UNIVERSITA' VITA-SALUTE
SAN RAFFAELE

CORSO DI DOTTORATO DI RICERCA
INTERNAZIONALE IN MEDICINA
MOLECOLARE

CURRICULUM IN NEUROSCIENCE AND EXPERIMENTAL NEUROLOGY

Modulation of CDKL5 expression and relevance in CDKL5 Deficiency Disorder: new insights from translation efficiency driven by the 5'UTR

Director of Studies: Daniele Zacchetti

Second Supervisor: Maria Passafaro

Tesi di DOTTORATO di RICERCA di Valeria Ruggiero
matr. 014477
Ciclo di dottorato XXXIV
SSD BIO/11

Anno Accademico 2020/2021

CONSULTAZIONE TESI DI DOTTORATO DIRICERCA

Il/la sottoscritto/*I*	Valeria Ruggiero
Matricola / *registration number*	014477
nat_ a/ *born at*	Tortona (AL)
il/*on*	22/01/1993

autore della tesi di Dottorato di ricerca dal titolo / *author of the PhD Thesis titled*

Modulation of CDKL5 expression and relevance in CDKL5 Deficiency Disorder: new insights from translation efficiency driven by the 5'UTR

Y AUTORIZZA la Consultazione della tesi / AUTHORIZES *the public release of the thesis*

X NON AUTORIZZA la Consultazione della tesi per 12 mesi /DOES NOT AUTHORIZE *the publicrelease of the thesis for 12 months*

a partire dalla data di conseguimento del titolo e precisamente */ from the PhD thesis date, specifically*

Dal / *from* 03/04/2022 Al / *to* 03/04/2023 Poiché /*because*:
□ l'intera ricerca o parti di essa sono potenzialmente soggette a brevettabilità/ *The wholeproject or part of it might be subject to patentability;*

□ ci sono parti di tesi che sono già state sottoposte a un editore o sono in attesa di pubblicazione/ *Parts of the thesis have been or are being submitted to a publisher or are in press;*

□ la tesi è finanziata da enti esterni che vantano dei diritti su di esse e sulla loro pubblicazione/ *the thesis project is financed by external bodies that have rights over it andon its publication.*

Data /Date .....20/03/2022.....

Firma /Signature .....

DECLARATION

This thesis has been composed by myself and has not been used in any previous application for a degree. Throughout the text I use both 'I' and 'We' interchangeably.

All the results presented here were obtained by myself, except for:

1) **CAGEr Analysis.**
   CAGEr Bioinformatic Analysis (Results, par.5.7, Fig.5.7.1 and Fig.5.7.2, par. 9.3), was performed in collaboration with Dr. Stefano De Pretis, Center for Omics Sciences at IRCCS Ospedale San Raffaele, Milan, Italy.

All sources of information are acknowledged by means of reference.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my Director of Studies Daniele Zacchetti and to Professor Nicoletta Landsberger. My thankfulness also goes to my colleagues Valerio Di Carlo, Alessandra Pisciottani, Maria Regoni, Michela Palmieri and Claudio Fagioli for supporting me with their knowledge and, perhaps more importantly, with their kindness. I also wish to thank Luca Biasetti for the great help, the support and the constructive advice provided throughout the last part of my PhD course.

ABSTRACT

CDKL5 is a protein kinase with important functions, acting as transcriptional regulator in the nucleus, modulating the cell cycle and apoptosis, and sensing DNA damages. This kinase, enriched in neurons, is involved in the correct development of the neuronal networks, and has a fundamental role in shaping synapses. Therefore, it does not come as a surprise that a decrease in its expression leads to a severe neurodevelopmental condition known as CDKL5 Deficiency Disorder (CDD), a rare X-linked epileptic encephalopathy. Most of the causative mutations of the disorder were reported in the CDKL5 coding region, but the identification of a novel SNP within the CDKL5 transcript leader, described by Evans et al. in 2005, led us to consider that its 5'UTR could also play a role in maintaining the physiological protein level by modulating the translational efficiency of the transcript. Other proteins with analogous functions in neurons are regulated by a translational control through their 5'UTRs. Indeed, we found that the CDKL5 5'UTR respects the typical features of a functional, highly-structured 5'UTRs and shows an impressive conservation throughout evolution. In addition, we observed that the silencing of eIF4B, a translational eukaryotic initiation factor involved in the unwinding of structured 5'UTR, correlated with a strong decrease of CDKL5. We analyzed the CDKL5 5'UTR by bioinformatic tools and verified the functionality of the various 5'UTR variants through a Dual Luciferases Reporter Assay, supporting a role of CDKL5 5'UTR in the translational modulation of the protein expression through cap-dependent and IRES-mediated mechanisms. Moreover, we obtained the first experimental hint pointing to a possible pathogenic role of the SNP found by Evans. Finally, we quantified the TSS usage of different CDKL5 transcript variants based on CAGE libraries, to better understand the meaning of the numerous alternative first exons of CDKL5 and their usage in the human tissues. Our work is the first comprehensive study about the 5'UTR of CDKL5, and not only demonstrates the importance of 5'UTR in the modulation of CDKL5, but also potentially open new options for therapeutic strategies to treat CDD.

Table of Contents

1. ACRONYMS AND ABBREVIATIONS

3'UTR : 3' Untranslated Region

4E-BP : eIF4B Binding Protein

5'UTR : 5' Untranslated Region

AD : Alzheimer's Disease

ADS : autism spectrum disorder

AMPA : α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid

Apaf-1 : apoptotic protease activating factor 1

ARHGEF : Rho guanine nucleotide exchange factor 6

ASD : Autism spectrum disorder

ASOs : Antisense Oligonucleotides

ATF4 : Activating transcription factor 4

BACE1 : Beta-secretase 1

BAG1 : BCL2 associated athano-gene 1

Bcl-2 : B-cell lymphoma 2

BDNF : Brain-derived neurotrophic factor

Bic : Bicuculline

c-Myc : basic helix-loop-helix protein 37

C/EBPβ : CCAAT/enhancer-binding protein beta

CAGE : Cap Analysis of Gene Expression

CDD : CDKL5 Deficiency Disorder

CDKL5 : cyclin dependent kinase like 5

CDS : coding sequence

CEP131 : Centrosomal Protein 131

CF : Cystic Fibrosis

CFTR : Cystic fibrosis transmembrane conductance regulator

CK : casein kinase

CLIP170 : Cytoplasmic linker protein 170

CXCL14 : C-X-C Motif Chemokine Ligand 14

DLG5 : Disks large homolog 5

Dnmt1 : DNA methyltransferase 1

EB2 : microtubule End-Binding 2

eIF : eukaryotic Initiation Factor

ERK : extracellular signal-regulated kinase

EST : Expressed sequence tag

Ex : Exon

FGFs : Fibroblast Growth Factors

Fluc : Firefly luciferase

FMRP : Fragile X mental retardation protein

FTL : L-ferritin

FXS : Fragile X syndrome

G4 : GQuadruplex

GABA : Gamma-aminobutyric acid

GAPDH : glyceraldehyde-3-phosphate dehydrogenase

HDCA4 : Histone deacetylase 4

HHCS : hereditary hyperferritinaemia-cataract syndrome

hnRNPQ : heterogeneous nuclear ribonucleoprotein Q

HSV1 : herpes simplex virus 1

IONO : Iomonycin

IRES : Internal Ribosome Entry Site

IRF6 : interferon regulatory factor 6

ITAF : IRES Trans-Acting Factor

LRRK2 : leucine-rich repeat kinase 2

m6A cap : N6-methyladenosine

m7G cap : 7-methylguanosine cap

MAP1S : microtubule associated protein 1S

MeCP2 : methyl-CpG binding protein 2

Met-tRNAiMet : initiator methionyl-transfer RNA

mGluR : metabotropic glutamate receptor

mRNA : messanger RNA

mTOR : mammalian target of rapamycin

MycN : basic helix-loop-helix protein 37

NMDA : N-methyl-D-aspartate

ORF : Opening Reading Frame

PD : Parkinson's Disease

PIC : pre-initiation complex

PPS : popliteal pterygium syndrome

PSD95 : Postsynaptic density protein-95

PTB : polypyrimidine tract-binding protein 1

PTEN : phosphatase and tensin homolog

RefSeq : Reference Sequence

RN : Renilla luciferase

RPL13 : ribosomal protein L13

rRNA : ribosomale RNA

RTT : Rett Syndrome

S6K1/2 : p70 S6 kinases

SCS : Saethre-Chotzen syndrome

Ser : serine

siRNA : Small Interference RNA

SMA : Spinal muscular atrophy

SMAD3 : Mothers against decapentaplegic homolog 3

SMN : Survival Motor Neuron

SNP : Single Nucleotide Polymorphism

TC : Ternary Complex

TGFβ : Transforming growth factor beta

TISU : translation initiator of short 5'UTR

Tk : thimydine kinase

TP53 : tumor protein p53

TPO : thrombopoietin

TRMP : TP53-regulated modulator of p27

TSS : Transcription Start Site

TV : Transcript Variant

TWIST1 : Twist Family BHLH Transcription Factor 1

uAUG : upstream AUG (start codon)

uORF : upstream Opening Reading Frame

VEGFs : Vascular Endothelial Growth Factor

VWS : Van der Woude syndrome

XIAP : X-linked inhibitor of apoptosis

YLPM1 : YLP motif-containing protein 1

IEG : immediate early gene

2. LIST OF FIGURES AND TABLES

INTRODUCTION

RESULTS AND DISCUSSION

MATERIALS AND METHODS

APPENDIX

3. INTRODUCTION

3.1 Translational Control

3.1.1 Principles of Translational Control

Translation is a key process in protein expression. On average, up to 20% of the energy of the cell is invested in protein synthesis – compared to the 15% spent for DNA transcription and replication combined (Buttgereit and Brand 1995). This great investment is also demonstrated by the fact that the two, most transcribed, categories of RNA are the ribosomal RNAs (rRNAs) and the messenger RNAs (mRNAs) encoding for ribosomal proteins. Therefore, the translation is considered the most- demanding cellular process (Rolfe and Brown, 1997). The importance of translation isto not due solely to the fact that it is a last and rather passive step in protein production,but also to its potential role in the fast and dynamic modulation of protein abundance (Robichaud et al., 2019; Hershey et al., 2019; Tahmasebi et al., 2019).

Protein expression starts from the transcription of the gene, a highly regulated process and, then, post-transcriptional mechanisms, implying the maturation of the transcript through several modifications (capping, polyadenylation, splicing, methylation, etc), determine how and how much transcript is available for translation. However, despite the importance of these mechanisms, the mature mRNA level is still not the unique determinant of protein translation. Some mRNAs are poorly translated even if their levels would predict a relevant abundance of the encoded protein, and some mRNAs are even translated at different rates according to the moment in the life of the cell (Biswas et al., 2018; Ingolia et al., 2018). So, specific mechanisms of protein synthesis, after mRNA transcription and maturation, can cause a differential translation efficiency and are fully recognized as key elements of the gene expression process.

The concept of translational control arose naturally from the mRNA itself. As early as the year 1961, at the birth of the concept of mRNA, Jacob and Monod stated that “the synthesis of individual proteins may be provoked or suppressed within a cell, under the influence of specific external agents, and (…) the relative rates at which different proteins are synthesized may be profoundly altered, depending on external conditions”. This regulation was immediately recognized as “absolutely essential to the survival of the cell”, due to the dynamic nature of the cell as system. Even though not directly linked to translational control, this intuition represented the seed from which the idea

that gene expression can be modulated also by translational efficiency started to take hold (Jacob and Monod, 1961). Not long after, in 1963, the term "translational control" was used for the first time to describe the differential protein expression from the viral genome of the RNA-phage MS2 in a cell-free translation system obtained from E.Coli (Ohtaka and Spiegelman, 1963). After this first evidence on phages, most of the studies about translational control were instead conducted in eukaryotic systems, in which more complex mechanisms of gene expression are expected. Three paradigmaticstudies in eukaryotic organisms contributed to define the features of translational control (Tahmasebi et al., 2019):

- First, the evidence that in eggs of sea urchins – as well as of other invertebrates - protein translation has a very low rate until fertilization, which triggers an increase in translational efficiency, measured by the speed of the incorporation of amino acids, without concomitant transcription of new mRNAs (Hultin, 1961; Nemer, 1962, Gross et al., 1964).
- Second, the indication that mammalian reticulocytes – immature red blood cells without nucleus –regulate their gene expression through translation, in the absence of transcription. In particular, the translation of globins is dynamically controlled by the availability of ferrous ions (Bruns and London 1965), and thus, in its absence, translation of globins is inhibited at the stage of translation initiation (Pelham and Jackson, 1976, Hardesty et al., 1963, Waxman and Rabinowitz, 1966).
- Third, eukaryotic cells regulate the expression of some genes through translational control in response to various stimuli or conditions, such as hormones (Eboué-Bonis et al., 1963, Garren et al., 1964, Tomkins at al., 1965), progression in cell cycle (Steward et al., 1968; Fan and Penman 1970), cell differentiation (Heywood 1970), environmental stress from heat shock or toxic substances (McCormick and Penman 1969; Thomas and Mathews 1984).

