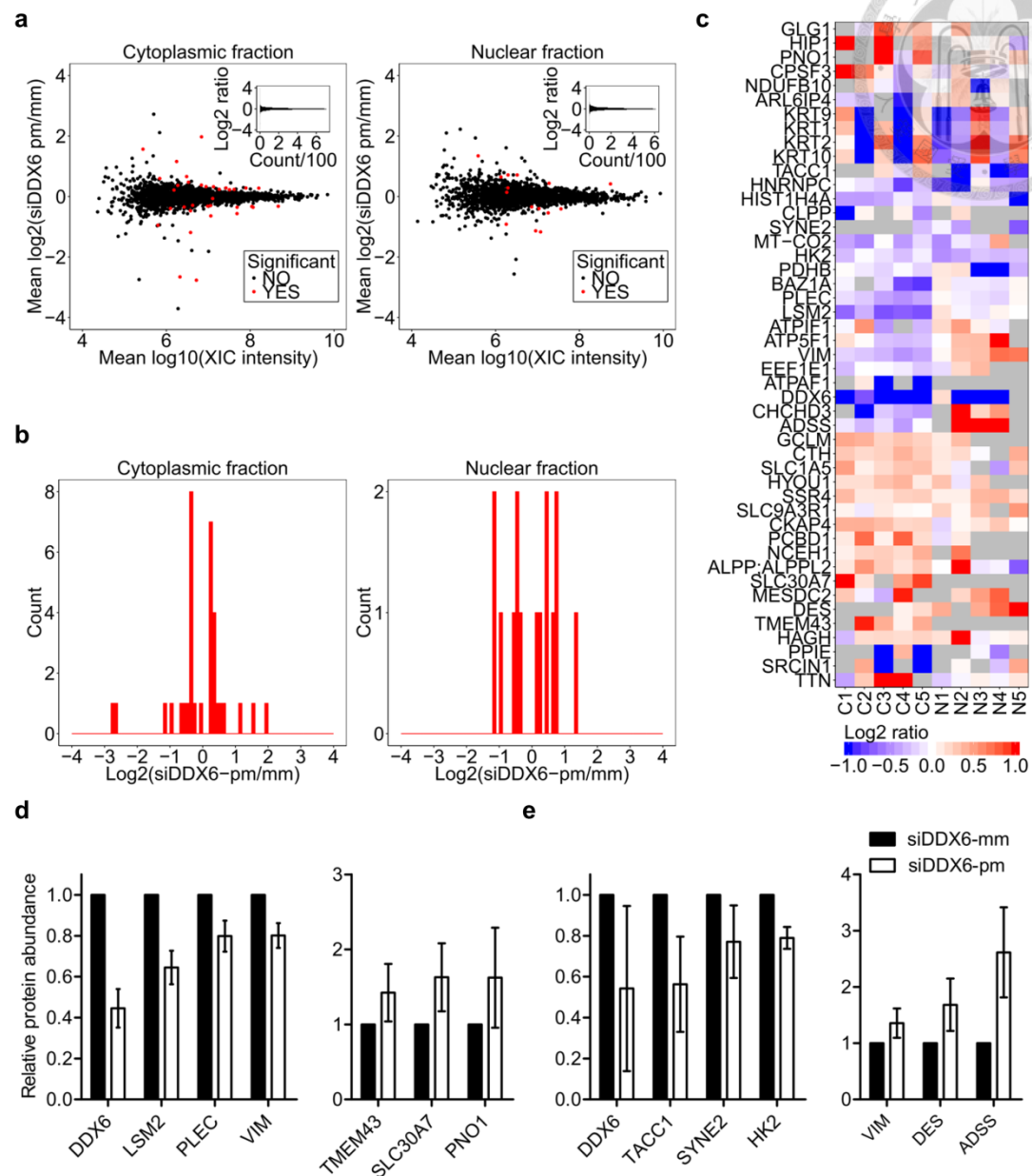


Figure 18. DDX6 has limited effect on the overall protein expression level.

a, The MA plots of the DDX6 KD quantitative MS dataset. The quantified proteins in the cytoplasmic fraction (left) and the nuclear fraction (right) were shown. The statistically significant proteins were determined by the MaxQuant software based on the significance B values and were colored in red. The Insets show the histograms of the

distributions of the log2-transformed protein expression ratio ($\log_2\text{-ERD}_{\text{protein}}$). **b**, The histogram showing the significant proteins in the cytoplasmic fraction (left) and the nuclear fraction (right). **c**, The heatmap showing the log2-transformed expression ratios of the significant proteins. The significant proteins from both fractions were plotted in the same heatmap. The log2-transformed expression ratios in both fractions and in all five biological replicates were plotted. NA values were filled with grey. **d**, The bar graphs showing the representative proteins in the cytoplasmic fraction (left) and the nuclear fraction (right). The representative proteins were selected based on statistical significance, smaller coefficient of variation, and larger difference from the hypothetical value. The bar heights and the error bars represent mean and standard deviation, respectively.

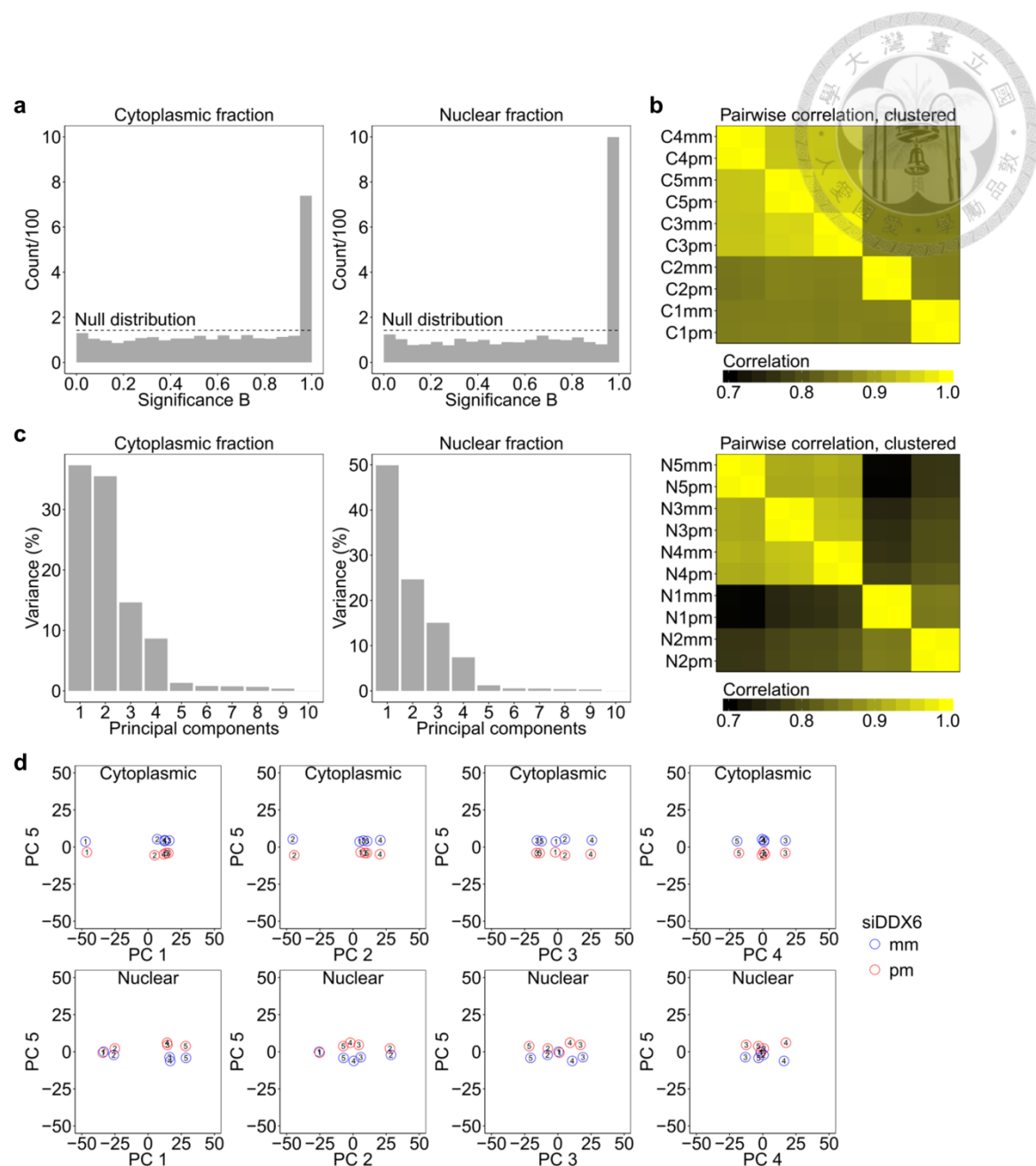


Figure 19. The statistical analyses of the quantitative MS dataset

a, The histogram showing the significance B distributions of the cytoplasmic fraction (left) and the nuclear fraction (right). The dashed lines represent the hypothetical distribution under null hypothesis. **b**, The heatmaps showing the pairwise correlations between the samples. The cytoplasmic fraction (top) and the nuclear fraction (bottom) are displayed separately for they formed distinct and lowly correlated clusters in the

unsupervised hierarchical clustering when computed together. The sample orders were determined by hierarchical clustering. **c and d**, PCA of the DDX6 KD quantitative MS dataset. **c**, The bar graphs showing the variance in the cytoplasmic fraction (left) and the nuclear fraction (right) explained by each PC. **d**, The scatter plots showing the mapping of the samples into the PC spaces. Samples of the cytoplasmic fraction and the nuclear fraction were plotted in the upper 4 panels and the lower 4 panels, respectively. The control samples (siDDX6-mm) were colored in blue while the DDX6 KD samples (siDDX6-pm) were colored in red. Replicates were labeled with numbers.

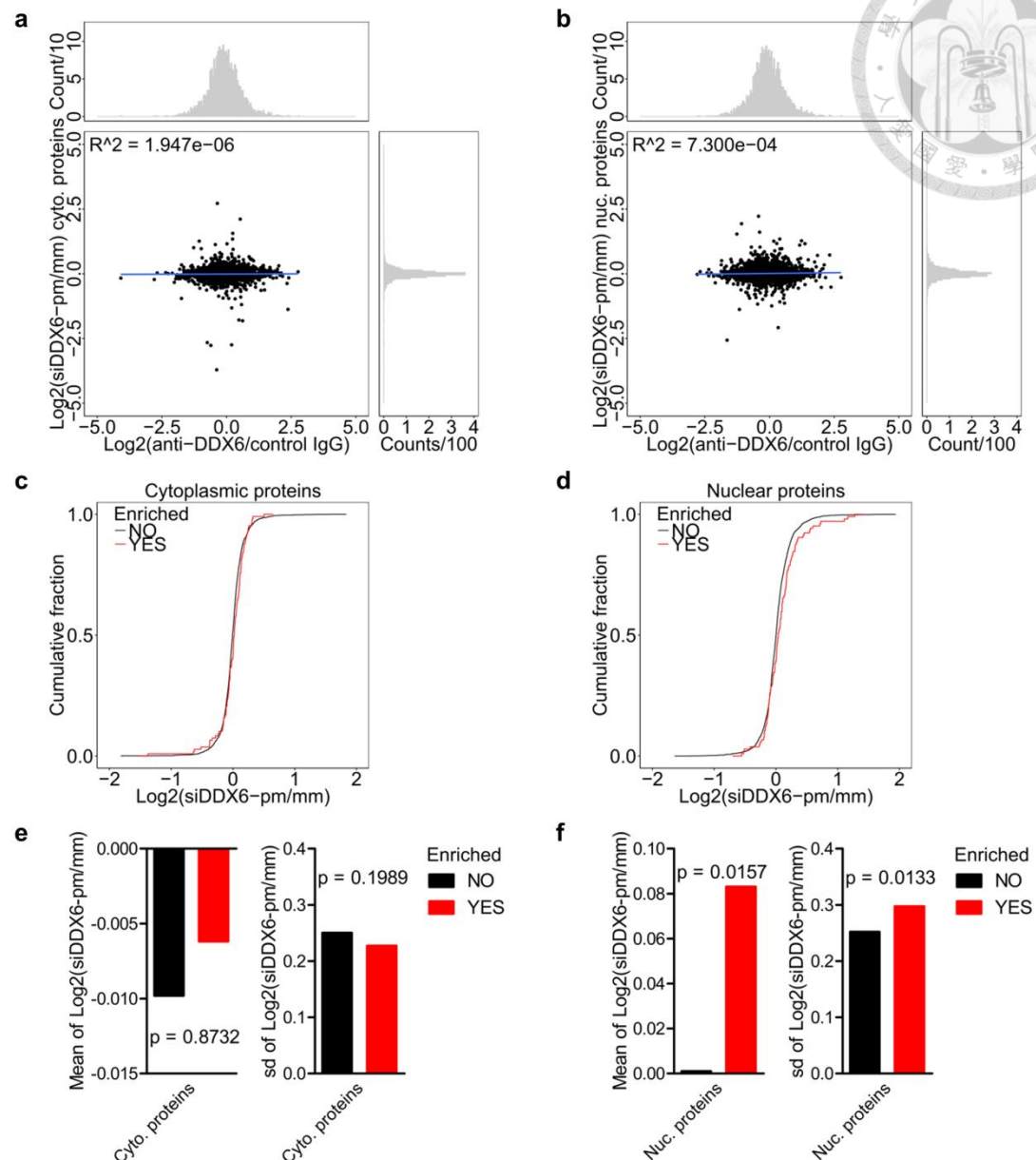
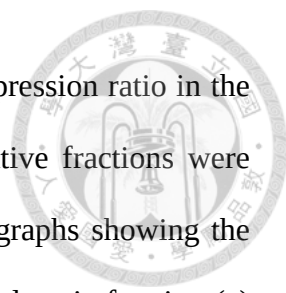


Figure 20. DDX6 regulates the protein expression noise of its bound mRNA.

a and b, The scatter plots showing the comparisons between the anti-DDX6 RIP-seq dataset and the cytoplasmic fraction (**a**), or the nuclear fraction (**b**), of the DDX6 KD quantitative MS dataset. The blue lines represent the regression lines. The histograms on the sides depict the distributions of the $\text{log}_2\text{-FED}$ (top) and the $\text{log}_2\text{-ERD}_{\text{protein}}$ (right). **c**



and d, The cumulative fractions of the log2-transformed protein expression ratio in the cytoplasmic fraction (**c**) and the nuclear fraction (**d**). The cumulative fractions were stratified according to the binary binding status. **e and f**, The bar graphs showing the means and the standard deviations of the log2-ERD_{protein} of the cytoplasmic fraction (**e**) and the nuclear fraction (**f**). Differences in the means and the standard deviations were statistically examined with Welch's t test and F test, respectively. sd: standard deviation. **c-f**, The enriched genes were colored in red.

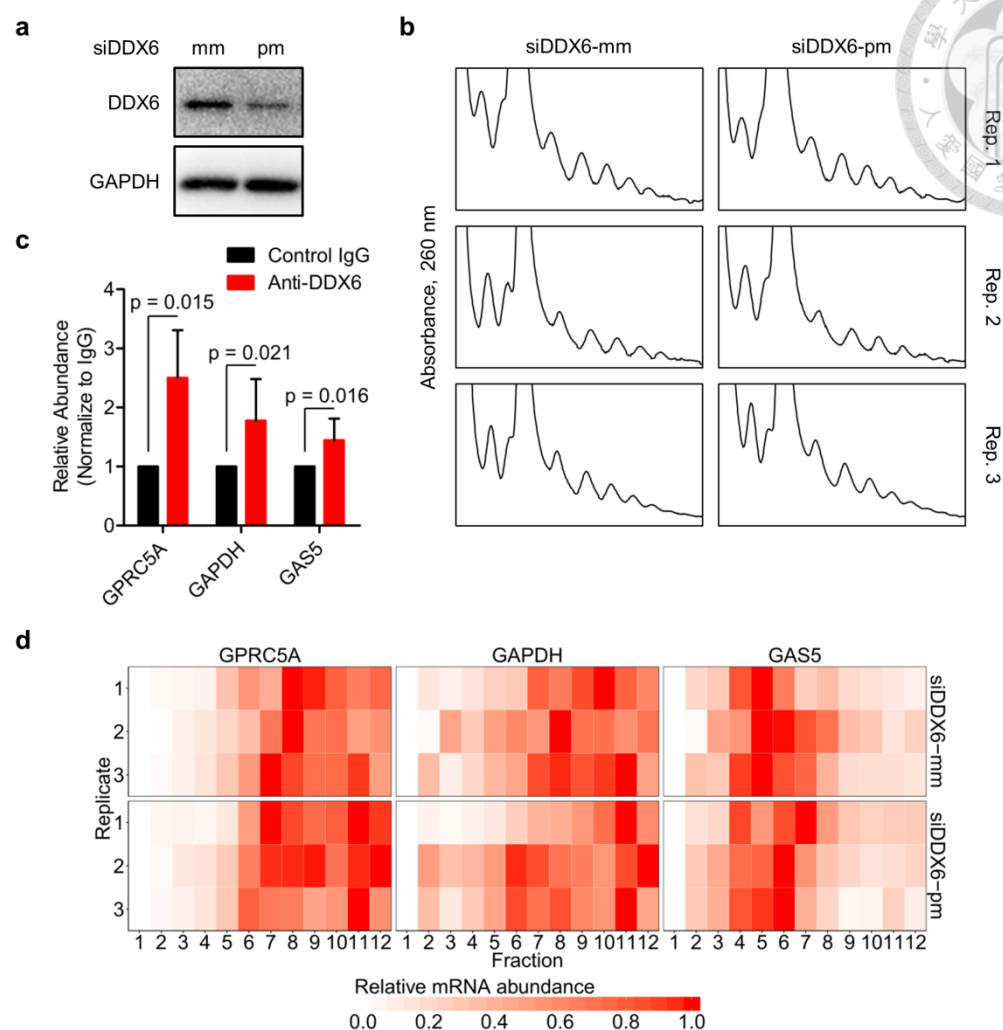
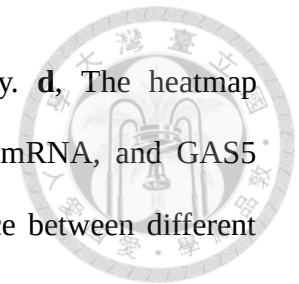


Figure 21. DDX6 depletion has limited effect on the polysome occupancy.

a, The representative western blot of DDX6 KD. GAPDH was used as internal control.

b, The polysome profiles in the control and the DDX6 KD cells. Three biological replicates were displayed. **a and b**, siDDX6-pm: siRNA with perfect-matching sequence to DDX6; siDDX6-mm: mismatched control siRNA. **c**, The bar graph showing the qPCR quantification of GPRC5A mRNA, GAPDH mRNA, and GAS5 lncRNA co-purified with anti-DDX6 IP relative to the control IgG. Statistical significance was assessed with right-tailed one sample t test. The bar heights and the

error bars represent means and standard deviations, respectively. **d**, The heatmap showing the qPCR quantification of GPRC5A mRNA, GAPDH mRNA, and GAS5 lncRNA in the 12 collected fractions. The relative RNA abundance between different fractions in each samples were normalized to [0,1]. Fraction 1-5: RNP; fraction 6-9: low molecular weight polysome; fraction 10-12: high molecular weight polysome.



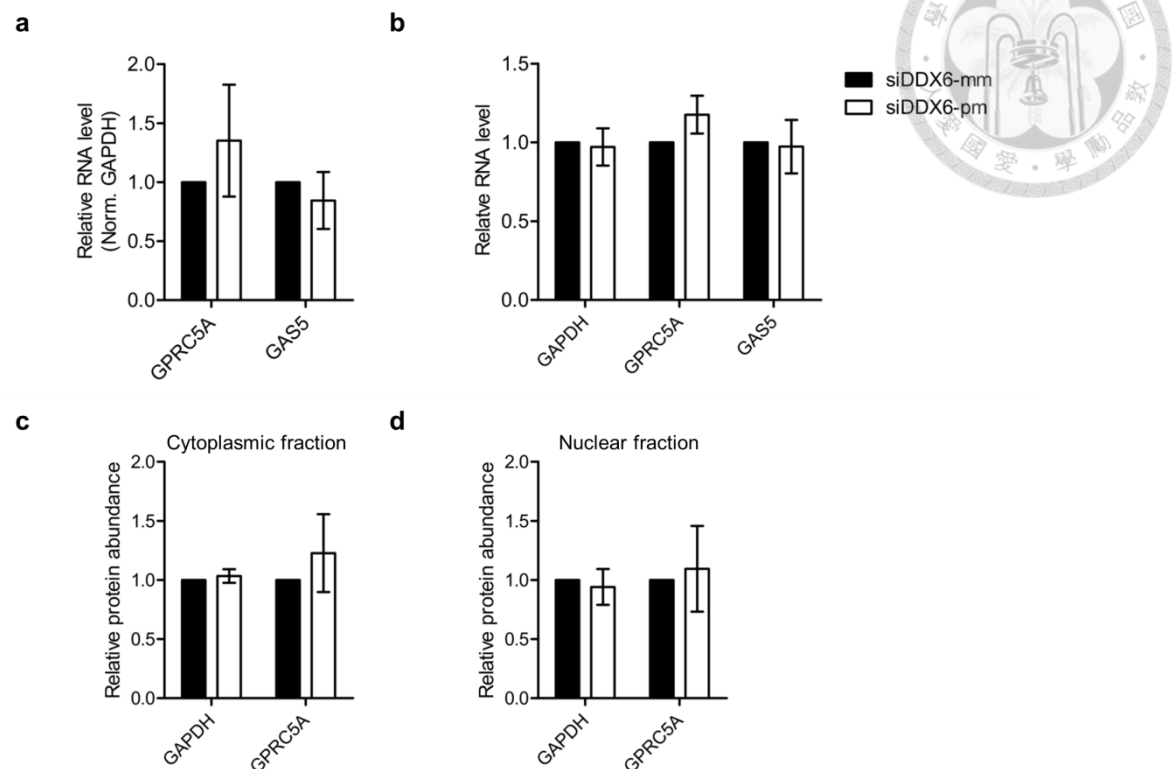


Figure 22. The expression analysis of GAPDH, GPRC5A, and GAS5 under DDX6 KD condition

a, The bar graph showing the relative RNA expression levels of GPRC5A and GAS5 under DDX6 KD measured by qPCR. **b**, The bar graph showing the RNA expression levels of GAPDH, GPRC5A, and GAS5 under DDX6 KD measured by RNA-seq. **c**, The bar graph showing the relative protein expression levels of GAPDH and GPRC5A in the cytoplasmic and the nuclear fractions under DDX6 KD measured by quantitative MS.

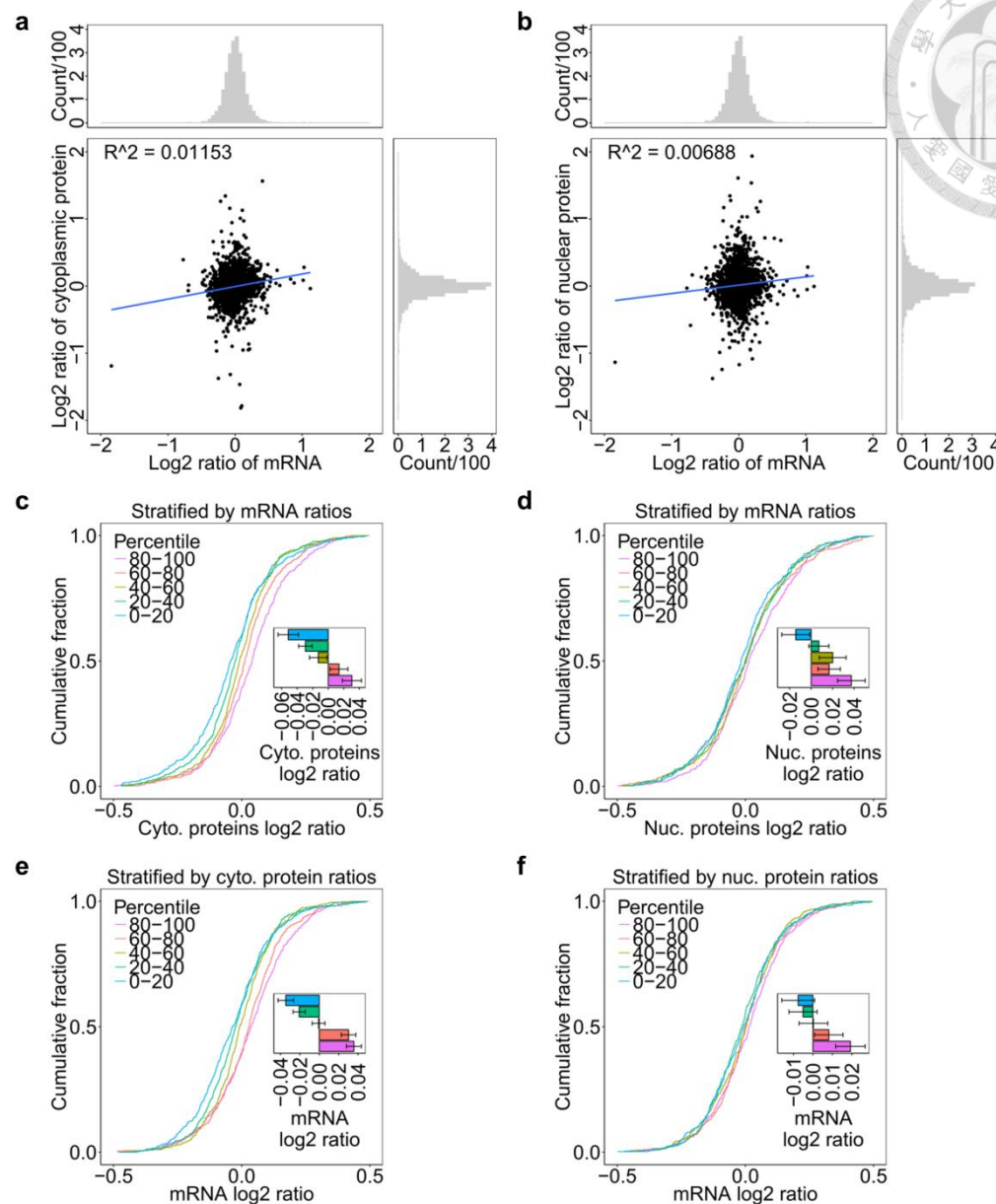
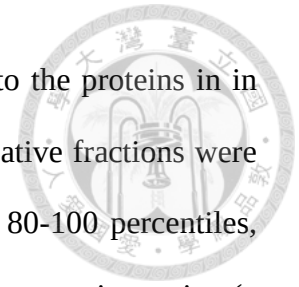


Figure 23. DDX6-regulated RNA and protein expression are weakly coupled.

a and b, The scatter plots showing comparisons between the DDX6 KD RNA-seq dataset and the cytoplasmic fraction (**a**), or the nuclear fraction (**b**), of the DDX6 KD quantitative MS dataset. The blue lines represent the regression lines. The histograms on the sides depict the distributions of the log2-ERD_{RNA} (top) and the log2-ERD_{protein} (right). **c-f**, The cumulative fractions of the log2-ERD_{protein} in the cytoplasmic fraction (**c**)

and the nuclear fraction (**d**), and the $\log_2\text{-ERD}_{\text{RNA}}$ corresponding to the proteins in in the cytoplasmic fraction (**e**) and the nuclear fraction (**f**). The cumulative fractions were stratified into 5 percentile groups: 0-20, 20-40, 40-60, 60-80, and 80-100 percentiles, according to the mRNA expression ratio (**c and d**) or the protein expression ratios (**e and f**). Insets: The bar graphs of the percentile groups. The bar lengths and the error bars represent the means and the standard errors of the mean, respectively.