All these findings strengthened the idea that translational control is an additional and relevant mechanism when the response to a stimulus must be rapid and responsive – especially in various tissues and cell types of higher organisms, or cells, like neurons, in which a complex morphology makes transcription not the best strategy for modulating gene expression, while the translational control can ensure the proper localization and rapidity of the response (Hershey et al., 2019; Tahmasebi et al., 2019). Even though a translational control can occur at the level of elongation and termination (Dever et al., 2018), the process is mainly regulated at the level of initiation, coherently

with the statement that it is more efficient to manage a pathway at its beginning than to interrupt it midstream changing its workflow (Tahmasebi et al., 2019). Indeed, translation initiation is the rate-limiting phase of protein synthesis, and complex mechanisms can change the translation efficiency of transcripts. For this reason, knowing the molecular events occurring in translation initiation is the starting point to understand the mechanisms of the translational control (Hershey et al., 2019).

3.1.2 Eukaryotic Translation Initiation: an introduction to mechanisms and factors

Translation initiation is a process with high complexity and the main step of regulation of the efficiency of protein synthesis. It ends with the interaction of the complete 80S ribosome with the mRNA and the initiator methionyl-transfer RNA (Met-$tRNA_i^{Met}$), starting translation (Merrick and Pavitt, 2018). The process is mediated by various specific proteins - some of which are called eukaryotic initiation factors (eIFs) - whose actions are often modulated by their phosphorylation state. Most of eukaryotic transcripts undergo a "canonical", cap-dependent translation, involving recognition of 5'capped-mRNA mediated by eIF4E (Hinnebusch, 2014; Borden et al., 2020), although alternative mechanisms can take place and will be described later.

In cap-dependent translation initiation, eIF4E binds the 7-methylguanosine cap (m^7G cap) structure on the 5' end of the leader sequence (also known as 5' untranslated region, 5'UTR) of the transcript and the other initiation factors of the eIF4F complex. eIF4F is a heterotrimeric complex formed by eIF4E, eIF4G, a big protein acting as a scaffold of the complex, and eIF4A, an RNA helicase whose activity is regulated by the binding of the other components of the eIF4F complex and accessory factors such as eIF4B (Merrick et al., 2015). A role of eIF4A is to unwind the mRNA secondary structures in the 5'UTR, the region spanned between the m^7G and the start codon in which, after the formation of the 43S pre-initiation complex (PIC) that contains the small ribosomal subunit 40S, the movement along the transcript in 3' direction to reach the start codon, known as scanning, occurs. After the recognition of the start codon happens, the interaction trough base-pairing between anticodon of the Met-$tRNA_i^{Met}$ and the start codon AUG – in a favorable Kozak consensus sequence – triggers the reorganization of the initiation complex, with the recruitment of eIF5B and, consequently, of the large ribosomal subunit 60S. Releasing of some of the initiation factors brings to the formation of the complete 80S ribosome, ready to start protein synthesis (**Fig. 3.1.2.1**).

The canonical mechanism of translation initiation includes several stages in which translational control can occur. Here, we report some of the best characterized ones to provide a brief description of the main regulatory factors in cap-dependent initiation.

eIF2 is the main carrier of Met-tRNA$_i^{Met}$ and represents a crucial regulatory switch in the modulation of the global cap-dependent initiation process. The affinity of Met-tRNA$_i^{Met}$ to the carrier greatly varies in dependance of the fact that eIF2 binds to GDP or GTP, showing an increase up to 50-fold in affinity when it is present GTP. The eIF2-GTP complex is unstable in comparison to eIF2-GDP, that even needs the aid of eIF2B (a guanine nucleotide exchange factor, GEF) to facilitate the GDP dissociation and thus allowing the GTP and Met-tRNA$_i^{Met}$ binding to form the ternary complex (TC, formed by eIF2, Met-tRNA$_i^{Met}$, and GTP). It has been demonstrated that in periods of cellular stress, eIF2 is controlled by kinases, that phosphorylate its α-subunit on the serine residue 51 (Ser51, in humans). The Ser51 phosphorylation transforms eIF2 in a competitive inhibitor of the action of eIF2B, stopping translation initiation (Rowlands et al., 1988; Pavitt et al., 1998; Marintchev et al., 2020). This represents an off switch, but not all the transcripts are sensitive to this block, and thus, selective mRNAs can escape the repression continuing to be actively translated (Hinnebush et al. 2016; Proud, 2018; Wek, 2018).

Three additional factors, eIF1, eIF1A and eIF3, aid recruiting the TC to the 40S ribosomal subunit (Majumdar et al. 2003; Olsen et al. 2003; Kolupaeva et al. 2005; Cheung et al. 2007).

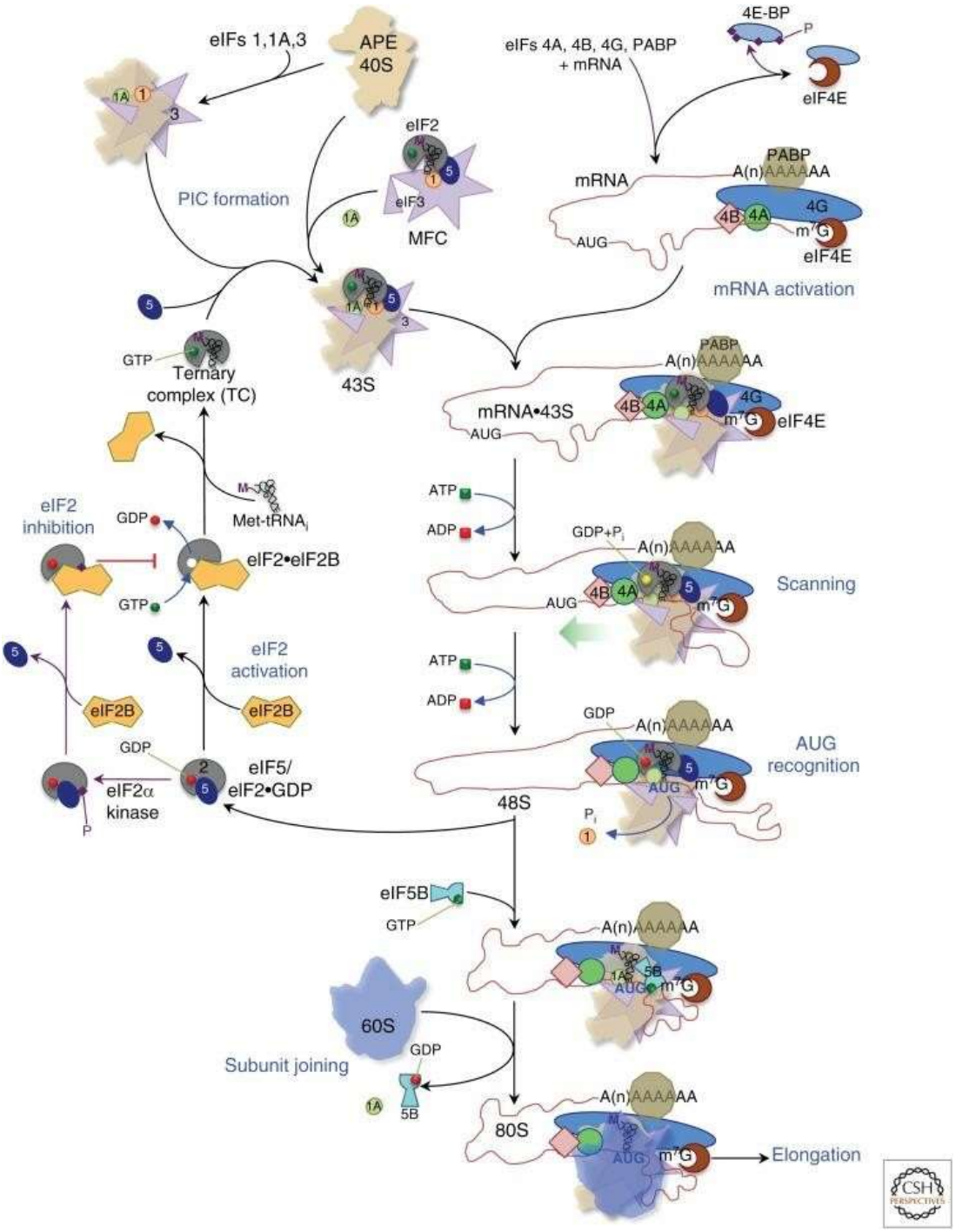


Fig. 3.1.2.1 Schematic representation of the canonical translation initiation pathway. *A series of consecutive steps (here reported in blue text and linked with black arrows) contribute to the canonical translation initiation process. Eukaryotic initiation factors (eIFs) and complexes act in a coordinated manner to the constitution of the functioning 80S ribosome (bottom right). The reactions of nucleotide hydrolysis releasing inorganic phosphate have a crucial role in moving forward the global process, constituting also points of regulation. For example, the regulatory reactions involving the inhibition eIF2 and eIF4E are reported in this diagram, since they represent crucial moments of control of the translation initiation. The*

timing of the release of eIFs (eIF4F, eIF4B or eIF3) from the ribosome at the end of the initiation process is not yet clear, and it is not shown in this figure. (from Merrick and Pavitt, 2018).

eIF1 and eIF1A play a role also in the scanning of the 5'UTR and the stringency of AUG recognition, as demonstrated by mutations that affects these functions (Cheung et al., 2007; Fekete e al., 2007; Hinnebusch, 2014). Additionally, eIF1A acts in recruiting the 60S subunit.

The recruitment of the mRNA is mediated by eIF4E, that is the main m7G cap recognition factor (Gross et al. 2003). Its access to eIF4G, is limited by the phosphorylation of eIF4E-binding proteins (4E-BPs) (Gross et al. 2003; Gruner et al. 2016; Proud 2018).

After the formation of the 43S PIC, other steps of modulation of translational efficiency are represented by the mRNA recruitment and the scanning of the 5'UTR, involving mainly the DEAD-Box helicase eIF4A, eIF4B, eIF4E and eIF4G. The action of eIF4A in the ATP-dependent unwinding highly structured 5'UTR allows the translational machinery to reach the main AUG in spite of the secondary structures that might impede the correct loading of the mRNA into the mRNA channel on the 43S PIC. This function of eIF4A is supported and, possibly, regulated by eIF4B (Abramson et al. 1987; Rozen et al. 1990; Harms et al., 2014).

Canonical translation initiation is not the unique way in which protein synthesis can start. There are alternative mechanisms of initiation, particularly relevant under various conditions or specific to a subset of transcripts, that involve different eIFs and control mechanisms. Examples of these non-canonical initiation mechanisms are:

- the use of the translation initiator of short 5'UTR (TISU), that allows a scanning-free mechanism of translational initiation (Haimov et al., 2017; Kwan and Thompson, 2019.
- ribosomal shunting, that consists in bypassing regions of the 5'UTR to reach the start codon (Koh at al., 2013; Sherril et al., 2011).
- the presence of N^6-methyladenosine (m^6A) in 5'UTR, that enhance the cap-independent translation (Peer et al., 2019).

However, the most studied non-canonical mechanism of initiation is the Internal Ribosome Entry Site (IRES), first identified and described in members of the Picornaviridae family (Jang et al. 1988; Pelletier and Sonenberg 1988). IRES are RNA

elements that recruit the ribosome internally to the transcript, in a cap-independent manner. They can show great variety in length and in the complexity of their folding containing stable secondary structures. It has been estimated that 10% of mammalian transcripts contain IRESs (Martinez-Salas et al. 2012; Walters and Thompson 2016; Weingarten-Gabbay et al. 2016) and generally they control the expression of proteins involved in early development, mitosis, apoptosis, tumorigenesis, viral infection and other stressor responses (Johannes and Sarnow 1998; Johannes et al. 1999; Coldwell et al. 2001; Holcik 2004; Bushell et al. 2006; Dresios et al. 2006; Xue et al. 2015; Vaklavas et al. 2016; Walters and Thompson 2016) (**Tab. 3.1.2.2**).

Cellular pathway	Name	References
Amino acid starvation	Cationic amino acid transporter 1 (CAT-1)	Fernandez et al. 2001
	Sodium-coupled amino acid transporter (SNAT2)	Gaccioli et al. 2006
Nutrient signaling	Human insulin receptor (HIR)	Spriggs et al. 2009
hypoxia	Hypoxia inducible factor 1α (HIF-1α)	Schepens et al. 2005
	Vascular endothelial growth factor (VEGF)	Morris et al. 2010
	Fibroblast growth factor 2 (FGF2)	Bonnal et al. 2003
Apoptosis survival	Apoptotic protease activating factor 1 (Apaf-1)	Sella et al. 1999
	B-cell lymphoma 2 (Bcl-2)	Marash et al. 2008
	BAG family molecular chaperone regulator 1 (Bag1)	Pickering et al. 2004
	Cellular inhibitor of apoptosis 1 (cIAP1)	Graber et al. 2010
	X-linked inhibitor of apoptosis (XIAP)	Riley et al. 2010
Oncogene	c-myc	Le Quesne et al. 2001
	c-Jun	Blau et al. 2012
	p58 PITSLRE	Cornelis et al. 2000
Mitosis	Cyclin-dependent kinase 1 (CDK1)	Marash et al. 2008
	p120 catenin (p120)	Silvera et al. 2009
	p53	Yang et al. 2006
DNA damage response	Serine hydroxymethyltransferase 1 (SHMT1)	Fox et al. 2009
	Replication protein A2 (RPA2)	Yin et al. 2013
Differentiation	Homeobox transcription factor (Hox)	Xue et al. 2015

Tab. 3.1.2.2 Eukaryotic internal ribosomal entry site (IRES). *The table shows a list of IRESs found in transcripts involved in specific cellular pathways, with the reference of the paper in*

which each IRES was described for the first time. The information in the table was taken from Kwan and Thompson, 2019.

The standard way to identify and validate putative IRESs is the dicistronic reporter assay, which employs a DNA vector encoding two reporter genes (commonly Renilla and firefly luciferases). The IRES sequence is inserted between the first and second cistron and, after the transfection of the vector, the reporter activity is monitored. The first cistron uses the canonical translation initiation mechanism, whereas the second cistron, if appropriately spaced, is expressed only in the presence of an IRES-mediated mechanism of translation. Therefore, the IRES activity is indicated as the expression of the second reporter gene. One important problem of this type of approach is to avoid false positive results when DNA transfection is performed, due to either the presence, within the putative IRES, of sequences that might act as cryptic promotes or alternative splicing events. Direct mRNA transfection of the dicistronic transcript can be performed to skip the transcription step and bypass these problems (Thompson et al., 2012; Kwan et al., 2019) (**Fig. 3.1.2.3**).

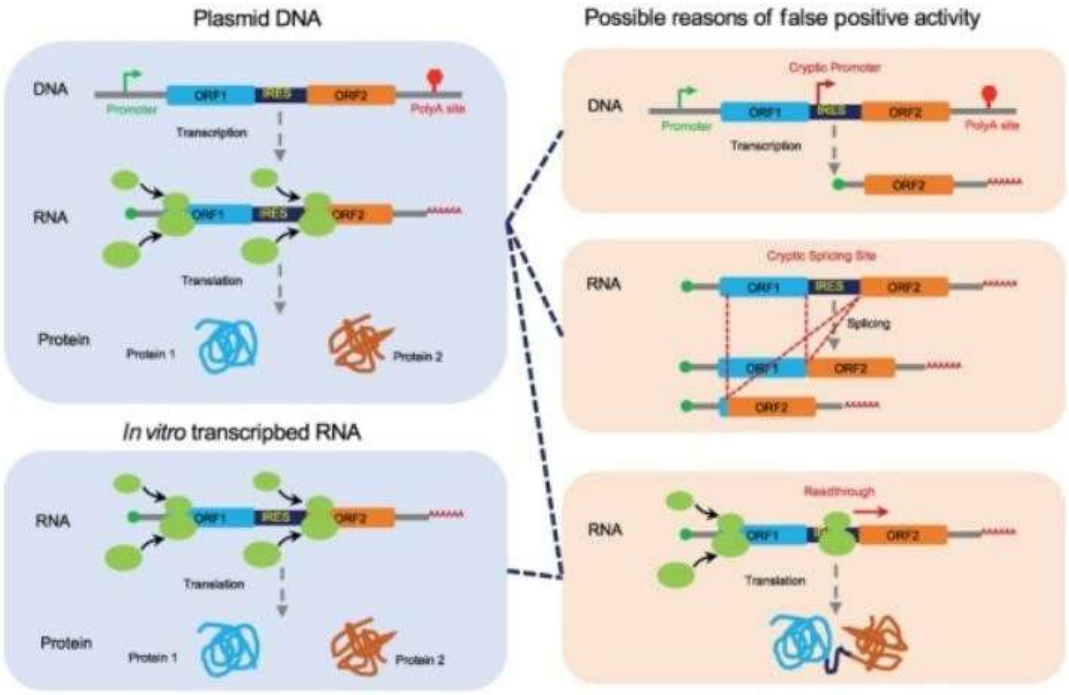


Fig. 3.1.2.3 Dicistronic reporter system in the IRES investigation. *Dicistronic vectors are the most popular reporter system used in the examination of IRES-mediated translation initiation. In the vector, two CDS are inserted downstream the promoter, and the putative IRES is inserted between the two CDSs. During translation, the first CDS is translated through cap-dependent initiation, while the second CDS is translated by the IRES-mediated initiation. Three events can*

cause false positive results: the presence of a sequence with, cryptic promoter activity within the putative IRES; the presence of splicing sites; and translational readthrough or reinitiation. Figure modified from Yang et al., 2019.

One of the most studied eukaryotic IRES is the one contained in the 5'UTR of the apoptotic protease activating factor 1 (Apaf-1) protein. Apaf-1 is involved in apoptosis during the early development and is essential for the correct brain development (Cecconi et al. 1998; Yoshida et al. 1998). It has been shown that the Apaf-1 IRES is most active in neurons, even though the transcript is present in almost all tissues. This result is in part due to the higher affinity of the neuronal IRES Trans-Acting Factor (ITAF) polypyrimidine tract-binding protein 1 (nPTB), compared with the non-neuronal counterpart (Coldwell et al. 2000; Mitchell et al. 2003; Andreev et al. 2009). Therefore, in eukaryotic transcripts, the IRES-mediated translation is greatly modulated by the presence of cell- or tissue-specific ITAFs, which can enhance or suppress translation and, also switch between the cap-dependent and the cap-independent initiation (Pichon et al., 2012; Kim et al., 2004). ITAFs availability does not depend exclusively on the cell type, but also from regulatory mechanism that change their expression or their phosphorylation status – similarly to canonical eIFs (Warnakulasuriyarachchi et al., 2004; Lewis et al., 2007; Chan, 2014). A further example of this type of translational control in IRES-mediated translation initiation is the of the fragile X mental retardation protein (FMRP). FMRP is an RNA binding protein ubiquitously present in brain and involved in the correct function and formation of synaptic network (Pfeiffer et al., 2009; Washbourne et al., 2014). The absence of this protein causes the Fragile X syndrome (FXS), a severe neurological disorder with intellectual disability and autism (Bassell at al., 2008; Sethna et al., 2013). It has been recently demonstrated that the transcript that encodes for FMRP undergoes both cap-dependent and IRES-mediated translation. The IRES-mediated translation is regulated by the presence of the heterogeneous nuclear ribonucleoprotein Q (hnRNPQ), that acts as ITAF. As expected, in condition in which hnPRPQ expression is upregulated, the

IRES-mediated FMRP expression increases, causing the axonal growth cone to collapse during axonal outgrowth (Choi et al. 2019).

3.1.3 Local Translation

The rapidity and responsiveness of the protein synthesis modulation obtainable with a translational control, make this regulation very suitable in cells with specific needs, such as neurons. In particular, the unique neuronal morphology and the function in communicating complex information far away from the soma, require a localized protein synthesis at the synapses – known as local translation - that must be tightly regulated to avoid severe dysfunctions.

Synapses are junctions formed between two specialized structures: the axon terminal of the presynaptic neuron, transmitting the information, and dendritic spines of the postsynaptic neuron that receives the information. During neuronal development, axons extend for great distances, up to hundreds of centimeters in vertebrates, to reach the postsynaptic targets and the growth cone – the specialized tip of the growing axon – is guided by multiple local clues. Also, dendrites are highly branched further contributing to the lengthening of the neuronal shape. Thus, while the neuronal body contains the nucleus of the cell, the 99% of cytoplasm is contained in the axon and the dendritic tree emanated from the body, resulting in a great distance among the various neuronal compartments (Holt et al., 2019). This leads to two conclusions that demonstrate the importance of the local translation and its regulation: first, it is more efficient to produce the required protein where its action is needed, avoiding delays and complications in protein transport; second, the distance between the neuronal body and thecompartments of neurites makes the transcriptional regulation of gene expression unfavorable, due to not only the slowness of transcription, but also the transport of required mRNAs at very distant cellular extremities.

In addition to the morphology of the neuron, the fast and dynamic nature of synaptic communication can benefit from a local translation mechanism of gene expression. Synaptic features such as the number of connections, the size and strength of synapses in the brain change in an experience-dependent manner, a process known as synaptic plasticity. Brain functions, such as learning and memory, depend on the dynamic variation in the connectivity between neurons, triggered by external stimuli. This type of long-lasting synaptic plasticity events, such as long-term memory, requires localized protein synthesis at the individual synapse or at a specific synapsis cluster. Therefore,

both the morphological complexity and the synaptic plasticity of neurons, require a translation control of local protein synthesis to set up, maintain and dynamically modify the proteome in axonal and dendritic compartments. As a result, the localization of mRNA and the local translation of these transcripts at the synapses have evolved to meet the local demand for new proteins at short time scales, allowing to fulfil essential neuronal functions as synaptogenesis, experience-dependent plasticity and the physiological maintenance of the axon and dendrites (Holt et al., 2019).

Local translation can be achieved thanks to two specific features: i) the transport of the transcripts in the place in which the local translation occurs; and ii) the activation of translation by a signaling cascade that transduces the synaptic stimulus (Moine et Vitale, 2018). The presence of polyribosomes – the ribosomes involved in active translation of the transcript – in synaptic compartments (Tennyson, 1970; Steward and Levy, 1982) and the microtubule-dependent mRNA transport to the distal sites of local translation have been thoroughly described. As previously demonstrated for the β-actin transcript in neurons, "zip-codes" localization elements can be detected on the 3'UTR of mRNAs. Zip-codes can bind RNA-binding proteins (RBPs), such as ZBP1, that in turn interact with motor proteins (Subramanian et al., 2011; Eliscovich et al., 2013). The zip-code sequences are highly heterogeneous (Jambhekar and Derisi, 2007) and there are also zip-code structures, such as the G-quadruplex (G4) that was identified in many transcripts locally translated in dendrites. A "sushi belt" model for mRNA localization in response to stimulation, pushing the transcript in the site of translation, predicts that mRNA transport through microtubules is triggered by request from a "hungry" activated synapse (Doyle and Kiebler, 2011). Of course, the correct transport of mRNA and ribosomes requires that the mRNA do not start to be translated before the arrival at the synapse localization. Translation inhibition during the transport assures the correctness of the processes and it happens preferentially by the low efficiency of the mRNA recruitment and the block of the scanning of the 5'UTR. Once the transcript is in the correct site, the phosphorylation state of eIFs, in response to synaptic activation, can resume translation. For example, the phosphorylation state of eIF2α depends on synaptic stimulations. When there is the need of an enhancement of the synaptic strength, such as under the treatment with BDNF, eIF2α is dephosphorylated, allowing to an increase in the local translation (Moon et al., 2018). Another mechanism involves the phosphorylation of eIF4E by MAPK kinases (Mnk1/2) at Ser209, with a decrease of its affinity to the capped transcripts. The extracellular

signal-regulated kinase (ERK), activator of Mnk1/2, triggers also the signaling of mammalian target of rapamycin (mTOR) (Kelleher at al., 2004), that has two main effectors: 4E-BPa and p70 S6 kinases (S6K1/2). Therefore, mTOR represent a main controller of the translation initiation because it alleviates the repression on eIF4E by phosphorylation of 4E-BPs (4E-BP1 and 4E-BP2), and activates S6K1/2, leading to the activation of translation through the consequent phosphorylation of S6 ribosomal protein (Gingras et al., 2004).

The neuronal, activation-mediated, local translation also acts with other, more specific mechanisms. For example, eIF4B enhances the translational efficiency of specific transcripts, especially of genes involved in cell survival and proliferationacting mainly as an adjuvant of the unwinding activity of eIF4A (Shahbazian et al., 2010). Like other regulators of translation, the activation of the eIF4B is mediated by phosphorylation. Two residues, Ser406 and Ser422, increases the association of eIF4B at the PIC. In neurons, the phosphorylation status of eIF4B is modulated by neuronal activation, but, interestingly, an additional, neuronal-specific phosphorylation at site Ser504 has been characterized (Bettegazzi et al., 2017). The phosphorylation at Ser504 is triggered by the activity of the metabotropic glutamate receptors (mGluRs), also reported to be important contributors to translation-mediated synaptic plasticity (Weiler et al., 1996; Luscher and Huber, 2010; Gladding et al., 2009). While the eIF4B phosphorylation at Ser422 and Ser406 is mediated by the P13K-mTOR and MAPK pathways, Ser504 phosphorylation is sustained by a different pathway that involves CKs and the activation of cPKCs (Bettegazzi et al., 2017). Thus, eIF4B can act as a point of convergence of the multiple signaling pathways involved in the control of local translation in neurons (Bettegazzi et al., 2017; Moine and Vitale, 2018), further suggesting that activation-mediated translation at synapses is tightly regulated by neuronal specific phosphorylations and eIFs integrated functions. Moreover, the finding of an altered Ser504 phopshorylation state of eIF4B in an epileptic rat model (Bettegazzi et al., 2017) suggests how a malfunctioning in these regulatory mechanisms of synaptic plasticity might lead to neurological diseases.

3.1.4 Translational dysregulation in Neuronal Disorders

Malfunctioning of eIFs could be related to the neurodegenerative diseases, such as Alzheimer's and Parkinson's Diseases (AD and PD) (Buffington et al., 2014; Jishi et al., 2021). In fact, the ribosome might be one of the targets of pathological tau

aggregates in the early stages of the AD progression, resulting in an impaired ribosomal activity and a decreased translational rate in neurons (Ding et al., 2005; Meier et al., 2016). Because of the need of continued and tightly regulated translation to properly maintain the synaptic activity, this impairment could be a contributing factor to memory loss and even neuronal death observed in AD patients (Jingshu et al., 2019). The eIF2α phosphorylation state is also greatly increased in AD patients (Ohno et al., 2014) and the genetic deletion of the eIF2α kinase PERK in an AD mouse model reduces the abnormal translation that impairs synaptic plasticity and memory deficits (Ma et al., 2013). These findings suggest a role for eIF2α in the AD pathogenesis, even if it is not the only eIFs potentially involved. In fact, the expression of other eIFs, such as eIF3η and eIF5, is reduced in the hippocampus of AD patients (Hernández-Ortega et al., 2016), suggesting that the impaired translation control is a general AD feature (Ghost et al., 2020). Similarly, one of the most common causative gene of both sporadic and familiar PD, the leucine-rich repeat kinase 2 (LRRK2), is known to regulate the translational efficiency of kinases phosphorylating 4E-BPs. Therefore, mutated LRRK2 isoform found in PD patients often cannot carry out the kinase function, causing a disregulation of protein synthesis at the synapse leading to pathology (Teleman et al., 2005; Imai et al., 2008).