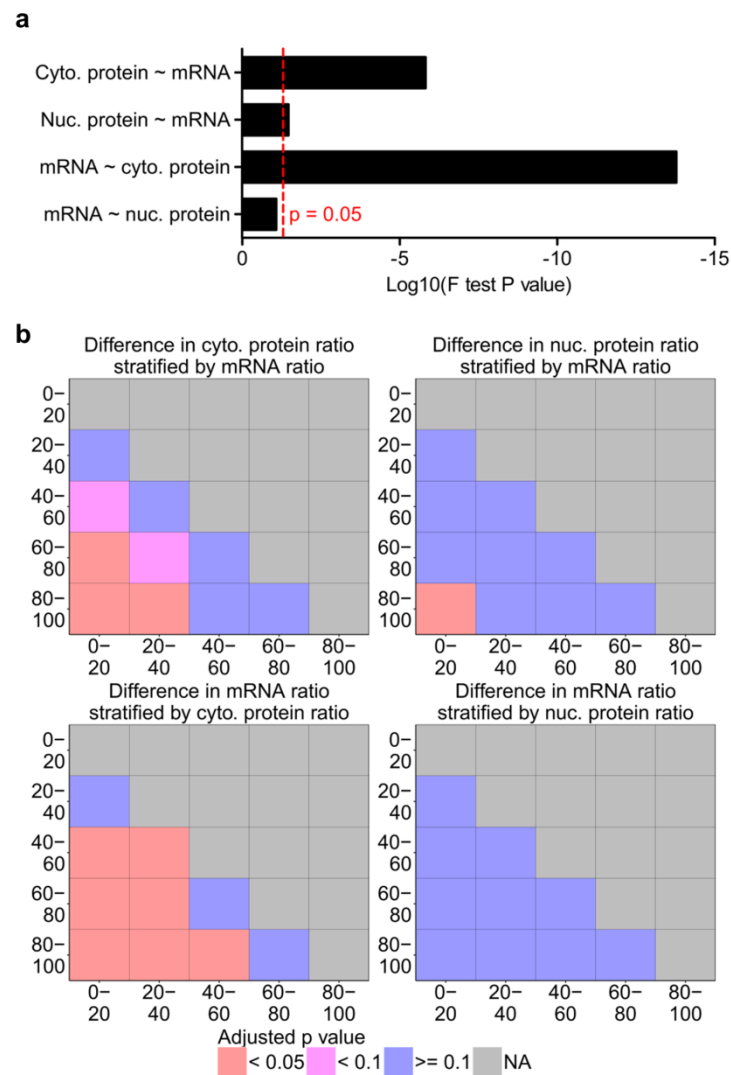


Figure 24. The ANOVA of the percentile groups

a, The bar graph showing the P values of the F test in ANOVA. The P value threshold at $p \leq 0.05$ was depicted as the dashed red line. Cyto. protein ~ mRNA: cytoplasmic protein expression ratio stratified by mRNA expression ratio, corresponding to Fig. 9c. The other groups follow the same naming pattern. **b**, Tile plots showing the post hoc t test results. All pairs were compared. P values were adjusted by the Bonferroni correction method.

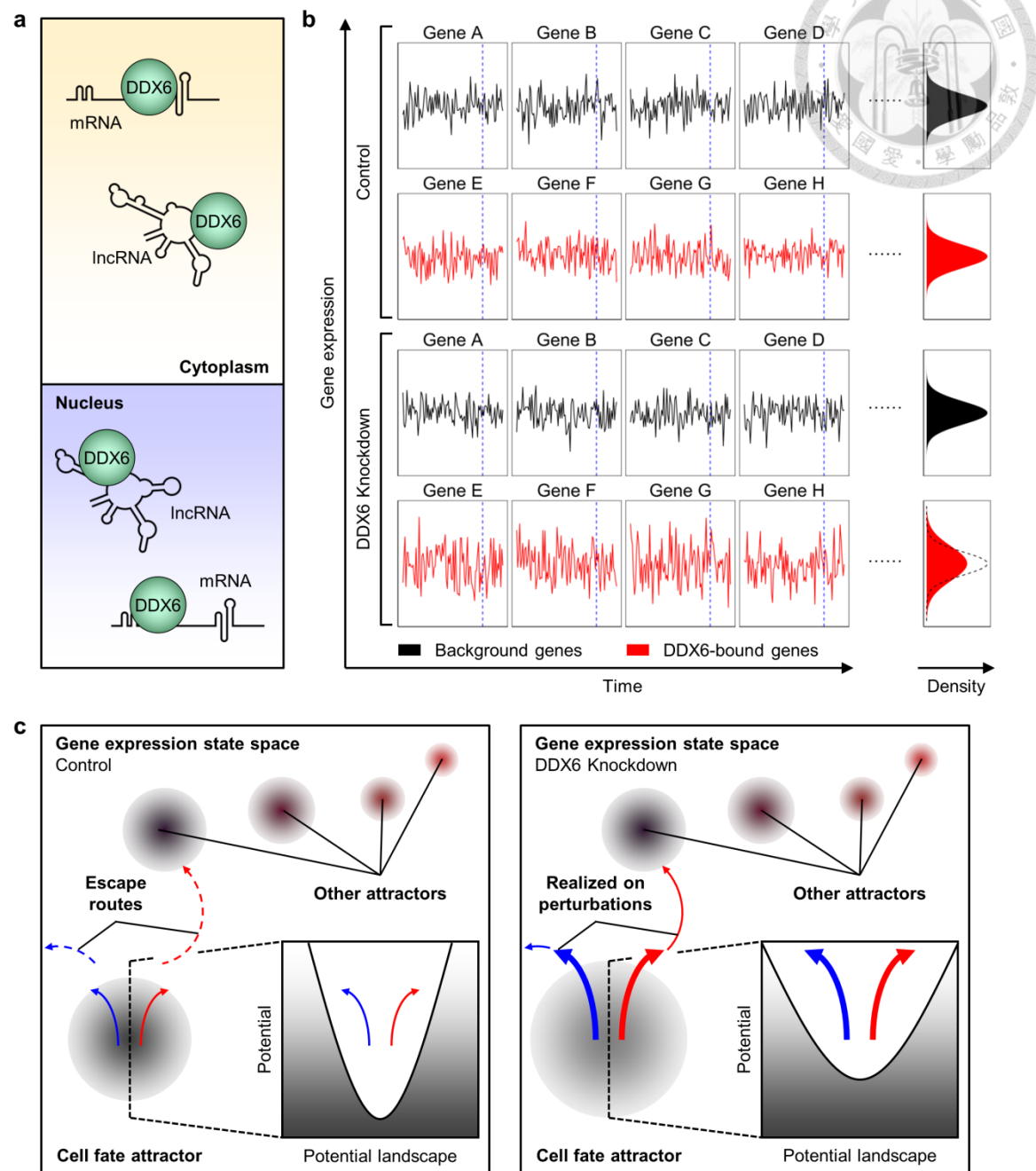
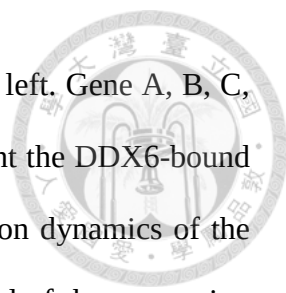


Figure 25. The graphical summary and the proposed model

a, The diagram showing DDX6 binds both cytoplasmic and nuclear RNA, including both mRNA and lncRNA. **b**, The noise reduction model for DDX6's role in post-transcriptional gene regulation. The gene expression dynamics of the background



genes and the enriched (DDX6-bound) genes were illustrated on the left. Gene A, B, C, and D represent the background genes. Gene E, F, G, and H represent the DDX6-bound genes. DDX6 KD likely increases the fluctuations in gene expression dynamics of the DDX6-bound genes. Increased fluctuation will result in wider spread of the expression distribution (depicted as the theoretical distributions on the right) when the overall gene expression is measured at a given time (depicted as the blue dashed lines). The increased uncertainty in gene expression will propagate to the expression ratios as we have observed in Fig. 5. c, The cell fate attractor stability model for DDX6's role in cellular physiology. Under the paradigm of systems biology, cell fates are understood as attractors in the high dimensional gene expression state space. Here, the gene expression state space is illustrated as a 2D plane whereas the cell fate attractors are depicted as the gradient-colored bubbles. DDX6 KD increases the noise in gene expression, thereby counteracting the attracting force of the cell fate attractor. This can be idealized as decreasing the “potential”, or the flattening of the “potential landscape”, of the attractor. The flattening of the potential landscape conceivably increases the escapability of the attractor (depicted by the curved arrows in red and blue, representing different routes of escape). Physiologically, this notion corresponds to the sensitization to the external stresses due to the loss of DDX6's activity.

7. References

Agarwal, V., Bell, G.W., Nam, J.W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife* 4.

Ascano, M., Jr., Mukherjee, N., Bandaru, P., Miller, J.B., Nusbaum, J.D., Corcoran, D.L., Langlois, C., Munschauer, M., Dewell, S., Hafner, M., *et al.* (2012). FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* 492, 382-386.

Audhya, A., Hyndman, F., McLeod, I.X., Maddox, A.S., Yates, J.R., 3rd, Desai, A., and Oegema, K. (2005). A complex containing the Sm protein CAR-1 and the RNA helicase CGH-1 is required for embryonic cytokinesis in *Caenorhabditis elegans*. *J. Cell Biol.* 171, 267-279.

Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. *Nature* 455, 64-71.

Barbee, S.A., Estes, P.S., Cziko, A.M., Hillebrand, J., Luedeman, R.A., Collier, J.M., Johnson, N., Howlett, I.C., Geng, C., Ueda, R., *et al.* (2006). Staufen- and FMRP-containing neuronal RNPs are structurally and functionally related to somatic P bodies. *Neuron* 52, 997-1009.

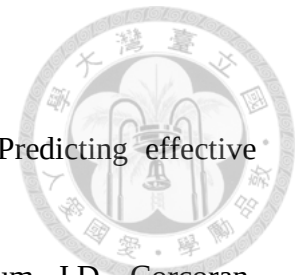
Benaglia, T., Chauveau, D., Hunter, D.R., and Young, D.S. (2009). mixtools: An R Package for Analyzing Mixture Models. *J. Stat. Softw.* 32, 29.

Bergkessel, M., and Reese, J.C. (2004). An essential role for the *Saccharomyces cerevisiae* DEAD-box helicase DHH1 in G1/S DNA-damage checkpoint recovery. *Genetics* 167, 21-33.

Boag, P.R., Atalay, A., Robida, S., Reinke, V., and Blackwell, T.K. (2008). Protection of specific maternal messenger RNAs by the P body protein CGH-1 (Dhh1/RCK) during *Caenorhabditis elegans* oogenesis. *J. Cell Biol.* 182, 543-557.

Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A.J. (2009). Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* 4, 484-494.

Braun, K.A., Vaga, S., Dombek, K.M., Fang, F., Palmisano, S., Aebersold, R., and Young, E.T. (2014). Phosphoproteomic analysis identifies proteins involved in



transcription-coupled mRNA decay as targets of Snf1 signaling. *Sci Signal* 7, ra64.

Braun, K.A., and Young, E.T. (2014). Coupling mRNA synthesis and decay. *Mol. Cell. Biol.* 34, 4078-4087.

Burden, C.J., Qureshi, S.E., and Wilson, S.R. (2014). Error estimates for the analysis of differential expression from RNA-seq count data. *PeerJ* 2, e576.

Cabili, M.N., Dunagin, M.C., McClanahan, P.D., Biaesch, A., Padovan-Merhar, O., Regev, A., Rinn, J.L., and Raj, A. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* 16, 20.

Carroll, J.S., Munchel, S.E., and Weis, K. (2011). The DExD/H box ATPase Dhh1 functions in translational repression, mRNA decay, and processing body dynamics. *J. Cell Biol.* 194, 527-537.

Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* 453, 544-547.

Chen, J.Y., Lin, J.R., Tsai, F.C., and Meyer, T. (2013). Dosage of Dyrk1a shifts cells within a p21-cyclin D1 signaling map to control the decision to enter the cell cycle. *Mol. Cell* 52, 87-100.

Cheng, Z., Collier, J., Parker, R., and Song, H. (2005). Crystal structure and functional analysis of DEAD-box protein Dhh1p. *RNA* 11, 1258-1270.

Chu, C.Y., and Rana, T.M. (2006). Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. *PLoS Biol.* 4, e210.

Collier, J., and Parker, R. (2005). General translational repression by activators of mRNA decapping. *Cell* 122, 875-886.

Collier, J.M., Tucker, M., Sheth, U., Valencia-Sanchez, M.A., and Parker, R. (2001). The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes. *RNA* 7, 1717-1727.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367-1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794-1805.

de Bruin, R.G., Shiue, L., Prins, J., de Boer, H.C., Singh, A., Fagg, W.S., van Gils, J.M., Duijs, J.M., Katzman, S., Kraaijeveld, A.O., *et al.* (2016). Quaking promotes monocyte differentiation into pro-atherogenic macrophages by controlling pre-mRNA splicing and gene expression. *Nature communications* 7, 10846.

Denzler, R., Agarwal, V., Stefano, J., Bartel, D.P., and Stoffel, M. (2014). Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Mol. Cell* 54, 766-776.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., *et al.* (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775-1789.

Dori-Bachash, M., Shema, E., and Tirosh, I. (2011). Coupled evolution of transcription and mRNA degradation. *PLoS Biol.* 9, e1001106.

Dutta, A., Zheng, S., Jain, D., Cameron, C.E., and Reese, J.C. (2011). Intermolecular interactions within the abundant DEAD-box protein Dhh1 regulate its activity in vivo. *J. Biol. Chem.* 286, 27454-27470.

Eichhorn, S.W., Guo, H., McGeary, S.E., Rodriguez-Mias, R.A., Shin, C., Baek, D., Hsu, S.H., Ghoshal, K., Villen, J., and Bartel, D.P. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol. Cell* 56, 104-115.

Encode Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74.

Ernault-Lange, M., Baconnais, S., Harper, M., Minshall, N., Souquere, S., Boudier, T., Benard, M., Andrey, P., Pierron, G., Kress, M., *et al.* (2012). Multiple binding of repressed mRNAs by the P-body protein Rck/p54. *RNA* 18, 1702-1715.

Fischer, N., and Weis, K. (2002). The DEAD box protein Dhh1 stimulates the decapping enzyme Dcp1. *EMBO J.* 21, 2788-2797.

Geissler, R., Simkin, A., Floss, D., Patel, R., Fogarty, E.A., Scheller, J., and Grimson, A. (2016). A widespread sequence-specific mRNA decay pathway mediated by hnRNPs A1 and A2/B1. *Genes Dev.* 30, 1070-1085.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466, 835-840.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., *et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129-141.

Haimovich, G., Medina, D.A., Causse, S.Z., Garber, M., Millan-Zambrano, G., Barkai, O., Chavez, S., Perez-Ortin, J.E., Darzacq, X., and Choder, M. (2013). Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell* 153, 1000-1011.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., *et al.* (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 22, 1760-1774.

Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 153, 654-665.

Hu, C.W., Lin, M.H., Huang, H.C., Ku, W.C., Yi, T.H., Tsai, C.F., Chen, Y.J., Sugiyama, N., Ishihama, Y., Juan, H.F., *et al.* (2012). Phosphoproteomic analysis of *Rhodospseudomonas palustris* reveals the role of pyruvate phosphate dikinase phosphorylation in lipid production. *J. Proteome Res.* 11, 5362-5375.

Huang, S. (2009a). Non-genetic heterogeneity of cells in development: more than just noise. *Development* 136, 3853-3862.

Huang, S. (2009b). Reprogramming cell fates: reconciling rarity with robustness. *BioEssays* 31, 546-560.

Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D.E. (2005). Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* 94, 128701.

Huang, S., Guo, Y.P., May, G., and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* 305, 695-713.

Iio, A., Takagi, T., Miki, K., Naoe, T., Nakayama, A., and Akao, Y. (2013). DDX6 post-transcriptionally down-regulates miR-143/145 expression through host gene NCR143/145 in cancer cells. *Biochim. Biophys. Acta* 1829, 1102-1110.

Jangra, R.K., Yi, M., and Lemon, S.M. (2010). DDX6 (Rck/p54) is required for efficient hepatitis C virus replication but not for internal ribosome entry site-directed translation. *J. Virol.* 84, 6810-6824.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36.

Kruk, J.A., Dutta, A., Fu, J., Gilmour, D.S., and Reese, J.C. (2011). The multifunctional Ccr4-Not complex directly promotes transcription elongation. *Genes Dev.* 25, 581-593.

Lebedeva, S., Jens, M., Theil, K., Schwanhauser, B., Selbach, M., Landthaler, M., and Rajewsky, N. (2011). Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol. Cell* 43, 340-352.

Lee, S., Kopp, F., Chang, T.C., Sataluri, A., Chen, B., Sivakumar, S., Yu, H., Xie, Y., and Mendell, J.T. (2016). Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* 164, 69-80.

Li, Q., Wennborg, A., Aurell, E., Dekel, E., Zou, J.Z., Xu, Y., Huang, S., and Ernberg, I. (2016). Dynamics inside the cancer cell attractor reveal cell heterogeneity, limits of stability, and escape. *Proc. Natl. Acad. Sci. USA* 113, 2672-2677.

Lu, Z., Guan, X., Schmidt, C.A., and Matera, A.G. (2014). RIP-seq analysis of eukaryotic Sm proteins identifies three major categories of Sm-containing ribonucleoproteins. *Genome Biol.* 15, R7.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet.journal* 17.

Mercer, T.R., and Mattick, J.S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* 20, 300-307.

Minshall, N., Thom, G., and Standart, N. (2001). A conserved role of a DEAD box helicase in mRNA masking. *RNA* 7, 1728-1742.

Mitchell, S.F., Jain, S., She, M., and Parker, R. (2013). Global analysis of yeast mRNPs. *Nat. Struct. Mol. Biol.* 20, 127-133.

Mukherjee, N., Corcoran, D.L., Nusbaum, J.D., Reid, D.W., Georgiev, S., Hafner, M., Ascano, M., Jr., Tuschl, T., Ohler, U., and Keene, J.D. (2011). Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol. Cell* 43, 327-339.

Nakamura, A., Amikura, R., Hanyu, K., and Kobayashi, S. (2001). Me31B silences translation of oocyte-localizing RNAs through the formation of cytoplasmic RNP complex during *Drosophila* oogenesis. *Development* 128, 3233-3242.

Navarro, R.E., Shim, E.Y., Kohara, Y., Singson, A., and Blackwell, T.K. (2001). *cgh-1*, a conserved predicted RNA helicase required for gametogenesis and protection from physiological germline apoptosis in *C. elegans*. *Development* 128, 3221-3232.

Nicklas, S., Okawa, S., Hillje, A.L., Gonzalez-Cano, L., Del Sol, A., and Schwamborn, J.C. (2015). The RNA helicase DDX6 regulates cell-fate specification in neural stem cells via miRNAs. *Nucleic Acids Res.* 43, 2638-2654.

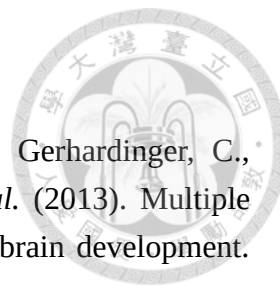
Pettersson, O.J., Aagaard, L., Andrejeva, D., Thomsen, R., Jensen, T.G., and Damgaard, C.K. (2014). DDX6 regulates sequestered nuclear CUG-expanded DMPK-mRNA in dystrophin myotonia type 1. *Nucleic Acids Res.* 42, 7186-7200.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135, 216-226.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* 2, 1896-1906.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81, 145-166.

Rouhana, L., Shibata, N., Nishimura, O., and Agata, K. (2010). Different requirements for conserved post-transcriptional regulators in planarian regeneration and stem cell



maintenance. *Dev. Biol.* **341**, 429-443.

Sauvageau, M., Goff, L.A., Lodato, S., Bonev, B., Groff, A.F., Gerhardinger, C., Sanchez-Gomez, D.B., Hacisuleyman, E., Li, E., Spence, M., *et al.* (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* **2**, e01749.

Schmiedel, J.M., Klemm, S.L., Zheng, Y., Sahay, A., Bluthgen, N., Marks, D.S., and van Oudenaarden, A. (2015). Gene expression. MicroRNA control of protein expression noise. *Science* **348**, 128-132.

Smillie, D.A., and Sommerville, J. (2002). RNA helicase p54 (DDX6) is a shuttling protein involved in nuclear assembly of stored mRNP particles. *J. Cell Sci.* **115**, 395-407.

Sun, M., Schwalb, B., Pirkl, N., Maier, K.C., Schenk, A., Failmezger, H., Tresch, A., and Cramer, P. (2013). Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Mol. Cell* **52**, 52-62.

Sweet, T., Kovalak, C., and Collier, J. (2012). The DEAD-box protein Dhh1 promotes decapping by slowing ribosome movement. *PLoS Biol.* **10**, e1001342.

Tanaka, K.J., Ogawa, K., Takagi, M., Imamoto, N., Matsumoto, K., and Tsujimoto, M. (2006). RAP55, a cytoplasmic mRNP component, represses translation in *Xenopus* oocytes. *J. Biol. Chem.* **281**, 40096-40106.

Tichon, A., Gil, N., Lubelsky, Y., Havkin Solomon, T., Lemze, D., Itzkovitz, S., Stern-Ginossar, N., and Ulitsky, I. (2016). A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nature communications* **7**, 12209.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46-53.

Wang, Y., Arribas-Layton, M., Chen, Y., Lykke-Andersen, J., and Sen, G.L. (2015). DDX6 Orchestrates Mammalian Progenitor Function through the mRNA Degradation and Translation Pathways. *Mol. Cell* **60**, 118-130.

Westmoreland, T.J., Olson, J.A., Saito, W.Y., Huper, G., Marks, J.R., and Bennett, C.B. (2003). DHH1 regulates the G1/S-checkpoint following DNA damage or BRCA1 expression in yeast. *J. Surg. Res.* 113, 62-73.

Wong, S.F., Agarwal, V., Mansfield, J.H., Denans, N., Schwartz, M.G., Prosser, H.M., Pourquie, O., Bartel, D.P., Tabin, C.J., and McGlinn, E. (2015). Independent regulation of vertebral number and vertebral identity by microRNA-196 paralogs. *Proc. Natl. Acad. Sci. USA* 112, E4884-4893.

Yoon, J.H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., Huarte, M., Zhan, M., Becker, K.G., and Gorospe, M. (2012). LincRNA-p21 suppresses target mRNA translation. *Mol. Cell* 47, 648-655.

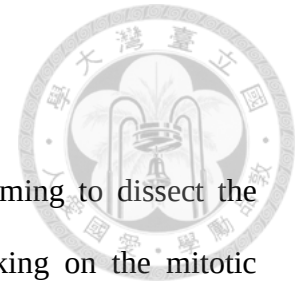




Chapter 4

An Unfinished Quest





1. **Abstract**

This chapter includes the data from an unfinished project aiming to dissect the molecular mechanism of the nuclear entry of DDX6 by hitchhiking on the mitotic chromosomes. First, we established a semi-automatic image analysis workflow to facilitate the analysis of the nucleocytoplasmic distribution of different constructs. Second, we analyzed the potential motifs in the DDX6 CTD that could potentially facilitate the hitchhiking mechanism.

2. **Introduction**

The DDX6 subfamily proteins were best known for their cytoplasmic localization and functions. However, since the early research, they have also been observed in the cell nuclei in various biological models (Haimovich et al., 2013; Iio et al., 2013; Kawahara et al., 2014; Kruk et al., 2011; Lodomery et al., 1997; Minshall et al., 2001; Pettersson et al., 2014; Smillie and Sommerville, 2002; Yoshida-Kashikawa et al., 2007). Previously, it was proposed that the nuclear localization of the DDX6 homologs is driven by the importin- α/β pathway owing to a putative nuclear localization signal (NLS) in the DDX6 N-terminal domain (NTD) (Smillie and Sommerville, 2002; Weston and Sommerville, 2006). However, in the Chap. 2 of this study, we demonstrated that instead of being a direct cargo of the importin- α/β pathway, DDX6 can enter the cell nuclei by two distinct mechanisms, namely the “hitchhiking” mechanism and the “piggyback” mechanism. It is clear that the piggyback transportation of DDX6 is facilitated by the NLS in the DDX6 transporter 4E-T. However, the molecular mechanism of the hitchhiking mechanism remains unclear. To elucidate this, we utilized the imaging-based analysis to quantify the nuclear localization ability and the dynamics

of the DDX6 CTD. Furthermore, we adopted a comparative approach in hope to identify the motifs that could potentially facilitate nuclear entry.



3. Materials and methods

3.1. Molecular cloning

The constructs appeared in the Chap. 2 were described in detail there. The YFP-ZP54 and the YFP-CBC1 were cloned and inserted in to the pEYFP-C1 vector by C. C. Yeh (C. C. Yeh's master thesis) and C. W. Huang (Y. H. Wang's master thesis). The CTD of Zp54 and CBC-1 were amplified from the plasmids using PCR with specific primers and then inserted back into the pEYFP-C1 vector. HIST1H2BJ (H2B) was cloned from the complementary (cDNA) prepared from the HeLa cells. The Dhh1p and the Me31B were cloned from the *S. cerevisiae* genomic DNA (gDNA) and the cDNA prepared from the *D. melanogaster* ovaries, respectively. The gDNA of yeast and the *Drosophila* ovaries were kind gifts from Prof. Wei-Pang Huang and Prof. Chao-Ti Ting in the Department of Life Science, National Taiwan University.

3.2. Cell culture and plasmid transfection

Cell culture and plasmid transfection were described in more detail in Chap. 1. For imaging, the transfected cells were process at 1 day, 2 day, or 3 day after transfection, as specified in the figure legends.

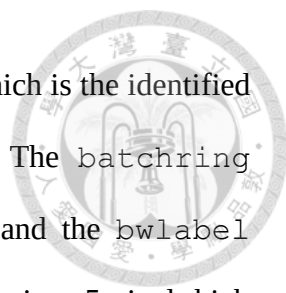
3.3. Sample preparation and microscopy

For F-actin staining, the cells were fixed with 4% paraformaldehyde (Sigma-Aldrich) in phosphate-buffered saline (PBS) for 10 min, penetrated with 0.25% Triton X-100 (Sigma-Aldrich) in PBS for 5 min, blocked with 2% bovine serum

albumin (BSA, Sigma-Aldrich)/0.1% Triton X-100 in PBS for 30 min prior to staining. We then stained the cells with Alexa Fluor 568 phalloidin (0.05 Unit per well of the 12-well plates; purchased from Thermo Fisher Scientific) for 30 min. The cells were then stained with Hoechst at 1 $\mu\text{g/mL}$ concentration. For live cell imaging, the cells were simply stained with Hoechst prior to imaging. For imaging, we used the DMI300 B (Leica) wide field fluorescent microscope as described in Chap. 2.

3.4. Image analysis

We built the “batchimageR” utility on the published “EBImage” package in R (Pau et al., 2010). The batchimageR is in essence a wrapper of the functions in the EBImage, which provides a toolkit for basic image analysis but was not optimized for batch image processing. For example, the `batchread` and the `batchexport` functions were custom scripts that drive the `readImage` and the `writeImage` functions of the EBImage for batch image input/output. Notably, however, the batch image signal normalization function (`batchcontrast2`) included in the batchimageR was based on several basic functions included in the R environment. In essence, `batchcontrast2` simply scales the signal distribution in range [0,1]. For image object segmentation, the `batchseg` and the `batchring` functions were applied. The `batchseg` function drives the `thresh`, the `opening`, the `fillHull` and the `bwlabel` functions for nuclear segmentation. Briefly, the `thresh` function identifies the nuclei by adaptive thresholding. The `opening` and the `fillHull` functions shape and fill the holes in the identified nuclei. The `bwlabel` function converts the processed images into binary masks. Also, the `batchseg` function drives the `propagate`, the `fillHull`, and the `bwlabel` functions to segment the cells and



creates the cellular masks. The `propagate` function uses seeds, which is the identified nuclei in this case, to find boundaries in the adjacent regions. The `batchring` function, also using the nuclear masks, commands the `dilate` and the `bwlabel` functions to create the expanded nuclear masks. The default setting gives 5 pixel thick cytoplasmic rings. For more details about the `batchimageR` utility, please refer to the original codes and the user's manual. For details about the `EBImage` functions, please refer to the original paper and the user's manual. Following the image object segmentation, we used the masks to mark the different regions of interest in the normalized images. This and the following steps were carried out using ImageJ. The cells', nuclei's, and the expanded nuclei's borders were manually selected using the Wand (selection) tool. During this step, we avoided the multi-nucleated cells, the cells with failed segmentation, and the dead cells. By failed segmentation, we mean the nuclei or the cells that should be, but failed to be, separated, and the cell borders that were fragmented due to low signal intensities. Moreover, the manual supervision improved the data quality. The signals in the selected regions were then measured by the built-in measurement function in ImageJ. We filtered out the cells whose mean-signals in the nucleus and the cytoplasmic ring were both lower the whole-image mean. The resulting dataset were subsequently analyzed using R.

3.5. Protein sequence and structural analysis

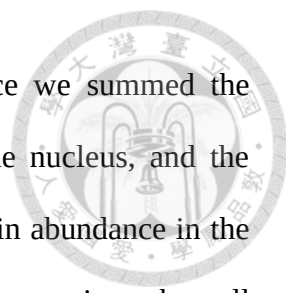
The protein sequences of the DDX6 homologs were retrieved from the protein database of NCBI. We used the NCBI Reference Sequence entries: the human DDX6 (NP_004388), the mouse Ddx6 (NP_031867), the *Xenopus* Xp54 (NP_001072584), the *Drosophila* Me31B (NP_523533), the *C. elegans* CGH-1 (NP_498646), and the *S.*

cerevisiae Dhh1p (NP_010121). For the zebrafish Zp54 and the *D. japonica* CBC-1 we used the sequences described in C. C. Yeh's master thesis and Y. H. Wang's master thesis. The sequence alignment and the phylogenetic trees were generated with the CLC Sequence Viewer 6 (CLC bio). The crystal structure 4CT5 was retrieved from the PDB database and visualized as described in Chap. 2.