Aberrant neuronal translation has also emerged as a common feature of neurodevelopmental disorders, further underlining the importance of translation regulation in neuronal health. Among these neurodevelopmental disorders there are the autism spectrum disorder (ADS) and the Fragile X Syndrome (FXS) (Jishi et al., 2020). In the case of ASD, the atypical neuronal network consists in hyperconnected neurons that show a pathological imbalance between the amount of excitatory and inhibitory synapses (Gatto et al., 2010; Supekar et al., 2013). The number of excitatory and inhibitory receptors is controlled by the expression of adhesion molecules, such as neuroligins and other scaffolding proteins at the synapse. The enhancement of neuroligins translation has been linked to ASD development as a consequence of the loss of 4E-BP2 (Gkogkas et al., 2013). The role of translational control through 4E-BPs/eIF4E mechanism in ASD etiopathology is also confirmed by the fact that the phenotype of 4E-BP2 KO mice presenting autistic-like behaviors is rescued by the inhibition of eIF4E (Gkogkas et al., 2013; Cao et al., 2018). Moreover, mutations that affect the mTOR pathway in ASD patients can also impair protein translation by affecting eIF4E/4E-BPs interactions. This suggests that also aberrant 4E-BP

phosphorylation state can lead to unphysiological translational changes and the development of ADS badly influencing the synaptic plasticity (Santini et al., 2014). It is worth of notice that most of the mutations associated with the development of ADS do not directly affect the composition of synapse but, rather, the pathways linked to protein synthesis, causing synapse damage by impairing local translation.

Regarding FXS, it has been reported that decrease in FMRP expression level or reduction in its activity lead to the enhancement of local translation, that results in autism-related phenotypes. Most of the characterized mutations found in FXS patients affect FMRP, PTEN, TSC1, TSC1 and eIF4E genes, that encode for translational effectors involved in mTOR pathway (Kelleher et al 2008). FXS is caused by the trinucleotide CGG expansion in the 5'UTR of the gene fmr1, encoding the protein FMRP. Patient with a great number of repetitions (>200) display a severe reduction in FMRP expression (Richter et al., 2015), whose absence leads to pathological increases in the translation of other proteins. In fact, FXS patients show increased global level of translation, whereas Fmr1-KO mice present greater density of dendritic spines due to enhanced local translation, independently from synaptic activation (Jacquemont et al., 2018; Comery et al., 1997).

In conclusion, the central role of translational dysregulation in neuronal diseases introduces two main questions in the understanding of these pathologies: first, if the rescue of the physiological translation through innovative therapeutic strategies could reverse the decline of brain functions showed in these disorders; and, consequently, what could be the most efficient and safe strategy to target translational control avoiding undesirable side effects (**Fig. 3.1.4**).

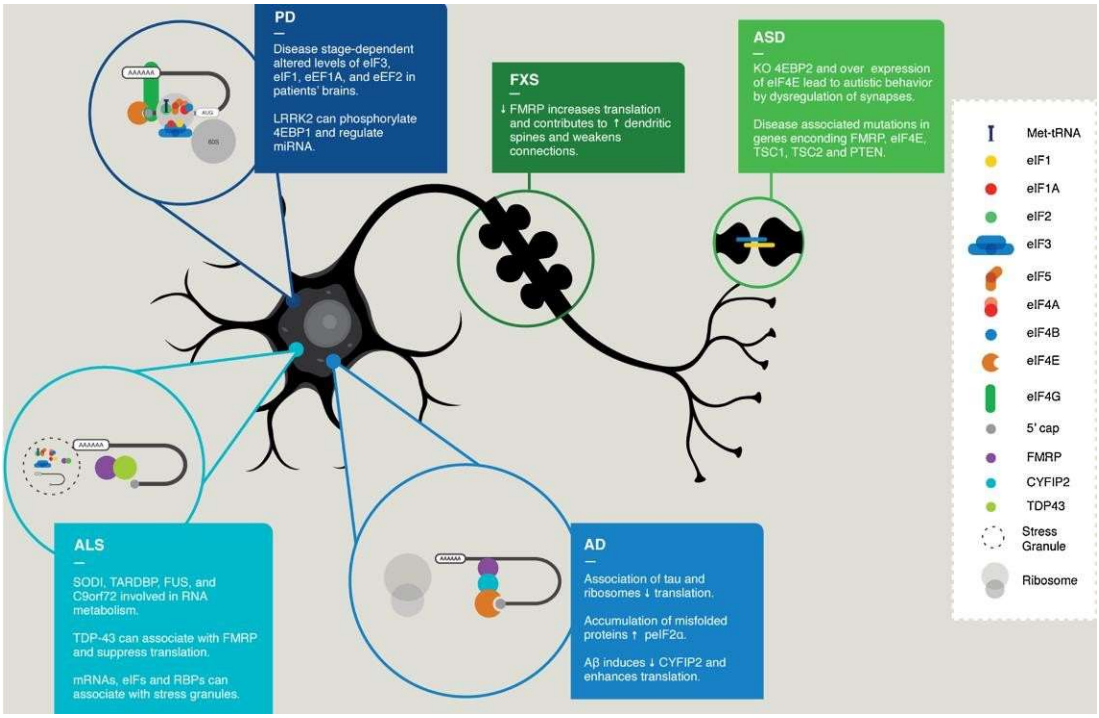


Fig. 3.1.4 Aberrant translational control involved in neurological disease*. In the figure are reported the different type of dysregulation of the translation control of mRNA linked to the development of neurological disease with a summary of how they affect neuronal health. Autism Spectrum Disorder (ASD) is associated to the loss of 4E-BP2 and the overexpression of eIF4E. Mutations in trans-acting protein, such as FMRP, eIF4E, TSC1, TSC2, and PTEN are also recognized to be causative of ASD. Fragile X Syndrome (FXS) is caused by CGG expansion in FMRP 5'UTR. Reduction in the physiological level of FMRP lead to increased translational efficiency in dendritic compartments. Alzheimer's Disease (AD) involves tau aggregation to the ribosomes, impairing translational efficiency. Increased phosphorylation of eIF2α is also reported. amyloid beta (Aβ) pathologically increases protein synthesis. Parkinson's Disease (PD) is associated with altered level of eIF1, eIF3, eEF1A, eEF2. Moreover, 4E-BP1 is a target of the kinase LRRK2. LRKK2 regulates also miRNA involving in translational control. Amyotrophic Lateral Sclerosis (ALS) is caused by mutations in genes involved in mRNA metabolism. This figure is taken from Jishi et al., 2020.*

3.2 5' Untranslated Region

3.2.1 Characteristic Features of Human 5'UTRs

Translational control involves trans-acting factors and mRNA cis-acting elements, both of which are conserved across eukaryotic species, suggesting how fundamental this type of control of gene expression is. 5' untranslated region (5'UTR) - also known as leader sequence - plays an important role in this process, containing different cis-

regulatory elements, proper to each transcript. 5'UTR represents an essential platform for the ribosome recruitment to the mRNA and is involved in the regulation of both cap-dependent and cap-independent translation.

During the evolution, the more eukaryotic genomes greatly increased their size, the more the UTRs showed an expansion in length, probably to be more accessible to more trans-acting factors or to contain multiple cis-regulatory elements to modulate in a tighter way gene expression. If the length of 3'UTRs has a great increase in eukaryotic evolution, the 5'UTR also shows a relevant increase, with a median length of 53-218 nucleotides among different eukaryotic species (Pesole at al., 2001; Lynch et al., 2005; Mignone et al., 2002). Among these, human 5'UTRs has the longest recorded median length. Nevertheless, 5'UTR length greatly varies among individual genes of a species ranging from a dozen to hundreds of nucleotides, suggesting a difference in the presence of regulatory cis-acting elements in specific mRNA subset (Hernàndez at al., 2010) (**Fig. 3.2.1.1**).

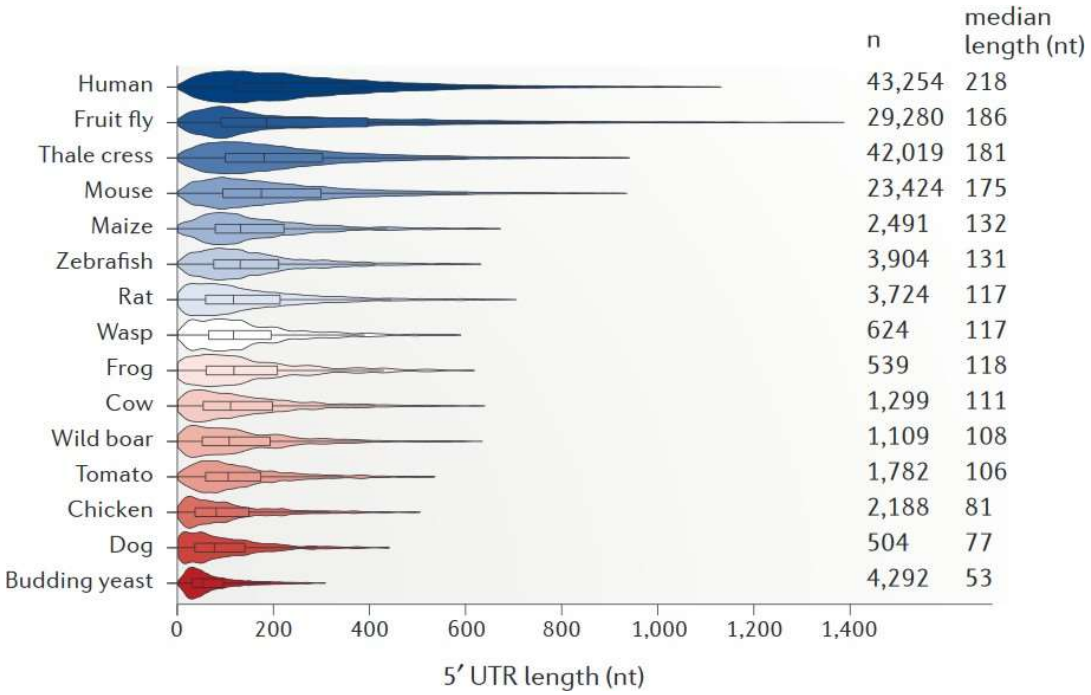


Fig. 3.2.1.1 5'UTR Length through Evolution. *The median length of 5'UTRs is increased trough evolution, ranging from 53 to 218 nts in the considered species. The violin plot is built from at least 100 RefSeq validated 5'UTR sequences for each specie. The 15 species took into consideration include (Homo sapiens), fruit fly (Drosophila melanogaster), thale cress (Arabidopsis thaliana), mouse (Mus musculus), maize (Zea mays), zebrafish (Danio rerio), rat (Rattus norvegicus), wasp (Nasonia vitripennis), western clawed frog (Xenopus tropicalis), cow*

(Bos taurus), wild boar (Sus scrofa), tomato (Solanum lycopersicum), chicken (Gallus gallus), dog (Canis lupus familiaris) and the budding yeast Saccharomyces cerevisiae. This figure is taken from Leppek, Das and Barna, 2018.

Another evolutionary feature of 5'UTRs compared with 3'UTRs, is the higher percentage of UTRs with introns, whose trend is conserved among the species (**Fig. 3.2.1.2 A**). Nearly 30% of genes of multicellular animals possesses 5'UTRs in which at least the first exon is totally untranslated, whereas the much longer 3'UTRs show a lower intron frequency, ranging from 1-11% (Pesole et al., 2001; Mignone et al., 2002). The most represented regulatory elements within 5'UTRs are secondary structures, uAUGs/uORFs and IRES (Hinnebusch et al., 2016).

Stable secondary structures, characterized by lower ΔGs, can lead to an inhibition of the translation efficiency of the transcript, impeding the ribosome translocation to the main AUG (Hinnebusch et al., 2016; Leppek et al., 2018). This blockage of scanning mechanism can be overcome increasing the eIF4A unwinding activity. In contrast with some 5'UTR cis-regulatory elements, it is very hard to estimate a percentage of secondary structures contained in 5'UTRs from a phylogenetic point of view. This difficulty is due to the heterogeneity in length, strength and base-pairing of 5'UTR structures and thus the absence of a consensus motif – especially for Watson and Crick (WC) structures. Even though it has been proposed that 5'UTR length could influence the stability and the complexity of the structures, the most useful parameters to predict secondary structures is the GC content, expressed as GC% (Pesole et al., 2001;Mignone et al., 2002). In fact, the GC base pairs are stronger than AU and GU pairs, because of the presence of three hydrogen bonds that held together the couple of nucleotides, bringing higher thermostability to the structure – and thus more negative ΔGs. 5'UTRs GC content in human transcript leaders is reported to be averagely 60% (Pesole et al., 2001), value used as indicative threshold to discriminate potential highlyregulatory from poorly regulatory 5'UTRs (Davuluri et al., 2000). Interestingly, GC content of 5'UTRs negatively correlates with their length (Pesole et al., 2001), strengthen the concept that the length is not the unique parameters to take into consideration in evaluating the strength of secondary structures. Secondary structures can be formed using also Hoogsteen base pairing, and, thus, G4s can also act as 5'UTRscis-regulatory structural motifs modulating translational efficiency. 5'UTR G4s can act as either inhibitory elements of translational efficiency, similarly to the WC secondary

structures, or IRES, even enhancing the protein synthesis of certain transcripts (Bugaut at al., 2012). Concerning IRES, these are suggested to be present in the 5'UTR of about 10% of human transcripts (Mignone at al., 2002), modulating the cap-independent translation of mRNA particularly relevant for the cell life (**Fig. 3.2.1.2B**).

About 50% of human 5'UTRs contains uAUGs that are predicted to reduce the translation efficiency of the main ORF (Mignone et al., 2002; **Fig. 3.2.1.2 B**). Their presence positively correlates with the length of 5'UTRs, and the Kozak context in which they lied is often not optimal enough to turn off the translation of the main ORF.

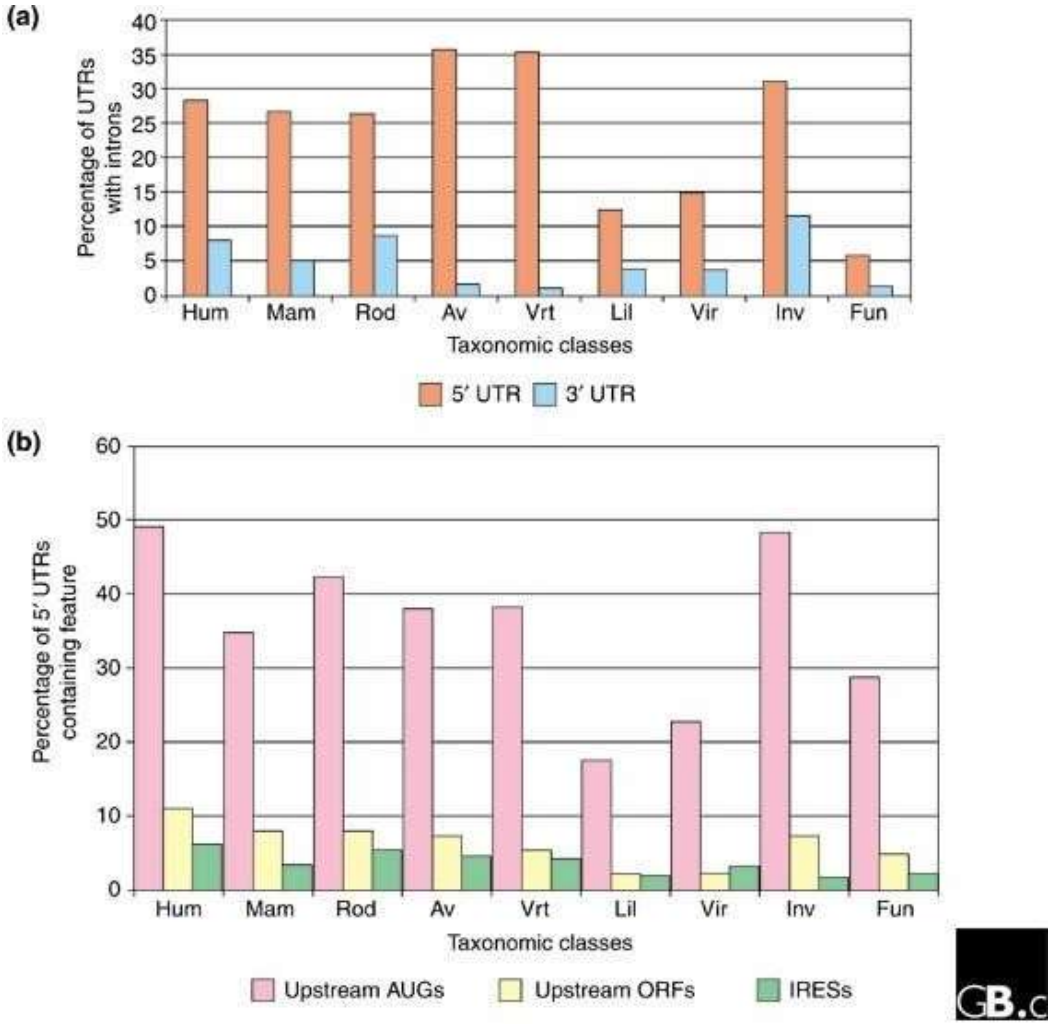


Fig. 3.2.1.2 5'UTR features in different taxonomic groups. *(A) Percentage of the introns contained in 5'UTR compared with 3'UTR. (B) Percentages of uAUGs, uORFs and IRES elements. Hum, human; mam, other mammals; rod, rodents; av, Aves; vrt, other vertebrates; lil, Liliopsidae; vir, other plants (Viridiplantae); inv, invertebrates; fun, fungi. Plots are taken from Mignone et al., 2002.*

Nevertheless, functional uAUGs represent a cis-acting regulatory element that can

reduce the expression level of the transcript. In fact, the recruitment of the ribosome on a uAUG, especially when it starts the translation of an uORF, strongly reduces the synthesis of the protein due to the inability of eukaryotic cells to reinitiate the translation from a downstream AUG – especially when uORFs are long and their distance from the main AUG short (Luukkonen et al., 1995; Wang et al., 2004).

During the translation initiation there are several steps in which translation control can occur. The richness of cis-acting elements on 5'UTR sequences increases the number of available tools to modulate protein synthesis of a specific transcript. Each cis-acting element can involve various trans-acting factors, whose action can be modulated through phosphorylation or their abundance in response to the different circumstances that the cell faces. Therefore, 5'UTRs containing a wide repertoire of cis-acting elements often belong to transcripts that undergo a tight translation control. In fact, most of the mRNAs subjected to translational control, are generally involved in the most important functions of the cell, and they share the presence of multiple cis-acting elements. This allows to predict transcripts that are undergoing translational control, expanding the possible repertoire of translationally controlled genes. The importance of these studies is also due to the growing evidence of mutations that, by disrupting 5'UTRs cis-acting regulatory elements, can cause human diseases. Therefore, the importance of the screening of 5'UTR region in diagnostic processes, the understanding of 5'UTRs influence in setting the expression level of proteins, and the development of new therapeutical strategies targeting 5'UTRs start to be field of studies of great interest.

3.2.2 Translational Control by upstream ORFs and upstream AUGs

uAUGs are defined as start codons upstream the main AUG, even if, given the fact that AUG is not the unique start codon used in eukaryotic translation, it would be more accurate referring to upstream start codon in the broad sense. Nevertheless, because of the very low frequency of usage of the alternative start codons, this traditional nomenclature has remained in use. uORF are short coding region starting from uAUGs, although not all uAUGs are in a good context for translation initiation. uAUGs can be divided in in-frame and out-of-frame uAUGs referring to the frame of the main start codon. Also, uORFs can be divided in different classes, consisting of totally upstream uORFs, whose stop codon is placed upstream the main start codon; frame uORFs, that share the same stop codon with a downstream ORF; overlapping uORFs, whose coding

sequence overlaps with other ORFs (Iacono et al., 2005). uAUGs are the most represented 5'UTR cis-acting elements, since genome-wide sequencing of mammalian transcripts confirmed their presence in about 50% of them (Ingolia et al., 2009; Lee et al., 2012). Nevertheless, their evolutionary conservation is generally moderate, as emerged from the comparison of subsets of human uAUGs and uORFs with mouse and rat genomes. In fact, only the 24% of uAUGs are conserved in all three species, while the conserved uORFs are estimated to be around 38%. (Iacono et al., 2005).

Despite the widespread occurrence of predicted uAUGs/uORFs in 5'UTRs, direct evidence showing that these elements perform a negative modulation of the downstream transcript translatability, are collected uniquely for specific subset of transcripts, encoding proteins such as transcription factors, growth factors and proto-oncogenes (Davuluri et al., 2000). Both uAUGs and uORFs show great variability of their efficiency as negative regulators of translation, changing in length - in the case of uORFs -, distance from 5'cap and main AUG, and Kozak context quality (Calvo et al., 2009). It has been experimentally assessed that regulatory uAUGs/uORFs display typical features, such as great distance between the 5'cap and the uAUG, conservation among species and a strong Kozak context (Kozak et al., 2001).

Currently, several translational control mechanisms involving uAUGs/uORFs are reported in literature, even if most of them can be traced back to two groups.

- The first group is characterized by the need of the translation of a growing peptide after the meeting between the ribosome and a functional uAUG/uORF. This causes the stalling of the elongating 80S ribosome, creating a roadblock for the scanning of the 5'UTR or the dissociation of the ribosomal subunits after the uORF translation, making difficult the reinitiation of translation of the main ORF.
- In the second group, the presence of the encoded peptide from the uORFs is irrelevant, because the regulatory element is the upstream start codon acting simply as interceptor of the scanning PICs.

The competitive barrier provided by such uAUGs is generally overcome by a mechanism known as "leaky scanning", an expression that means that there is a percentage of the scanning PIC that does not recognizes the uAUG as start codon and will continue the scanning to reach the downstream main AUG. Consequently, uAUGs lying in better Kozak context show a stronger inhibitory effect, because intercepts the major number of scanning PICs (Lee et al., 2012; Calvo et al., 2008). The ability of

PICs to leaky scan uAUGs is not yet completely understood, but it is favored in stress conditions, demonstrating the possible dynamic regulation by these cis-acting elements. The increased PICs ability to overcome uAUGs correlates with the increase of the phosphorylation of the Ser51 of eIF2α (Andreev DE et al., 2015), leading to the decrease of the TC assembly, and thus, probably, allowing the bypassing of uAUGs by the scanning PICs (Hinnebusch et al., 2005). A well-known case of differential, uORF-mediated translation control occurring in stress response is provided by the transcript of the General Control Nondepressible (GCN4) transcription factor in yeast. The GCN4 5'UTR contains four uORFs, the first of which is always translated regardless of the nutritional condition. In physiological condition, the rapid reloading of the ribosome allows the translation of the other three uORFs, inhibiting the translation of the main ORF, encoding GCN4. In condition of starvation that activates the kinases of eIF2α, such as GCN2 (for amino acid deprivation) and PERK (for ER-stress), there is a deceleration in ribosome reloading, due to the decreased abundance of TC level provided by eIF2α. Therefore, these events modulate the leaky scanning ability of scanning PICs, increasing the possibility of overcoming the translation of the last three uORFs, and thus increasing the initiation of the translation of the main ORF. This mechanism allows to strategically translate GCN4 only when required in condition of nutritional shortage (Mueller et al., 1986; Hinnebusch et al., 2005). This role of uAUGs/uORFs in repressing translation of specific transcripts that dynamically evade this limitation under stress condition using the phosphorylation of eIF2α as a switch, is used in different contexts in humans, starting from the mammalian equivalent of GCN4, ATF4, activated under condition of hypoxia and amino acid deprivation (Blais et al., 2004; Palii et al., 2009; Lewerenz et al., 2011). Other examples experimentally assessed in human are: BACE1, under condition of reduced energy and endoplasmic reticulum stress condition (O'Connor et al., 2008; Mouton-Liger et al., 2012), CD36, under condition of high glucose (Hamel et al., 2001), and other transcripts, some of which are summarized in the **Fig. 3.2.2.1** (Barbosa et al., 2013). Therefore, this mechanism appears important in brain functions such as learning and memory, and in the development of neurodegenerative diseases and cancerogenesis (Hinnebusch et al., 2016).

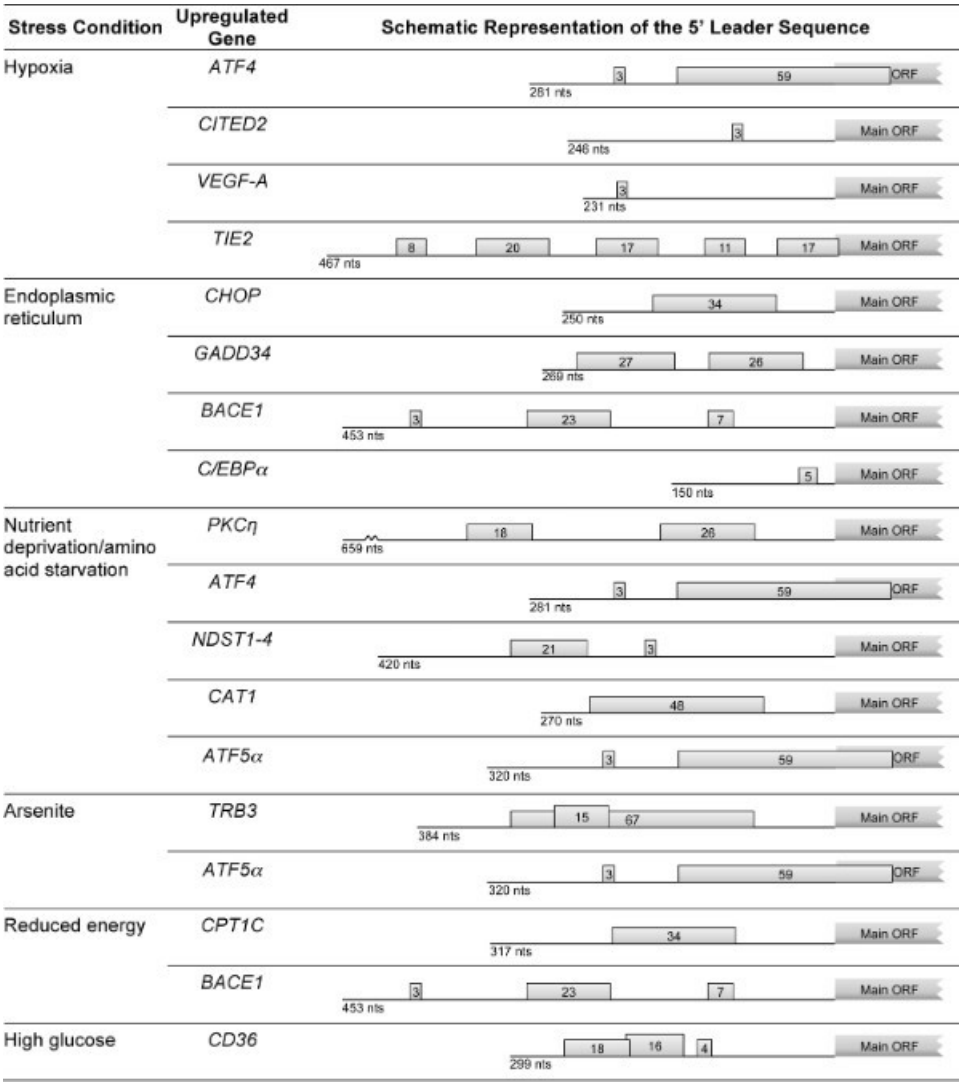


Fig. 3.2.2.1 Examples of human transcripts involving uORFs in differential translational control under stress condition. *For each transcript, it has been reported the schematic representation of 5'UTR with the length in nucleotides (nts); boxes represent uORFs, where number indicated the lengths in codons. Transcripts are grouped depending on the type of stress condition that trigger the evading from the uORF inhibition effect, with a mechanism like the one described for GCN4 in yeast. Figure is taken from Barbosa et al., 2013, where are reported all the references for the transcripts.*