4. **Results**

4.1. **A semi-automatic workflow for the imaging-based analysis of the nucleocytoplasmic distribution**

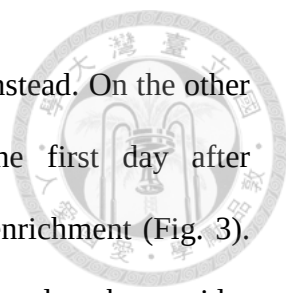
The population-wide heterogeneity in the nucleocytoplasmic distribution of YFP-PKM-DDX6 CTD sparked our interest (Chap. 3). We wondered if the cells exhibiting the nuclear-positive (type 2 and 3 combined) phenotype increases over time as more cell cycles lapse. To study this, we first tried to elevate our sample size. We established a semi-automatic workflow for image analysis based on R and ImageJ, two open source platforms (Fig. 1a). We established an automated image processing workflow based on several custom scripts and the R package “EBImage”, which provides the basic tool kit for image analysis. To command the EBImage more efficiently, we wrote several custom scripts to wrap up the key functions in the EBImage. We termed this utility the “batchimageR”. The batchimageR normalizes the image signal distributions and segments the image objects, including the cells or the nuclei (Fig. 1b). For nuclear segmentation, we used the Hoechst signals. Notably, the nuclear segmentation was iterated to create two masks (Fig. 1c), with one fits to the Hoechst signals (the nuclear area) and the other expands outward (the expanded nuclear area). For cellular segmentation, we used the nuclei mask as seeds which allowed the EBImage algorithm to expand outward to find the cells' borders



from the cytoskeleton staining. To estimate the protein abundance we summed the signal intensities of each pixel for each area, namely, the cell, the nucleus, and the expanded nucleus. We simply used subtraction to estimate the protein abundance in the cytoplasm, the cytoplasmic ring, and the nucleus (Fig. 1d). However, since the cell shape of the HeLa cells were not controlled in our experiments, the cellular area became a major factor in determining the estimated abundance of the cellular or the cytoplasmic protein abundances. The non-uniformity of the cell body thickness also affected the estimation. Thus, in the following sections, we used the cytoplasmic ring to estimate the cytoplasmic protein abundance as described previously (Spencer et al., 2013). This could circumvent the cell shape problem and was less affected by the cell area. With this, we indexed the nucleocytoplasmic distribution with the ratio of the mean signal intensities in the cytoplasmic ring versus the nucleus (C.R./Nuc mean-intensity ratio).

4.2. Time-dependent phenotypic shift in the cells expressing YFP-PKM-DDX6 CTD

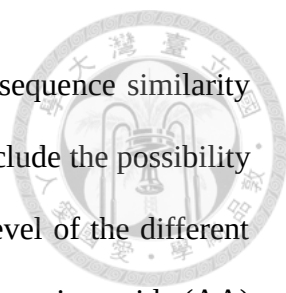
We then applied this imaging analysis workflow to YFP-PKM-DDX6 CTD. We obtained an image dataset with the HeLa cells expressing, YFP-PKM, YFP-DDX6 CTD, and YFP-PKM-DDX6 CTD over a three day time course. We found that YFP-PKM maintained a similar cytoplasmic distribution over the three days, whereas YFP-DDX6 CTD maintained a similar level in the first two days (Fig. 2). We attributed the change in the inferred nucleocytoplasmic distribution to the decline in the expression level. That is, on day 3, YFP-DDX6 CTD was expressed at significantly lower level overall, and thus the signal strength difference in the nucleus and the cytoplasm was diluted by the background. Interestingly, we also observed the nucleocytoplasmic heterogeneity in the cells expressing YFP-DDX6 CTD. A small fraction of the cells did not exhibit



explicit nuclear localization, and were rather uniformly distributed instead. On the other hand, YFP-PKM-DDX6 CTD was majorly cytoplasmic on the first day after transfection (Fig. 2), with a minor fraction showing clear nuclear enrichment (Fig. 3). On day 2 after transfection, this bimodal distribution disappeared, replaced by a wider distribution centered on the uniform nucleocytoplasmic distribution (the type 2 phenotype, Fig. 2 and 3) that persisted to the third day after transfection. These results indicated that despite YFP-DDX6 CTD and YFP-PKM-DDX6 CTD appeared to bind similarly to the mitotic chromosomes and to enter the cell nuclei through the same hitchhiking mechanism (Chap. 2), their dynamics were not identical.

4.3. The nucleocytoplasmic localization of the DDX6 homologs CTD

To further explore the nature of the DDX6 CTD in nuclear localization, we compared the nucleocytoplasmic localization of the CTD of several DDX6 homologs, including the zebrafish Zp54, the *Drosophila* Me31B, the *D. japonica* DjCBC-1, and the *S. cerevisiae* Dhh1p (Fig. 5a). Since DjCBC-1 and Dhh1p contain longer C-terminal disorder region (C-terminal tail, CT), we constructed the CT deletion mutants (CTD- Δ CT) in DjCBC-1 and DDX6 to assess the role of the CT in nuclear localization (Fig. 5a). As a result, all constructs were observed with the nuclear-enriched phenotype as well as more the uniformly distributed phenotype (Fig. 5b-h). We then applied the image analysis workflow to quantify the nucleocytoplasmic distribution of the DDX6 homologs CTD and the Δ CT mutants (Fig. 6a and 7). We found that the Δ CT mutation decreased the nuclear enrichment (Fig. 6a and 7) but did not abolish the nuclear localization like the CTD- Δ 1 and CTD- Δ 2 (Chap. 2). On the other hand, we compared the sequence similarity data (Fig. 6b) with the nucleocytoplasmic distribution. As a



result, we found that the nuclear-enrichment is correlated with the sequence similarity relative to the human DDX6 (Fig. 6c). Even though we could not exclude the possibility that this correlation at least partially resulted from the expression level of the different constructs, this result prompted us to investigate if there are any key amino acids (AA) that regulate the nucleocytoplasmic distribution.

4.4. Mutational analysis of the DDX6 CTD motifs

To find the potential candidate AA, we first compared the sequences of the DDX6 homologs CTD (Fig. 8). We searched for the AA that differ in polarity or charge on the molecular surface in different DDX6 homologs (Fig. 8c, 9, and 10), as we hypothesized such differences can alter the interaction of the DDX6 CTD with other macromolecules, which can in turns affect its interaction with mitotic chromosomes. As a result, we found several candidates, including Q356, F362, Y363, R369, E371, H372, L414, F428, and P459 (Fig. 9 and 10). We then mutated several of these residues to match the residues of their counterparts in DjCBC-1 (Fig. 11a), because YFP-CBC1 CTD was the least nuclear-enriched as compared to other homologs. Consequently, however, none of the mutants showed significant abrogation of the nuclear localization (Fig. 11b). We also tested if the conserved helicase motifs, the proteins binding sites, and the phosphorylation sites in the DDX6 CTD affect nuclear localization. However, none of the mutants showed abrogated nuclear localization (Chap. 2).

5. Discussion

With the semi-automatic image analysis (Fig. 1), we discovered that, as compared to YFP-DDX6 CTD, YFP-PKM-DDX6 CTD suffers a time lag in nuclear entry (Fig. 2). This finding indicated that even though YFP-DDX6 CTD and YFP-PKM-DDX6 CTD

alike can bind the mitotic chromosome in approximately 80% of the nocodazole-synchronized cells (Chap. 2), their dynamics were different. This is consistent with the fact that on the second and the third day after transfection, YFP-PKM-DDX6 CTD possessed the widest C.R./Nuc ratio distribution among all the constructs assessed (Fig. 2 and 6a). This is also consistent with our findings in the time-lapse imaging, in which we observed the divergent behavior for YFP-PKM-DDX6 CTD during M phase progression whereas the behavior of YFP-DDX6 CTD was more uniform (Chap. 2).

On the other hand, we showed that the nuclear localization could be a conserved feature of the DDX6 homologs CTD (Fig. 5). Furthermore, as the image analysis clearly suggests that a small fraction of the cells expressing YFP-DDX6 CTD exhibited more uniform nucleocytoplasmic distribution (Fig. 2), we confirmed this heterogeneity in the human DDX6 CTD as well as in the other homologs (Fig. 5). In addition, we showed that the non-conserved C-terminal tails in some DDX6 homologs are dispensable for the nuclear localization (Fig. 5). Furthermore, we observed a correlation between the nuclear-enrichment and the sequence similarity relative to the human DDX6 (Fig. 6b and c), which pointed to the possibility that some AA that differ in the DDX6 homologs could be important for nuclear localization (Fig. 8-10). However, our further analysis did not support this hypothesis (Fig. 11). Furthermore, the nuclear localization of the DDX6 CTD was not abrogated by mutations in the conserved helicase motifs, known protein binding sites, and the consensus phosphorylation sites (Chap. 2). It is possible that the exact motif responsible for the nuclear entry remains to be found. Or, it is also possible the nuclear entry cannot be attributed to a single motif. Nonetheless, an alternative approach is to directly assess the interactome of the DDX6 CTD by

proteomic profiling, which may provide further clues to how the DDX6 CTD interact with the chromosomes and facilitate the nuclear entry through the hitchhiking mechanism.



6. Figures

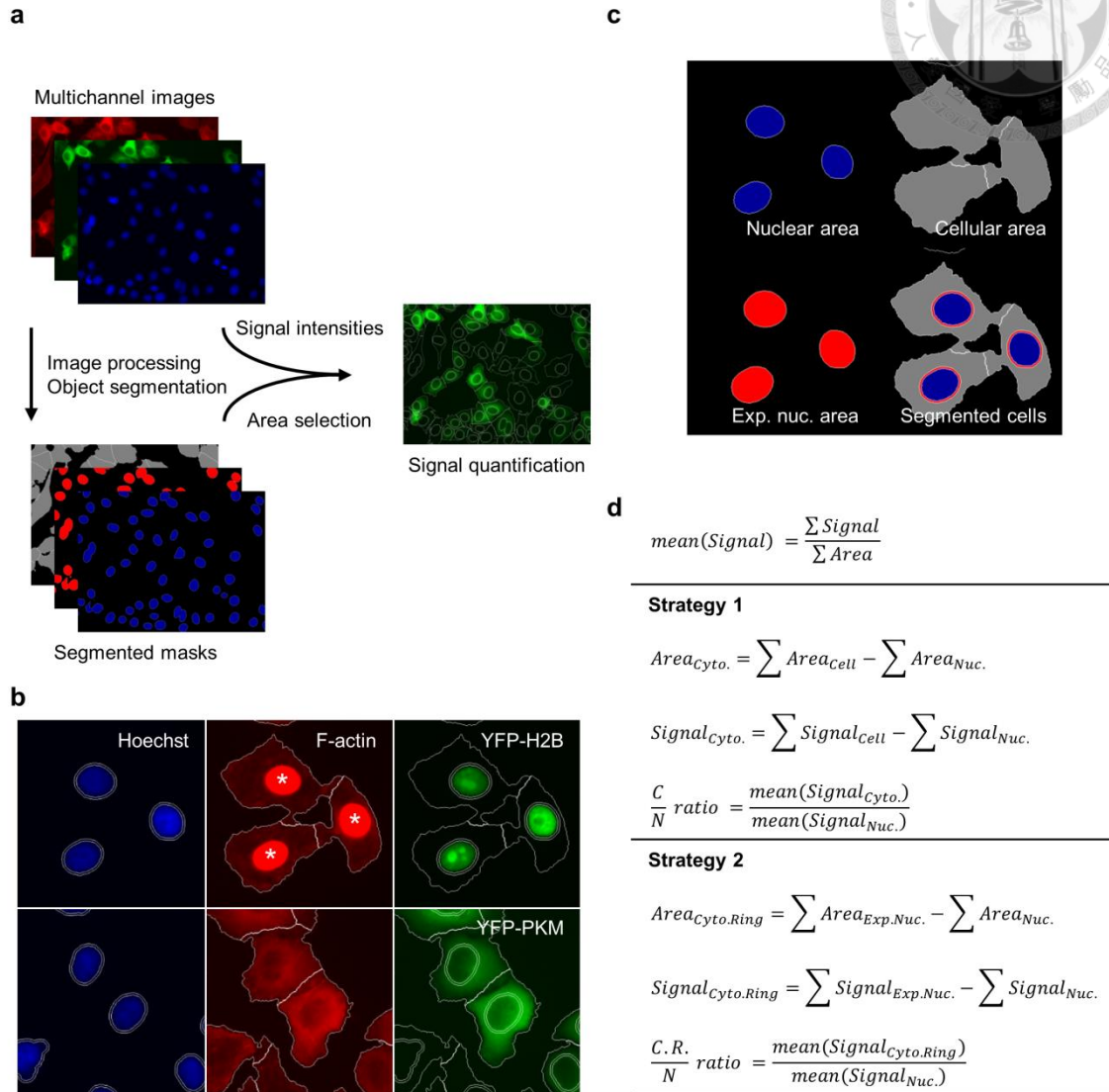


Figure 1. An imaging-based assessment of the protein nucleocytoplasmic distribution

a, The diagram showing the workflow for the semi-automatic image analysis applied in this chapter. **b**, The representative images showing the segmentation results. The cells expressing YFP-H2B or YFP-PKM (green) were fixed and stained with the fluorophore-conjugated phalloidin (red) and Hoechst (blue). The Hoechst signals were used for the segmentation of the nuclei. The segmentation of the nuclei was iterated,

creating a “nuclear area” and an “expanded nuclear area”. The phalloidin signals were used for the segmentation of the cells. The segmentation borders were shown in white. Asterisks: the bleed-through YFP-H2B signals. **c**, The representative images of the segmented cells. The nuclear mask was colored in blue. The expanded nuclear mask was colored in red. And the cellular mask was colored in grey. In the segmented cells, the red ring around the nuclei represents the cytoplasmic ring (cyto. ring, C. R.). **d**, The description of the different strategies of estimating the protein nucleocytoplasmic distribution. Strategy 1: calculation of the cytoplasmic and the nuclear mean-signal ratio. Strategy 2: calculation of the cytoplasmic ring and the nuclear mean-signal ratio.

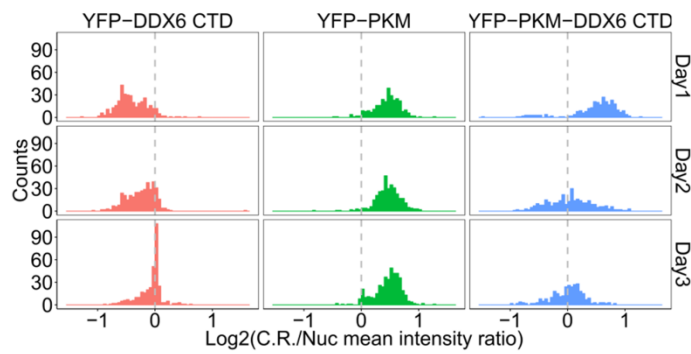


Figure 2. The time-dependent alteration in the nucleocytoplasmic localization of YFP-PKM-DDX6 CTD

The histograms displaying the nucleocytoplasmic distribution of YFP-DDX6 CTD, YFP-PKM, and YFP-PKM-DDX6 CTD over 3 days. The cells transfected with each of the three constructs were imaged on the first, the second, or the third day after transfection. The grey dashed lines correspond to no enrichment. Data from three biological replicates were summarized in this figure.