An additional possible outcome of the presence of in-frame uAUGs is the production of different protein isoforms encoded from the same genes but with a different N-terminus. Often these isoforms play specific roles in context such as tissue differentiation and development - as reported, for example, in mouse liver when C/EBPβ transcript is translated in three isoforms from consecutive in-frame uAUGs (Wethmar et al., 2010a). Moreover, the ribosomal stalling at the uORF stop codon can

trigger the nonsense-mediated mRNA decay (NMDA), due to the recognition of the uORF stop codon as a premature one, leading to the transcript loss, as reported for the STN1 transcript (Torrance et al., 2018). There are also intriguing cases in which the presence of an uORF positively correlates with the translation of the main ORF. This finding, on contrary to what expected from the competition with the uORF - is due to a possible additional regulatory role of the "near-cognate" uORF that promotes the reinitiation ability of the ribosome, leading to the efficient translation of the main protein under specific condition and with the involvement of various eIFs (Munzarovà et al., 2011; Ingolia et al., 2012).

The occurrence of interesting mutation of uAUGs/uORFs in human disease is well-described, since more than 3700 variants identified in 5'UTRs alter the uORFs in disease-associated variant databases (Ye at al., 2015). It is easily predictable that among the most harmful reported mutations there are the ones that eliminate or create an upstream start codon, even if also mutations altering Kozak context are reported to be causative of diseases (Wethmar et al., 2010 b; Ye at al., 2015). A well-known example in this context is the formation of a new upstream start codon in the SOX9 5'UTR, that reduces the protein expression causing a rare and severe developmental disease called campomelic dysplasia (von Bohlen et al., 2017).

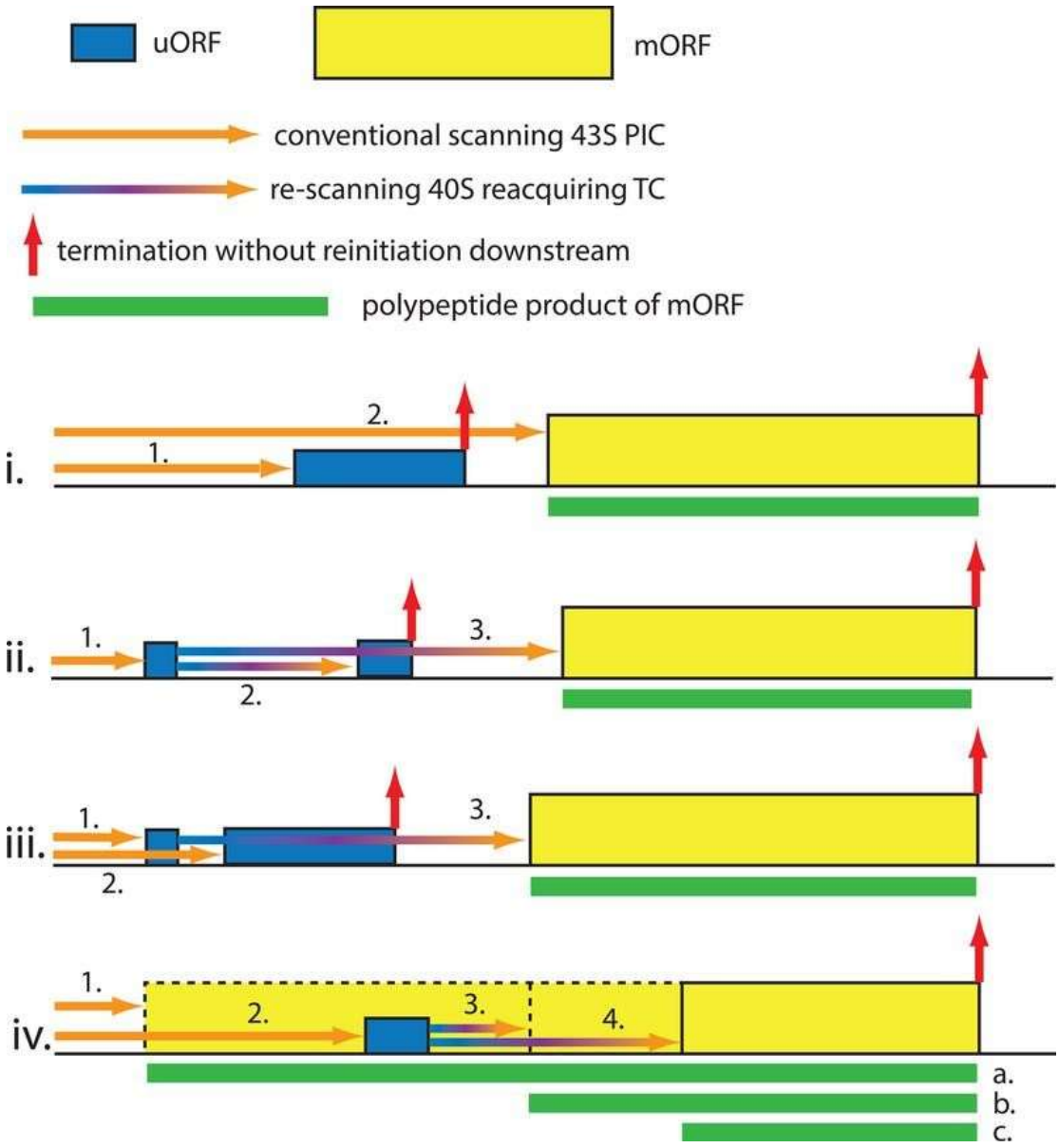


Fig. 3.2.2.2 uORF-mediated translational control in different gene architectures. *i) scanning PICs fail to reinitiate the translation of the main ORF, even if a certain percentage successfully leaky-scan the uORF in a not optimal Kozak consensus sequence and thus reach the main ORF. ii) scanning PICs initiate the translation of an uORF that does not preclude reinitation. Reinitiation can involve the translation of a second inhibitory uORF or the translation of the main ORF. iii) the translation of the first uORF allows the translation of the main ORF. On the contrary, leaky scan of first uORF allows the inititation of a second inhibitory uORF, precluding the translational initiation of the main ORF. iv) scanning PICs initiate the translation of a first uORF that is in frame with main ORF, resulting in the translation of a long protein isoform, with a particular N-terminus. Leaky scanning causes the not recognition of the first start site and initiate the translation of a second uORF that can inhibit the translation of the main ORF or can produce another protein isoform. The not optimal Kozak context of all the uORF allows the uniquely the translation of the main ORF. Figure is taken from Hinnebusch et al., 2016.*

An example of the loss of a functional uAUGs/uORF leading to development of a disease is the disruption of the uORF in the thrombopoietin (TPO) 5'UTR resulting in the increase of the TPO protein level that causes the hereditary thrombocytosis (Kondo et al., 1998; Ghilardi et al., 1998). uAUG/uORF alterations associated with the development of a pathological status are also found in the interferon regulatory factor 6 (IRF6) gene associated with Van der Woude syndrome (VWS) and the popliteal pterygium syndrome (PPS), where the appearance of a new uORF caused by a SNP inhibits by 70-100% the production of the protein (Kondo et al., 2002), and in the development of Saethre-Chotzen syndrome (SCS), where aberrant translational control of the twist-related protein 1 (TWIST1) is caused by two different SNPs that lead to the formation of a novel uAUG in a favorable Kozak context (Zhou et al., 2018). Proto-oncogenes and tumor suppressors such as c-MYC, B-cell lymphoma (Bcl-2), phosphatase and tensin homolog (PTEN) and tumor protein p53 (TP53) have mutations in uAUGs/uORFs, associated with the development of cancer (McGrillvray et al., 2018). A comprehensive list of diseases caused by mutations in uAUGs/uORFs is reported in the review by Barbosa et al., 2013, where there are briefly reported the pathogenic role of each of them.

3.2.3 Functional 5'UTR Structures

5'UTRs, especially those with a high GC content, can form thermostable cis-acting secondary structures that, in turn, might impair the translational efficiency by impeding the mRNA scanning (Kozak, 1986; Leppek et al., 2018).

The first study assessing the influence of a 5'UTR secondary structure on translation efficiency is dated 1985. It demonstrated how localized mutagenesis, aimed to increase the thermostability of a hairpin in the 5'UTR of the gene thymidine kinase (tk) of the herpes simplex virus 1 (HSV1), changed its translational rate (Pelletier and Sonenberg, 1985). A similar effect was later observed in the translation of the rat ornithine decarboxylase (ODC), in which a long stem-loop of 115 nts, turned out to set the protein level (Manzella et al., 1990). As emerged from these studies, the presence of a single structural motif can severely impair translation. So, when looking at potential effects of 5'UTRs, every structural element in the sequence must be considered. Accordingly,

even if the total GC content of 5'UTR is considered descriptive of the potential regulatory role of the 5'UTR, it must be taken into consideration also the relative GC content of single secondary structures to avoid losing potential local regulatory elements. In fact, local variations in GC content can lead to strong folding, and, thus, to a steep drop in translation efficiency downstream the hairpin (Babendure et al., 2006). Moreover, the position of the secondary structures is another important feature in predicting the strength of the inhibitory effect on translation: more the structures are placed in the proximity of 5'cap, more the translation of the gene downstream the hairpin usually shows a decrease in its translation, with the greater effect in the first 30nts downstream the 5'cap (Babendure et al., 2006).

Even if hairpins formed by Watson and Crick base-pairing are traditionally considered the structural cis-acting element of 5'UTRs, non-canonical secondary structures formed by Hoogsteen base-pairing, such as Gquadruplexes (G4s), can also play a similar role in translational control. The first evidence of a block of ribosome scanning caused by a G4 in 5'UTR sequence was found in the NRAS proto-oncogene transcript, and later strengthen by a similar case in the zinc finger protein Zic1 5'UTR (Kumari et al., 2006; Arora et al., 2018). Similar to what was assessed for WC secondary structures, G4 thermostability, as well as the proximity to the 5' of the leader sequence, plays an important role in determining the strength of inhibition (Kumari et al., 2006).

In addition to the role of steric hindrance for the ribosomal accessibility to the transcripts, secondary structures can also act as platforms for protein binding of trans-acting factors (Leppek et al., 2018). The most studied case is represented by the translational regulation of ferritin and other proteins involved in the iron homeostasis. The translation of the corresponding transcripts is mediated by the binding of the iron-regulatory proteins 1 and 2 (IRP1 and IRP2) on a small stem-loop closed to the 5'cap, named iron responsive element (IRE). When an IRP binds the IRE, the translation initiation of the mRNAs is prevented, because of the sterical impediment of the complex that blocks the ribosome recruitment at the 5'cap (Hentze et al., 1987; Gray et al., 1994; Muckenthaler et al., 1998). A similar mechanism involves G4s, which can act as binding site for trans-acting factors enhancing the inhibitory effect on translation, as observed in for the protein phosphatase PP2Ac, in which the negative regulator FMRP binds a 5'UTR G4 (Pany et al., 2019).

RNA secondary structures in the same 5'UTR have the possibility to interact with each other to form highly ordered structures, such as pseudoknots. Pseudoknots are defined

as intramolecular RNA structures formed at least by two stem-loop helices (Leppek et al., 2018). The action of 5'UTR pseudo-knots in translational control is documented in the expression of the human interferon gamma (IFNG), where a pseudo-knot dynamically acts in response to the PKR activity levels to avoid IFNG overexpression (Ben-Asouli et al., 2002; Cohen-Chalamish et al., 2009). The interplay between secondary structures can also interest Watson and Crick stem-loops and Hoogsteen G4s, forming hybrid high-order inhibitory structures, such as the one studied in the hepatocyte nuclear factor 4-alpha 1 (HNFα1) (Guo et al., 2018).