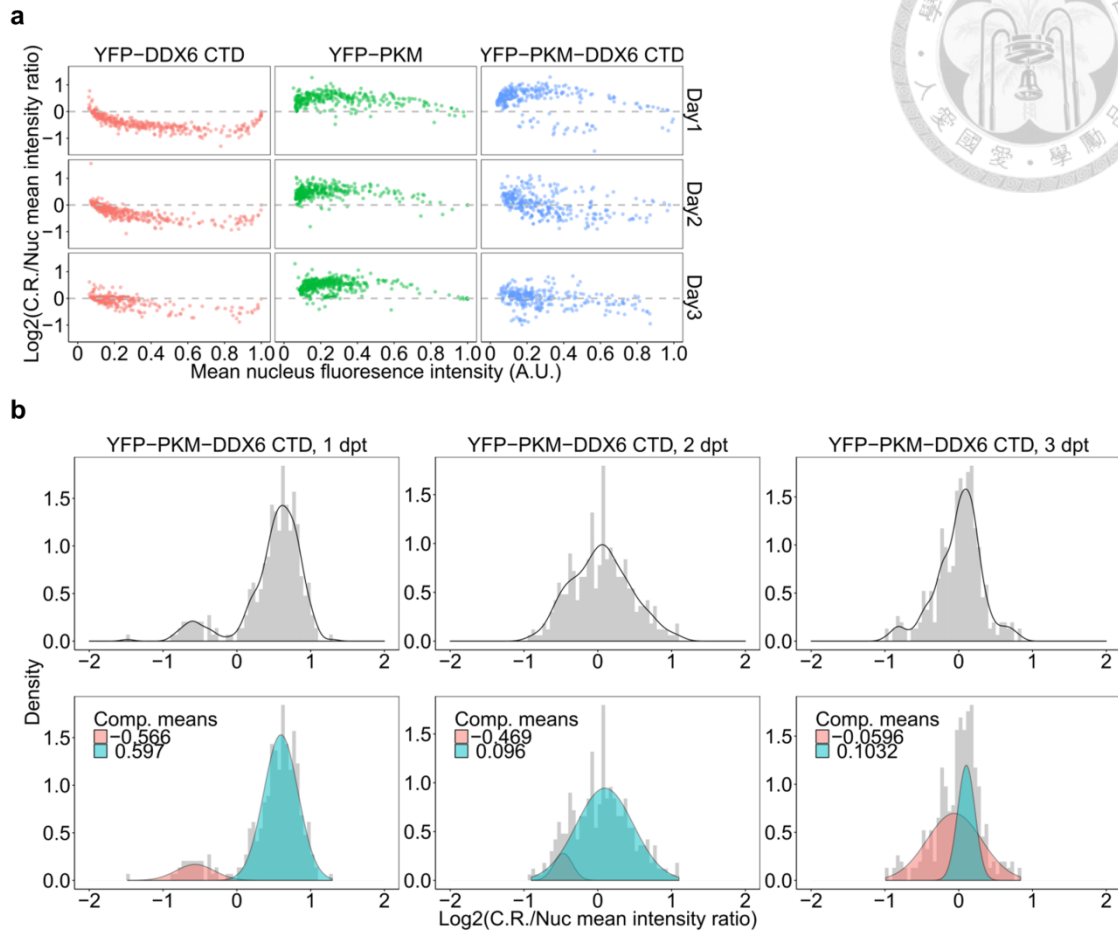


Figure 3. The analysis of the YFP-PKM-DDX6 CTD image dataset

a, The MA plot of the dataset. **b**, The decomposition of the nucleocytoplasmic distribution data with Gaussian mixture model. The upper panels show the density plots of the nucleocytoplasmic distribution across the three days after transfection. The lower panels show the corresponding density plots after Gaussian mixture decomposition. Comp. means: the means of the two component distributions.

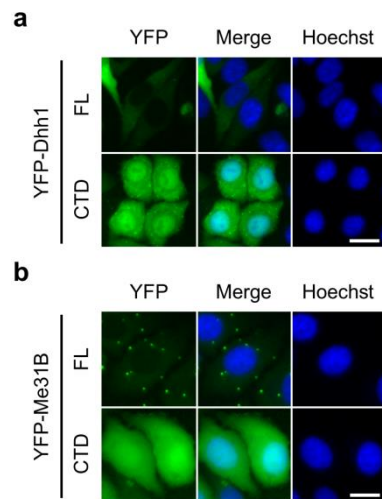


Figure 4. Cloning and expression of the YFP-Dhh1 and YFP-Me31B

a, The representative images of the YFP-Dhh1 FL and CTD. **b**, The representative images of the YFP-Me31B FL and CTD. **a and b**, The YFP signals were shown in green. The nuclei were stained with Hoechst and were shown in blue. Scale bars = 20 μm .

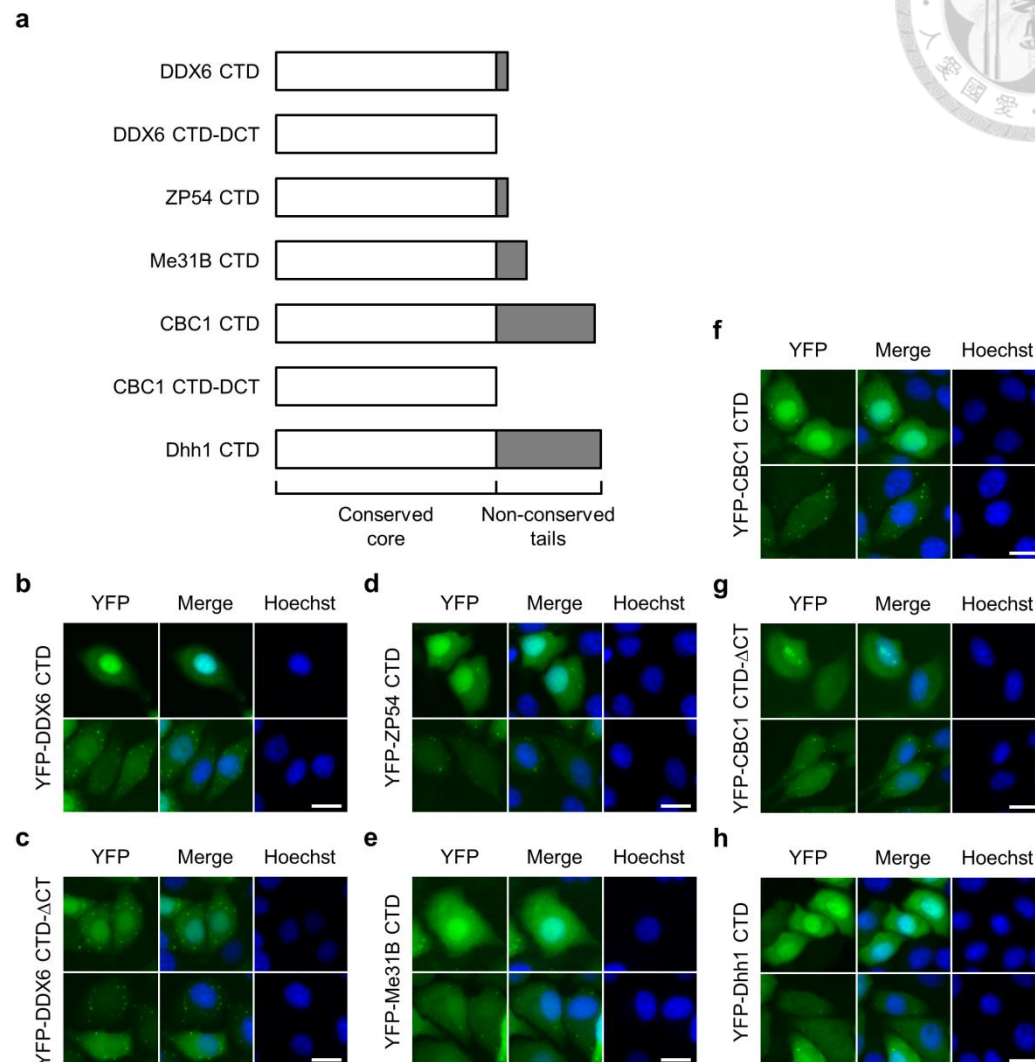


Figure 5. The population-wide heterogeneity in the nucleocytoplasmic localization of the CTD and the mutants of the DDX6 homologs

a, The diagram depicting the DDX6 homologs CTD and C-terminal tail truncation mutants. **b-h**, The representative images of the YFP-tagged CTD and mutants of the DDX6 homologs. The YFP signals were shown in green. The nuclei were stained with Hoechst (blue). Scale bars = 20 μ m. **b**, The CTD of the human DDX6; **c**, the truncation mutant of the human DDX6 CTD (301-473); **d**, the CTD of the zebrafish Zp54; **e**, The

CTD of the *Drosophila* Me31B; **f**, the CTD of the *D. japonica* CBC-1; **g**, the truncation mutant of the CBC-1 CTD (423-504); **h**, the CTD of the *S. cerevisiae* Dhh1p.



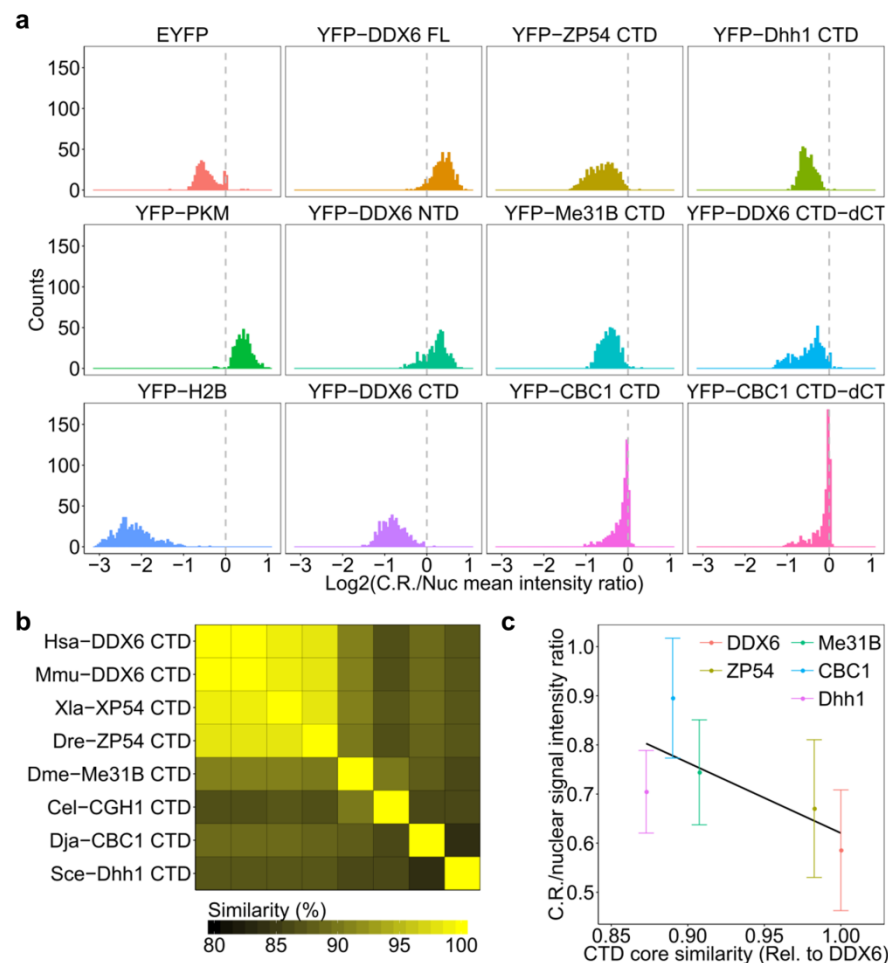


Figure 6. The nucleocytoplasmic distribution of the DDX6 homolog CTD scales with the similarities compared to the human DDX6

a, The histogram showing the nucleocytoplasmic distribution of the transfected constructs. The transfected cells were imaged two days after transfection. Data from three biological replicates were summarized in this plot. The grey dashed lines correspond to no enrichment. **b**, The heatmap showing the sequence similarity matrix of the conserved region of the DDX6 homologs CTD. **c**, The scatter plot showing the correlation between the nucleocytoplasmic distribution and the sequence similarities as compared to the human DDX6. Points denote the means and the error bars represents the standard deviations. The black line represents the regression line.

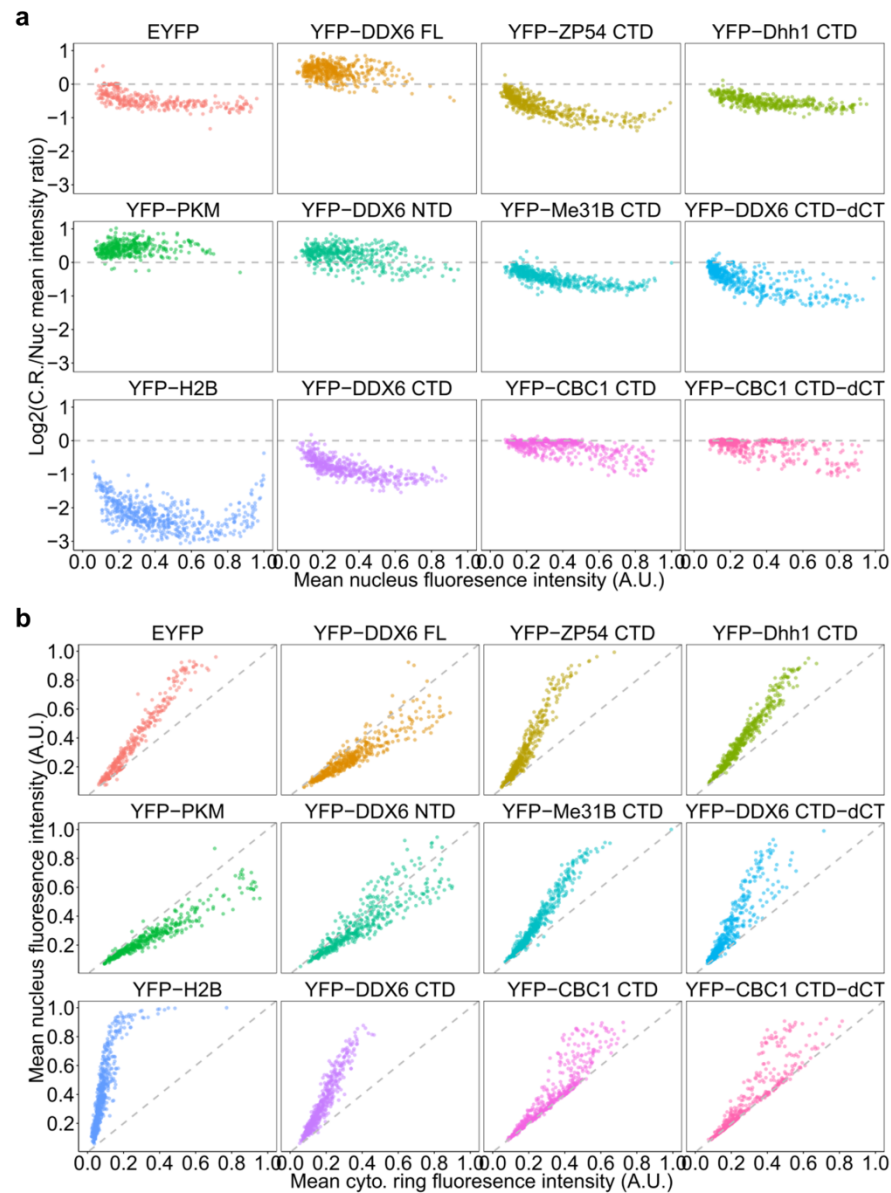


Figure 7. The analysis of the DDX6 homologs CTD image dataset

a and b, The MA plot (**a**) and the scatter plot (**b**) of the dataset. The grey dashed lines correspond to no enrichment.

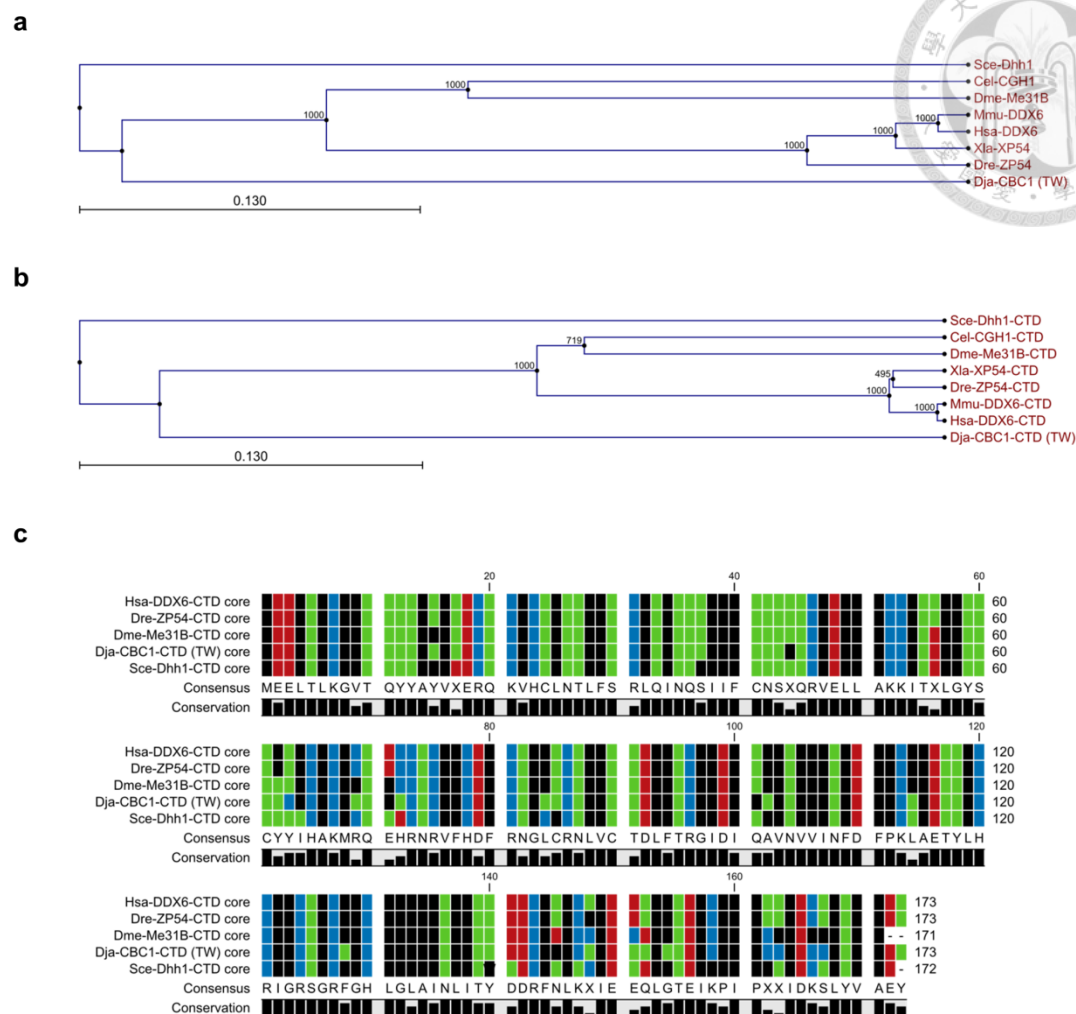


Figure 8. The sequence comparison between the DDX6 homologs

a and b, The phylogenetic tree based on the protein sequences of the full length DDX6 homologs (**a**) and the DDX6 homologs CTD (**b**). **c**, The sequence alignment of the core (conserved) region of the DDX6 homologs CTD. The AA were colored by charge and polarity. Black: neutral, nonpolar; green: neutral, polar; blue: positively charged, polar; red: negatively charged.

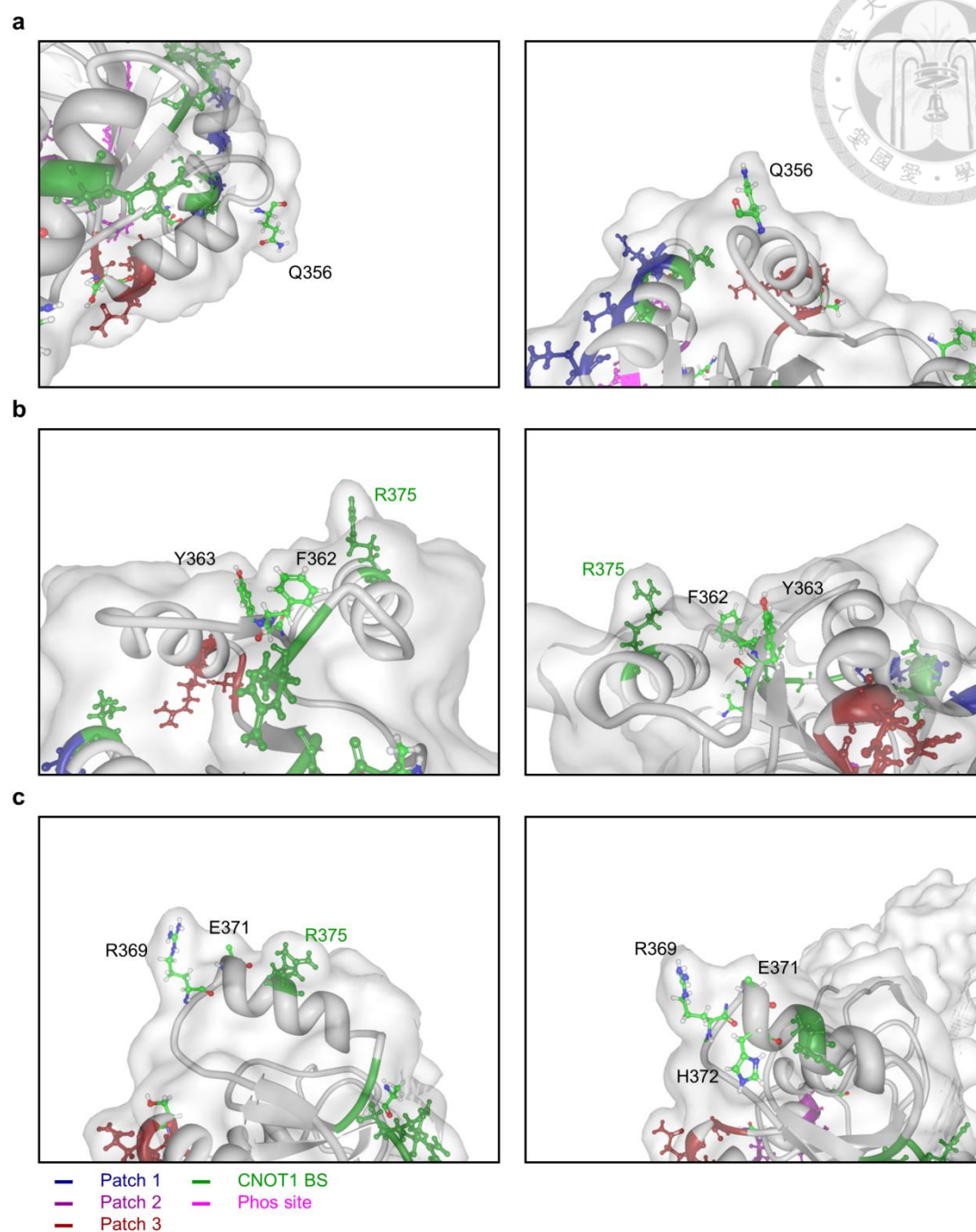


Figure 9. The visualization of several candidate residues

a-c, The visualization of Q356 (**a**), F362/Y363 (**b**), and R369/E371/H372 of the human DDX6 on the molecular surface. Structure: 4CT5.

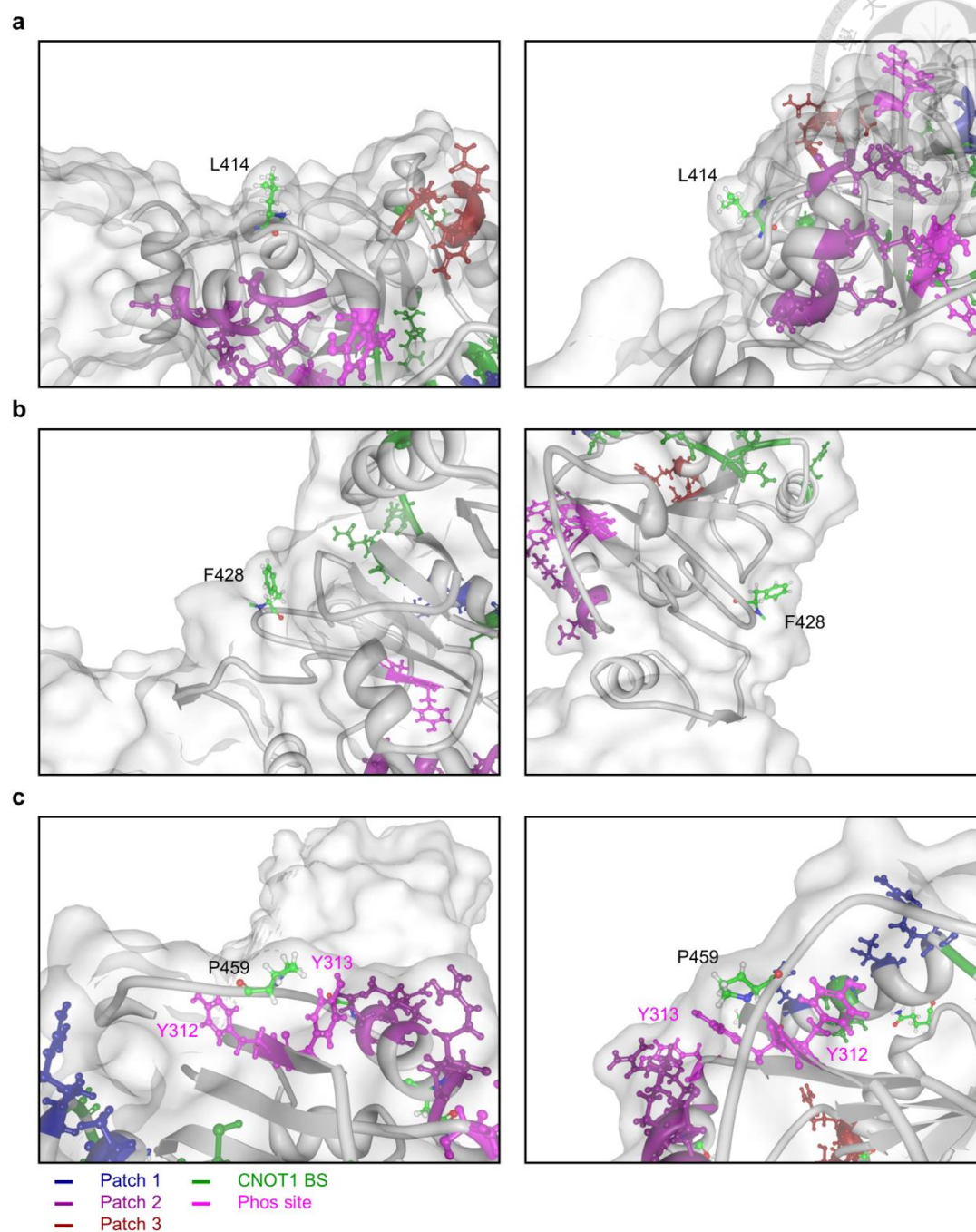


Figure 10. The additional candidate residues

a-c, The visualization of the additional candidate residues, including L414 (a), F428 (b), and P459 (c) of the human DDX6 on its molecular surface.

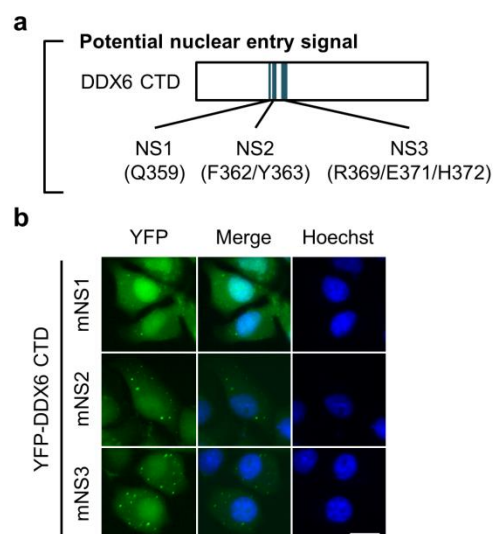
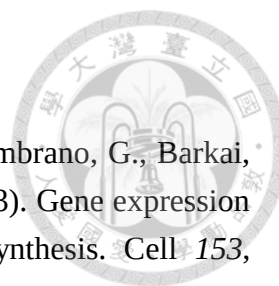


Figure 11. The nuclear localization of YFP-DDX6 CTD is independent of the candidate residues.

a, The diagram depicting the mutations of the candidate residues. **b**, The representative images of the YFP-DDX6 CTD mutants carrying the mutations of the candidate residues. The YFP signals were shown in green. The nuclei were stained with Hoechst and were shown in blue. Scale bar = 20 μ m.