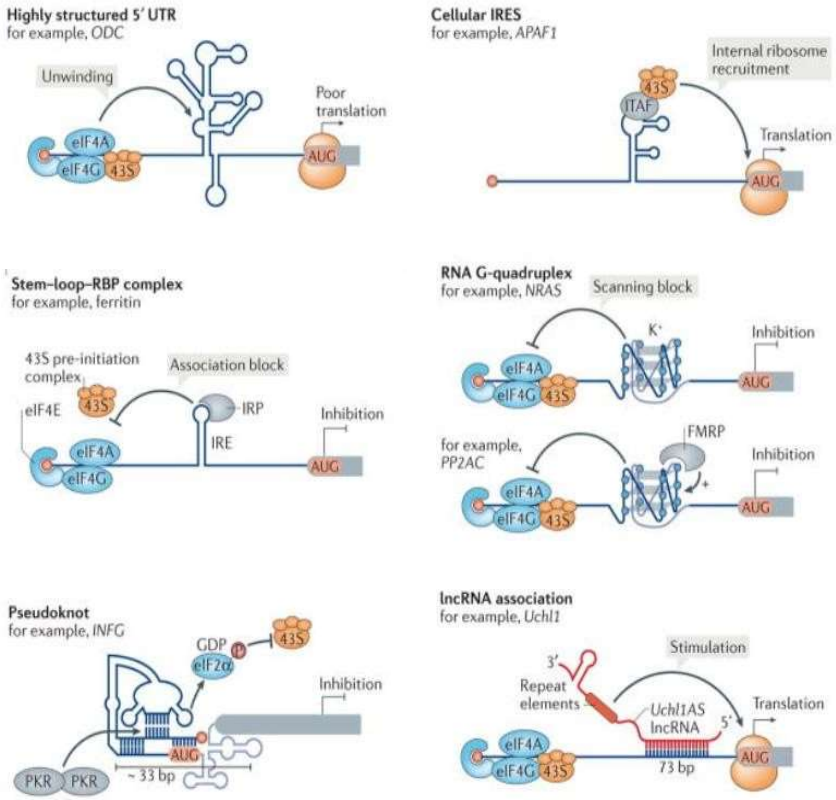


Fig. 3.2.3 Cis-acting regulatory structures in cap-dependent translation initiation. *Different types of translational control mechanisms involving cis-acting structures in cap-dependent initiation. IRE element is reported as example of stemloop-RBP complex. Highly structured region of 5'UTR, as the stem-loop in the ODC 5'UTR, is presented as example of steric barrier to the scanning PICs. 5'UTR G4 can act similarly to Watson and Crick secondary structures, creating a block of the scanning, such as the G4 in NRAS transcript, or recruiting RBPs, such as the 5'UTR G4 in PP2AC that binds FMRP. In addition, also high-order structures, such as*

the pseudoknots in the INFRG 5'UTR, resulting from the interplay of secondary structure can contribute to translational control. Modified from Leppek et al., 2018.

Despite the great heterogeneity in the composition of 5'UTR inhibitory structures, it is interesting to note that there is a major interactor which has the task of relax them: the DEAD-Box helicase eIF4A. This fundamental eukaryotic initiation factor is involved in the unwinding of both WC and Hoogsteen secondary structures, thereby acting as translational control mechanism (Wolfe et al., 2014). Ribosome profiling on cancer cells treated with silvestrol – a selective inhibitor of eIF4A activity - showed the reduction of the translational efficiency of e great number of transcripts, confirming the functional correlation between eIF4A and the presence of cis-acting regulatory elements in the 5'UTRs (Wolfe et al., 2014; Parsyan et al., 2011; Leppek et al., 2018). Defined as the "tactician" of the 5'UTR structures-mediated translation control, eIF4A expression is modulated in response to specific conditions and its activity can be enhanced by other modulatory proteins, such as eIF4B - on which the majority of the signaling cascades involved in translational control converges (Parsyan et al; 2011; Wolfe et al., 2014; Rubio et al., 2014).

Similarly to what reported for uAUGs/uORFs, and as demonstrated by Pelletier and Sonenberg with the expression of the tk transcript, mutations that impair the stability of 5'UTR secondary structures are associated with pathologies, such as the hereditary hyperferritinaemia-cataract syndrome (HHCS), in which the transition c.-151A > G causes the weakening of the IRE element in the L-ferritin (FTL) 5'UTR (Van de Sompele et al., 2017) or cancer-associated mutations in 5'UTR G4s as the ones of Bcl-2 and CXCL14 (Zeraati et al., 2017).

In addition to the translational control performed by eIF4A, eIF4B and other eIFs involved in solving the structural complexity of 5'UTRs, secondary structures are also fundamental elements in the formation of Internal Ribosome Entry Site (IRES), mediating the most studied mechanism of cap-independent translational initiation.

3.2.4 Internal Ribosome Entry Site

While cap-dependent initiation occurs in the 95-97% of the cases, most of the remaining initiates their translation in a cap-independent manner by IRESs (Komar et al., 2015; Godet et al., 2019). In IRES-mediated translation initiation the 40S ribosomal subunit is not recruited at the 5'cap structure but binds directly to specific sequences in

the 5'UTR. Currently, two general mechanisms of IRES-mediated initiation are proposed: the "direct landing", in which the 40S ribosomal subunit directly lands in the proximity of the main AUG, without the involvement of a proper scanning step; and the model "land and scan", in which there is a scanning step to reach the main AUG (Belsham et al., 200).

Eukaryotic IRESs are generally found in transcripts that encode proteins regulating cell proliferation and apoptosis, as well as stress-response, such as in hypoxia or lack of nutrients (Komar and Hatzoglou, 2011). In comparison with the viral IRESs, eukaryotic ones have less structured secondary structures (Martineau et al., 2004; Bonnal et al., 2005; Morfoisse et al., 2016), and do not share a common consensus sequence, making difficult to classify IRESs or to predict new endogenous IRESs on the basis on the 5'UTR primary sequence (Yang et al., 2019). Nevertheless, an accepted classification of eukaryotic IRESs distinguishes two types based on either the involvement of ITAFs in ribosome recruitment (Komar and Hatzoglou, 2011; Meyer et al., 2015) or the direct binding between the IRES and the 18S rRNA – similarly to what happens with the Shine-Dalgarno sequence in bacteria (Dresios et al., 2006) (**Fig. 3.2.4**).

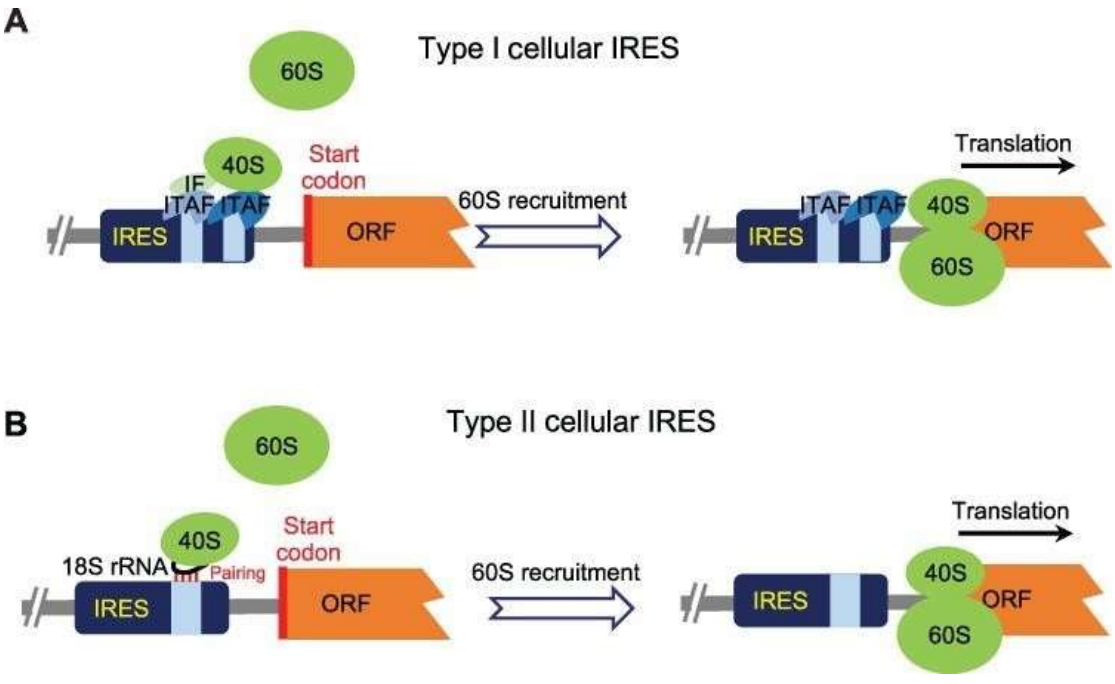


Fig. 3.2.4 IRES-mediated translation initiation mechanisms. *(A) Type I eukaryotic IRES: cis-acting elements in the IRES sequence are bound by ITAFs that in turn interact with the 40S ribosomal subunit. (B) Type II eukaryotic IRES: a short cis-element sequence inside the IRES*

base pair to the 18S rRNA, leading to the direct interaction between IRES and 40S ribosomal subunit. Figure taken from Yang et al., 2019.

However, the precise mechanisms by which eukaryotic IRESs and ITAFs work in translation is still under debate (Kozak, 2005; Kozak, 2007; Gilbert et al., 2010; Jackson et al., 2013). Some ITAFs act as RNA chaperones that remodel the mRNA secondary structure in the proximity of an IRES, allowing the ribosome recruitment (Mitchell et al., 2003; Pickering et al., 2004). Alternatively, ITAFs could function as adaptors for the interaction between the IRES motif and the eIFs (Stoneley et al., 2004; King et al., 2010).. More than fifty ITAFs have been reported to regulate translation, and a long non-coding RNA (lcnRNA), the TP53-regulated modulator of p27 (TRMP), has been shown to act as an ITAF, suggesting the role of other non-coding RNAs in this mechanism (Yang et al., 2018; Godet et al., 2019). The efficiency of eukaryotic IRESs is strictly dependent to the quantity and the quality of ITAFs: cellular-specific ITAFs can differentially modulate IRES-mediated translation in development and differentiation (Yang et al., 2019). In addition to ITAFs, eIFs are also involved as regulators of IRES-dependent translation, as demonstrated for eIF2, eIF3, and eIF5, acting in the stabilizing the ribosome recruitment on the IRES (Pestova et al., 1996). Moreover, "land and scan" IRESs, such as the one contained in the 5'UTR of myc and Apaf-1, require the unwinding activity of eIF4A and the enhancing activity of eIF4B (Spriggs et al., 2009).

To date, IRESs were found in the 5'UTRs of transcription factors involved in development, such as the homeotic gene antennapedia (Oh et al., 1992; Xue et al., 2015), proto-oncogenes such as c-Myc (Nanbru et al., 1997; Stoneley et al., 1998; Stonely et al., 2000), angiogenic growth factors such as fibroblast growth factors (FGFs) and vascular endothelial factors (VEGFs) (Vagner et al., 1995; Huez et al., 1998; Prats et al., 2002; Audigier et al., 2008; Morfoisse et al., 2016), underlining the role of IRESs in translational control of transcripts that need a tight regulation of gene expression. The switch between cap-dependent and IRES-mediated initiation of most of these transcripts occurs following a trigger, as reported for VEGFA, VEGFC, p53, where the condition of cellular hypoxia leads to the overexpression of eIF4G and 4E-PBs, sequestering the cap-binding protein eIF4E, and, thus, favoring the cap-independent initiation (Bellodi et al., 2010; Morfoisse et al., 2014). The IRES-mediated initiation can modify the quantity of the protein synthesized, but also produce protein isoforms

differing in the N-terminus, similarly to what has been described for uAUGs/uORFs. For example, the alternative IRES-mediated initiation of the FGF2 transcripts produces up to five different FGF2 proteins, differing in function and localization (Godet et al., 2019); likewise, the VEGFA transcript contains two IRESs that produce two isoforms with different intracellular localizations (Huez et al., 2001). 5'UTR IRESs are involved in the translation of a subset of transcripts regulating apoptosis, such as XIAP (Holcik), Apaf-1 (Coldwell et al., 2000), p53 (Yang et al., 2006) and Bag1 (Dobbyn et al., 2008). Since XIAP and Apaf-1 have opposite functions in apoptosis, their relative abundance, regulated by distinct IRESs, represents an essential point of control of the process. IRESs are also involved in the development of the synaptic network in brain, in spermatogenesis – where the IRES-mediated initiation of FGF2 has a crucial role (Audigier et al., 2008; Gonzalez-Herrera et al., 2006) – and in the cellular differentiation and tissue regeneration – through the translational control of FGF1 in muscle (Conte et al., 2009).

Due to the importance of the cellular processes regulated by IRES-mediated translation, similarly to what happens with uAUGs/uORFs and inhibitory secondary structures, mutations in IRESs can lead to the onset of diseases. For example, a SNP in the c-Myc IRES has been found to cause the overexpression of c-Myc in multiple myeloma (Chappell et al., 2000; Shi et al., 2016). SNPs in the connexin 32 and VEGFA IRESs have been linked to the development of two severe neurodegenerative disorders - respectively the Charcot-Marie-Tooth disease and ALS (Lambrechts et al., 2003; Péladeau et al., 2021).