7. **References**

Haimovich, G., Medina, D.A., Causse, S.Z., Garber, M., Millan-Zambrano, G., Barkai, O., Chavez, S., Perez-Ortin, J.E., Darzacq, X., and Choder, M. (2013). Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell* 153, 1000-1011.

Iio, A., Takagi, T., Miki, K., Naoe, T., Nakayama, A., and Akao, Y. (2013). DDX6 post-transcriptionally down-regulates miR-143/145 expression through host gene NCR143/145 in cancer cells. *Biochim. Biophys. Acta* 1829, 1102-1110.

Kawahara, C., Yokota, S., and Fujita, H. (2014). DDX6 localizes to nuage structures and the annulus of mammalian spermatogenic cells. *Histochem. Cell Biol.* 141, 111-121.

Kruk, J.A., Dutta, A., Fu, J., Gilmour, D.S., and Reese, J.C. (2011). The multifunctional Ccr4-Not complex directly promotes transcription elongation. *Genes Dev.* 25, 581-593.

Ladomery, M., Wade, E., and Sommerville, J. (1997). Xp54, the *Xenopus* homologue of human RNA helicase p54, is an integral component of stored mRNP particles in oocytes. *Nucleic Acids Res.* 25, 965-973.

Minshall, N., Thom, G., and Standart, N. (2001). A conserved role of a DEAD box helicase in mRNA masking. *RNA* 7, 1728-1742.

Pau, G., Fuchs, F., Sklyar, O., Boutros, M., and Huber, W. (2010). EBIImage--an R package for image processing with applications to cellular phenotypes. *Bioinformatics* 26, 979-981.

Pettersson, O.J., Aagaard, L., Andrejeva, D., Thomsen, R., Jensen, T.G., and Damgaard, C.K. (2014). DDX6 regulates sequestered nuclear CUG-expanded DMPK-mRNA in dystrophin myotonia type 1. *Nucleic Acids Res.* 42, 7186-7200.

Smillie, D.A., and Sommerville, J. (2002). RNA helicase p54 (DDX6) is a shuttling protein involved in nuclear assembly of stored mRNP particles. *J. Cell Sci.* 115, 395-407.

Spencer, S.L., Cappell, S.D., Tsai, F.C., Overton, K.W., Wang, C.L., and Meyer, T. (2013). The proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit. *Cell* 155, 369-383.

Weston, A., and Sommerville, J. (2006). Xp54 and related (DDX6-like) RNA helicases: roles in messenger RNP assembly, translation regulation and RNA degradation. *Nucleic Acids Res.* 34, 3082-3094.

Yoshida-Kashikawa, M., Shibata, N., Takechi, K., and Agata, K. (2007). DjCBC-1, a conserved DEAD box RNA helicase of the RCK/p54/Me31B family, is a component of RNA-protein complexes in planarian stem cells and neurons. *Dev. Dynam.* 236, 3436-3450.

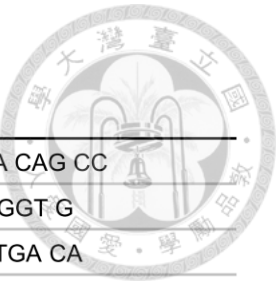




Appendices

A Catalog of Resources



Table 1. List of quantitative primer sets


Gene	Directionality	Sequence (5'-3')
GAPDH	F	CCA TGA GAA GTA TGA CAA CAG CC
	R	GGG TGC TAA GCA GTT GGT G
ACTB	F	CTG GAA CGG TGA AGG TGA CA
	R	AAG GGA CTT CCT GTA ACA ATG CA
18S rRNA	F	GGC CCT GTA ATT GGA ATG AGT C
	R	CCA AGA TCC AAC TAC GAG CTT
TBP	F	TGC ACA GGA GCC AAG AGT GAA
	R	CAC ATC ACA GCT CCC CAC CA
DDX6	F	GCT GGG AAA AGC CAT CTC CTA
	R	GGT CTA GCC GTT CAA GTA AGG G
MALAT1 (set 1)*	F	GAA TGC TTG AAG TAC CCC TGG
	R	CTG GAA AAG CTA GGG AAA AGT GG
MALAT1 (set 2)*	F	ACC ATC GTT ACC TTG AAA CCG
	R	GAT CTA GCA CAG ACC CTT CAC
NEAT1	F	ACA GAG TCA CCA GTT TTC CG
	R	ACT TGA TAA CAC CCA CAC CC
MYC	F	CCT GGT GCT CCA TGA GGA GAC
	R	CAG ACT CTG ACC TTT TGC CAG G
GPRC5A	F	CTC ACA AAG CAA CGA AAC CC
	R	TGT CCC CTG TCT CTT CAA AAC
AGO2 (set 1)*	F	CGG AAG CCC GTG TTT GAC GG
	R	CGC AGG ACA CCC ACT TGA TGG
AGO2 (set 2)*	F	CCT CCC ATG TTT ACA AGT CG
	R	TCT TTG TCC TGC CAC AAT G
GAS5	F	CTT GCC TGG ACC AGC TTA AT
	R	CAA GCC GAC TCT CCA TAC CT

Asterisks: Alternative qPCR primer sets that yield similar results. In this thesis, results obtained from a single set were reported.

Table 2. List of small interfering RNA oligonucleotides

Name	Target gene	Target sequence (sense; 5' – 3')	Working conc.	Reagent
siDDX6-mm ^a	DDX6 (mis-matched)	GCAGA AACGG UAUGA GAUU	25 nM	2000/ RNAiMAX
siDDX6-pm	DDX6	GCAGA AACCC UAUGA GAUU	25 nM	
siControl ^b	Non-targeting to human genes	UAGCG ACUAA ACACA UCAA	25 nM	
siAGO2	AGO2	GCACG GAAGU CCAUC UGAA	25 nM	
siMALAT1-N	MALAT1	GAAAG CGAGU GGUUG GUAA	50 nM	

a, Control for DDX6 knockdown.

b, Control for AGO2 and MALAT1 knockdown.

Table 3. List of antibodies for Western blots

Primary antibodies					
Target	Manufacturer	Cat. No.	Host species	Working conc.	Incubation time (hr) at room temperature
DDX6	Bethyl laboratories	A300-460A	Rabbit	1:5000	0.5 ~ 1
LMNA	Abcam	ab108595	Rabbit	1:10000	0.5
GAPDH	GeneTex	GTX100118	Rabbit	1:10000	0.5
CALR	EMD Millipore	06-661	Rabbit	1:5000	3
AGO2	Abnova	H00027161-M01	Mouse	1:1000	3
PATL1	Abcam	ab122526	Rabbit	1:1000	3
DCP1A	Bethyl laboratories	A303-590A	Rabbit	1:5000	0.5 ~ 1
GFP-tag	Clontech	632375	Mouse	1:5000	1 ~ 3
Myc-tag	EMD Millipore	05-724	Mouse	1:1000	1 ~ 3
POLR2A	Bethyl laboratories	A300-653A	Rabbit	1:2500	3
POLR2A Phospho-Ser2	Bethyl laboratories	A300-654A	Rabbit	1:5000	3
eIF4E	Cell signaling	9742	Rabbit	1:1000	1
CNOT1	Proteintech	14276-1-AP	Rabbit	1:1000	3
ACTB	GeneTex	GTX109639	Rabbit	1:5000	0.5 ~ 1
Secondary antibodies, HRP-conjugated					
Mouse IgG	Santa Cruz	SC-2004	Goat	1:5000	0.5
Rabbit IgG	Santa Cruz	SC-2005	Goat	1:5000	0.5
Secondary antibodies, HRP-conjugated (For IP verification)					
Native rabbit IgG	Rockland	18-8816-33	Mouse	1:5000	0.5
Native mouse IgG	Rockland	18-8817-33	Rat	1:5000	0.5

Table 4. List of antibodies for immunofluorescence

Primary antibodies				
Target	Manufacturer	Cat. No.	Host species	Working conc.
DDX6	Bethyl laboratories	A300-460A	Rabbit	200 ng per 100 μ g lysate in 1 mL reaction
DCP1A	Bethyl laboratories	A303-590A	Rabbit	200 ng per 100 μ g lysate in 1 mL reaction
GFP-tag	Wako	012-20461	Mouse	200 ng per 100 μ g lysate in 1 mL reaction

Control antibodies				
NA	Santa Cruz	SC-2027	Rabbit	200 ng per 100 μ g lysate in 1 mL reaction
NA	Santa Cruz	SC-2025	Mouse	200 ng per 100 μ g lysate in 1 mL reaction

Table 5. List of antibodies for immunoprecipitation

Primary antibodies				
Target	Manufacturer	Cat. No.	Host species	Working conc.
DDX6	Bethyl laboratories	A300-460A	Rabbit	1:1000
DCP1A	Bethyl laboratories	A303-590A	Rabbit	1:1000
Myc-tag	EMD Millipore	05-724	Mouse	1:1000
Phospho-SC35	Abcam	ab11826	Mouse	1:1000

Secondary antibodies, Alexa fluor-conjugated				
Mouse IgG	Thermo Fisher Scientific	A-21202 (Alexa 488)	Donkey	1:1000
Mouse IgG	Thermo Fisher Scientific	A10037 (Alexa 568)	Donkey	1:1000
Rabbit IgG	Thermo Fisher Scientific	A-21206 (Alexa 488)	Donkey	1:1000
Rabbit IgG	Thermo Fisher Scientific	A10042 (Alexa 568)	Donkey	1:1000

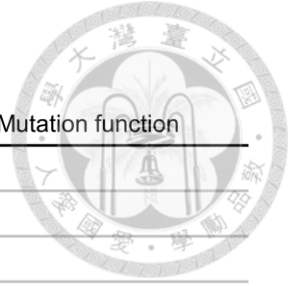


Table 6. List of expression vectors and constructs

Name	Vector	Segment	Point mutation	Mutation function
EYFP	pEYFP-C1			
ECFP	pECFP-C1			
YFP-DDX6 FL	pEYFP-C1	1-483		
YFP-DDX6 sm	pEYFP-C1	1-483	Silent	Evade siRNA
YFP-DDX6 NTD	pEYFP-C1	1-300		
YFP-DDX6 CTD	pEYFP-C1	301-483		
CFP-DDX6 sm	pECFP-C1			
YFP-DDX6 NTD mNES	pEYFP-C1	1-300	L158A/L160A	Disrupt putative NES
YFP-DDX6 FL mNES1	pEYFP-C1	1-483	L158A/L160A	Disrupt putative NES
YFP-DDX6 141-483 WT	pEYFP-C1	141-483		Disrupt putative NES
YFP-DDX6 141-483 mNES1	pEYFP-C1	141-483	L158A/L160A	Disrupt putative NES
YFP-DDX6 CTD-D1	pEYFP-C1	301-420		Truncate DDX6 CTD
YFP-DDX6 CTD-D2	pEYFP-C1	361-483		Truncate DDX6 CTD
YFP-DDX6 CTD-DCT	pEYFP-C1	301-473		Truncate DDX6 CTD
YFP-DDX6 CTD mut1	pEYFP-C1	301-483	Q320A/H323A/T327A/R331A	Disrupt FDF-binding pocket
YFP-DDX6 CTD mut2	pEYFP-C1	301-483	R443A/F444A/K447A/E450A/E451A	Disrupt FDF-binding pocket
YFP-DDX6 CTD mut3	pEYFP-C1	301-483	S343D/Q345D/R346D	Disrupt FDF-binding pocket
YFP-DDX6 CTD R386E	pEYFP-C1	301-483	R386E	Disrupt FDF-binding pocket
YFP-DDX6 CTD mPS1	pEYFP-C1	301-483	Y312A/Y313A	Disrupt phosphorylation site
YFP-DDX6 CTD mPS2	pEYFP-C1	301-483	Y440A	Disrupt phosphorylation site
YFP-DDX6 CTD mPS3	pEYFP-C1	301-483	Y473A/S475A	Disrupt phosphorylation site
YFP-DDX6 CTD F340A/N342A	pEYFP-C1	1-300	F340A/N342A	Disrupt helicase motif
YFP-DDX6 CTD R396A/G397A	pEYFP-C1	1-300	R396A/G397A	Disrupt helicase motif
YFP-DDX6 CTD H420A/R421A	pEYFP-C1	1-300	H420A/R421A	Disrupt helicase motif
YFP-DDX6 CTD G427A/R428A	pEYFP-C1	1-300	G427A/R428A	Disrupt helicase motif

Continues on the next page.

Table 6. List of expression vectors and constructs (continued)


Name	Vector	Segment	Point mutation	Mutation function
Myc-DDX6 FL	pKR43	1-483		
Myc-DDX6 NTD	pKR43	1-300		
Myc-DDX6 CTD	pKR43	301-483		
YFP-PKM	pEYFP-C1	1-531		
YFP-H2B	pEYFP-C1	1-126		
YFP-DBR1 CTD	pEYFP-C1	241-544		
YFP-PKM-DBR1 CTD	pEYFP-C1	1-531/ 241-544		
YFP-PKM-DDX6 CTD	pEYFP-C1	1-531/ 301-483		
YFP-DDX6 NTD-DBR1 CTD	pEYFP-C1	1-300/ 241-544		
YFP-DDX6 Rec	pEYFP-C1	1-300/ 301-483	VL insertion between L300 and M301	Sall site
YFP-AGO2	pEYFP-C1	1-825		
YFP-4E-T	pEYFP-C1			
YFP-PATL1	pEYFP-C1	1-770		
YFP-LSM14A	pEYFP-C1	2-463		
YFP-EDC3	pEYFP-C1	1-508		
GFP-DCP1A	pEGFP-C1			
GFP-DCP2	pEGFP-C1			
GFP-CNOT1	pEGFP-C1			
YFP-ZP54 CTD	pEYFP-C1	302-484		
YFP-Me31B	pEYFP-C1	1-459		
YFP-Me31B CTD	pEYFP-C1	262-459		
YFP-CBC1	pEYFP-C1			
YFP-CBC1 CTD	pEYFP-C1	253-504		
YFP-CBC1 CTD-DCT	pEYFP-C1	253-425		
YFP-Dhh1	pEYFP-C1	1-506		
YFP-Dhh1 CTD	pEYFP-C1	250-506		