3.2.5 Emerging therapeutics strategies involving drugs and antisense oligonucleotides targeting the 5'UTR

These findings, regarding the impact of mutation in 5'UTR cis-acting regulatory elements, increase the attention on transcripts leader sequences in the diagnostic processes of human pathologies and promote the study of new therapeutic strategies. The idea of specific 5'UTR as therapeutic targets is innovative, because it might avoid side effects of drugs acting on general cellular processes. Moreover, knowing that dysregulated translation of a transcript leads to cancerogenesis, suggests a therapeutic approach targeting eukaryotic initiation factors with new chemotherapeutics (Lindqvist and Pelletier 2009). Among these factors, one of the top target candidates is the DEAD-Box helicase eIF4A, given its role in modulating the translation efficiency of numerous

oncogenes and tumor suppressors (Cunningham et al., 2018). The use of eIF4A specific inhibitors, such as silvestrol, in treating sarcomas and other malignant cancers, is still considered an option, even if it has limitations in terms of specificity, due to the wide range of transcripts regulated by eIF4A (Chang et al., 2020). The action of silvestrol on eIF4A is also taken into consideration in viral infection, due to its broad-spectrum influence on translation that interests not uniquely the cancer context, but potentially every context in which transcripts regulated by the eIF4A unwinding activity are involved (Muller et al., 2018), although redundancies in eIFs action could potentially lead to less than expected effects (Parsyan et al., 2011). For this reason, in order to increase specificity, it could be convenient to consider a new strategy that implies a direct action on 5'UTR elements. Recently, the use of antisense oligonucleotides (ASOs) in correcting the protein expression levels, targeting the 5'UTR cis-acting regulators, emerged and could be potentially applied to several diseases. ASOs are synthetic antisense oligonucleotides (15-20 nts), chemically modified to gain resistance to endogenous nucleases and better patient compliance (Bennett et al., 2019). Their action is performed by binding to a specific region of the target mRNA using WC base pairing. Their use has already been approved by the Food and Drug Administration (FDA) and the European Medicines Agency (EMA) for clinical application in spinal muscular atrophy (SMA), hereditary transthyretin amyloidosis and familiar hypercholesterolemia (Benson et al., 2011; Crooke et al., 2017; Bennett et al., 2019). In the context of translation, ASOs action is based on the specific binding to cis-acting elements on the 5'UTR, resulting generally in the inhibition of their functions. For example, an ASO can prevent the translation initiation at uAUGs, promoting the translation of the main ORF, or it can avoid the inhibitory effect of secondary structures such as hairpins and G4s (Rouleau et al., 2015; Liang et al., 2016; Liang et al., 2017). An interesting case in the employment of ASOs is their use to increase the protein level of CFTR as new therapeutic strategy to treat Cystic fibrosis (CF). CF is an autosomal monogenic disease caused by aberrant CFTR protein production, folding, trafficking or activity (Mall et al., 2014). One of the strategies to treat CF is the increase of CFTR protein level and promising results have been obtained with the demonstration that the CFTR 5'UTR can be targeted by ASOs, impeding cis-acting inhibitory elements on the sequence and producing a significative increase in CFTR translational efficiency (Sasaki et al., 2019).

Recently, the utilization of ASOs targeting the 5'UTR has been also considered for the

treatment of SMA, a recessive neuromuscular disorder caused by the loss-of-function of the SMN1 protein (Lefebvre et al., 1995). All SMA patients have at least one copy of a SMN1 paralog, SMN2, which resulted from a gene duplication event at the SMN locus on chromosome 5 (Calucho et al., 2018). SMN2 does not fully compensate for the loss of SMN1, because the majority of the SMN2 transcripts does not retain the exon 7 (Monani et al.1999). As consequence, only 10% of the SMN2 transcripts encodes for a functional SMN protein. Therefore, the primary therapeutic strategy for SMA consists in increasing the expression level of SMN, targeting the splicing of SMN2 with ASOs that enhance the SMN2 exon 7 inclusion (Hua et al., 2007). The increase of SMN2 containing exon 7 can be even combined with the simultaneous upregulation of SMN2 translation efficiency through ASOs acting on the 5'UTR (Winkelsas et al., 2021). In fact, previous work on the SMN2 5'UTR found structured cis-acting regulatory elements and a functional uORF that can be targeted by ASOs to increase the rate of translation of SMN2 transcripts containing the exon 7, as confirmed by the consequent increase in the expression of Gemin6 and Gemin8 that are related to the levels of a SMN functional protein (Winkelsas et al., 2021).

The precise mechanism of action by which 5'UTR ASOs modulate translation is still not completely elucidated, nevertheless their specificity is an undeniable advantage of their use as a treatment. Required conditions to consider the use of 5'UTR ASOs as a treatment are a deep knowledge of the etiology of the disease of interest and the characterization of the 5'UTRs of the causative genes. Monogenic diseases, in which the pathology is caused by a shortage in the functional protein are the best candidates for this type of approach, similarly to what has been described for CF.

The monogenic CDKL5 Deficiency Disorder (CDD), caused by the loss-of-function of the kinase Cyclin-dependent kinase-like 5 (CDKL5), is a severe neurodevelopment disorder currently without a cure. We believe that, considering the features of this disorder, the 5'UTR could represent an interesting therapeutic target. However, the lack of information about the CDKL5 5'UTR has been, until now, a limit to this approach.

3.3 Cyclin-dependent kinase-like 5 and CDKL5 Deficiency Disorder

3.3.1 Cyclin-dependent kinase-like 5

Cyclin-dependent kinase-like 5 (CDKL5) is a serine/threonine protein kinase expressed mainly in the brain (Lin et al., 2005). Mutations in the gene leading to dysfunctional protein variants have been associated to the onset of neurodevelopmentaldisorders such as X-linked West syndrome (Kalscheuer et al., 2003) and Atypical Rett Syndrome named Hanefeld variant (Pini et al., 2012; Chahil et al., 2021; Guerrini et al., 2021). Recently, these brain conditions sharing the involvement of a mutated CDKL5 gene have been grouped under the umbrella term of CDKL5 Deficiency Disorder (CDD)(Jakimiec et al., 2017).

CDKL5 has been first identified in 1998, when it was named STK9, standing for "serine-threonine kinase 9" (Montini et al., 1998). Its current name comes from the similarity between its kinase domain and the ones of kinases belonging to the proteins of the CDK family (Manning et al., 2002). The CDKL5 gene is located on the X chromosome (Xp22.13) and is 228-kb long, comprising 21 exons. An abundance of CDKL5 transcript variants has been reported, and at least five different protein isoforms, reported as hCDKL5_1, hCDKL5_2, hCDKL5_3, hCDKL5_4, hCDKL5_5 (Hector et al., 2016) could be translated (**Fig. 3.3.1.1 A**). All these isoforms are present in every tissue, although with different degrees of expression (Hector et al., 2016), except for testis, where the unique detected isoform was hCDKL_5, also known as "long isoform", because it is the longest one composed of 1030 amino acids (**Fig. 3.3.1.1 B**). The long isoform is one of the most expressed variants, and it is reported as one of the two Reference Sequences of CDKL5 on NCBI (National Centre for Biotechnology Information), where it is named as isoform 1 (NP_003150.1 and NP_001032420.1). The remaining four isoforms do not differ much between each other in terms of length, but among them, hCDKL5_1 (960 amino acids) is the most expressed, especially in the brain, and it was reported on NCBI as isoform 2 (NP_001310218.1) (Fichou et al., 2011; Williamson et al., 2012; Hector et al., 2016).

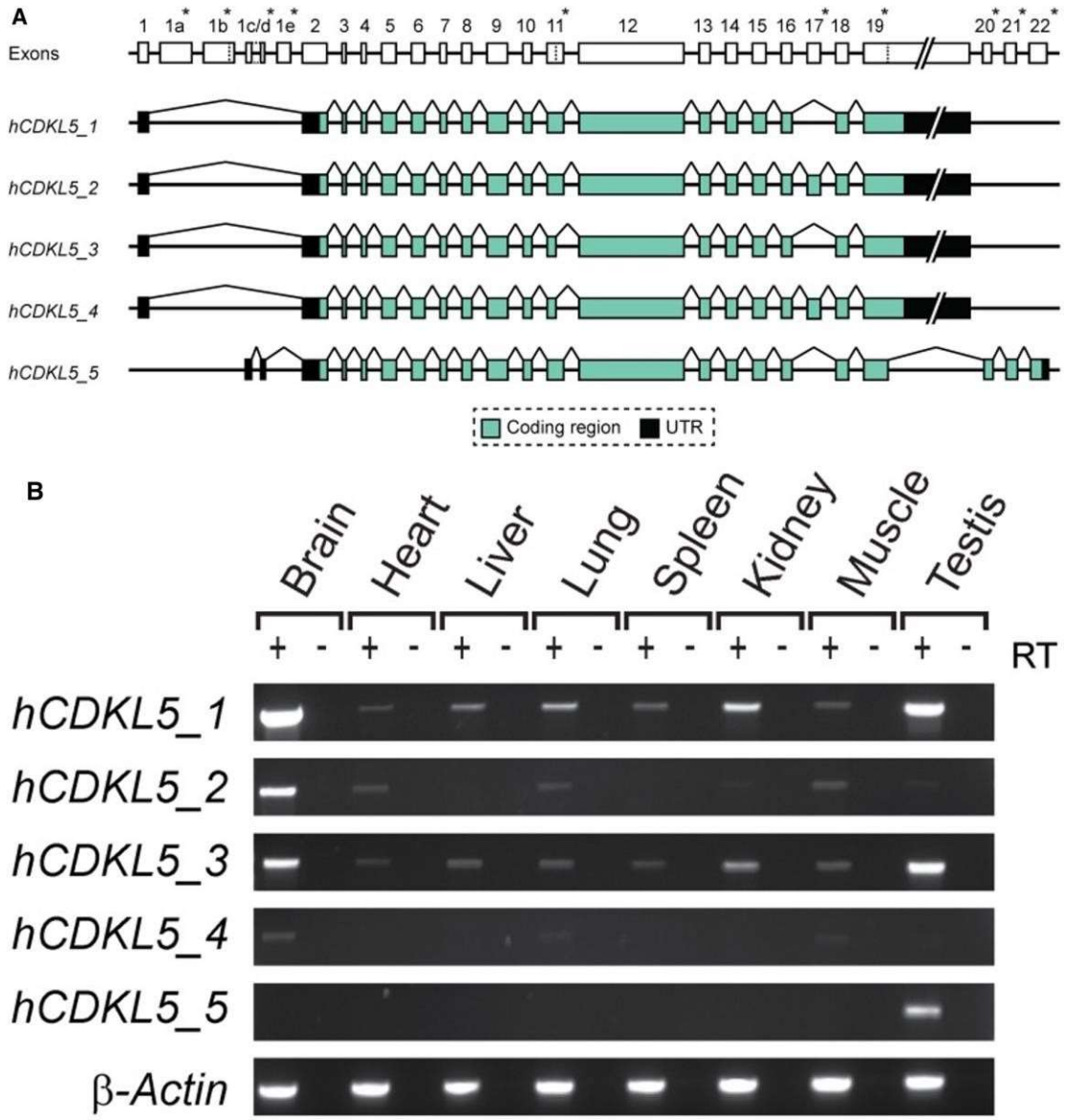


Fig. 3.3.1.1 CDKL5 Isoforms. *(A) Schematic representation of the structure of the human CDKL5 exon composition of the five different coding isoforms (hCDKL5_1 to HCDKL5_5). Black lines connecting exons indicate splicing events, while asterisk indicate where differences between the different TVs are found. Introns and 3'-UTR are not in proportion. (B) CDKL5 protein isoforms expression in human tissues. The analysis was performed through RT-PCR, using β-Actin as loading control. Figures are taken from Hector et al., 2016.*

Concerning the CDKL5 N-terminus, it includes the ATP binding site, the serine/threonine kinase active site (S/T) and the Thr-Glu-Tyr (TEY) motif – through which CDKL5 is self-regulating its phosphorylation function (Bertani et al., 2006). The C-terminus, instead, contains two nuclear localization sequences (NLS) and a nuclear export signal (NES) that regulates the intracellular localization of CDKL5 via a cytoplasmic-nuclear shuttling process (Jdila et al., 2018), allowing the kinase to operate

in both cellular compartments (**Fig. 3.3.1.2**).

All the reported isoforms share the same N-terminus, containing the catalytic domain (1-297 amino acids), indicating that CDKL5 has a unique start codon. However, most of the transcript variants of the gene differ at the 5'sequence of the mRNA, showing a variety in the 5'UTR.

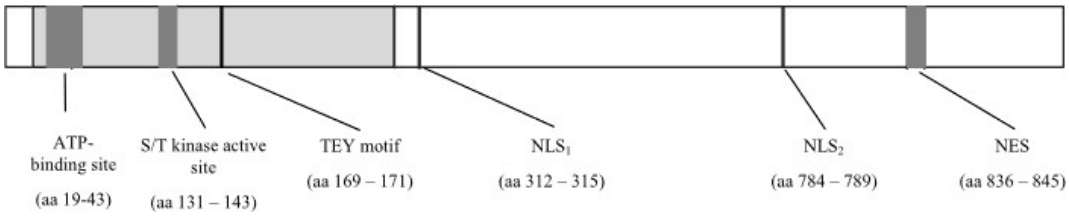


Fig. 3.3.1.2 Diagram of the CDKL5 protein. *In this representation the N-terminus and the C-terminus of the protein are schematically reported. N-terminus, representing the catalytic region, is shown in grey, whereas C-terminus region is colored in white. TEY motif Thr-Glu-Tyr motif, NLS1 and NLS2 nuclear localization signals 1 and 2, NES nuclear export signal. Modified from Fehr et al., 2015.*

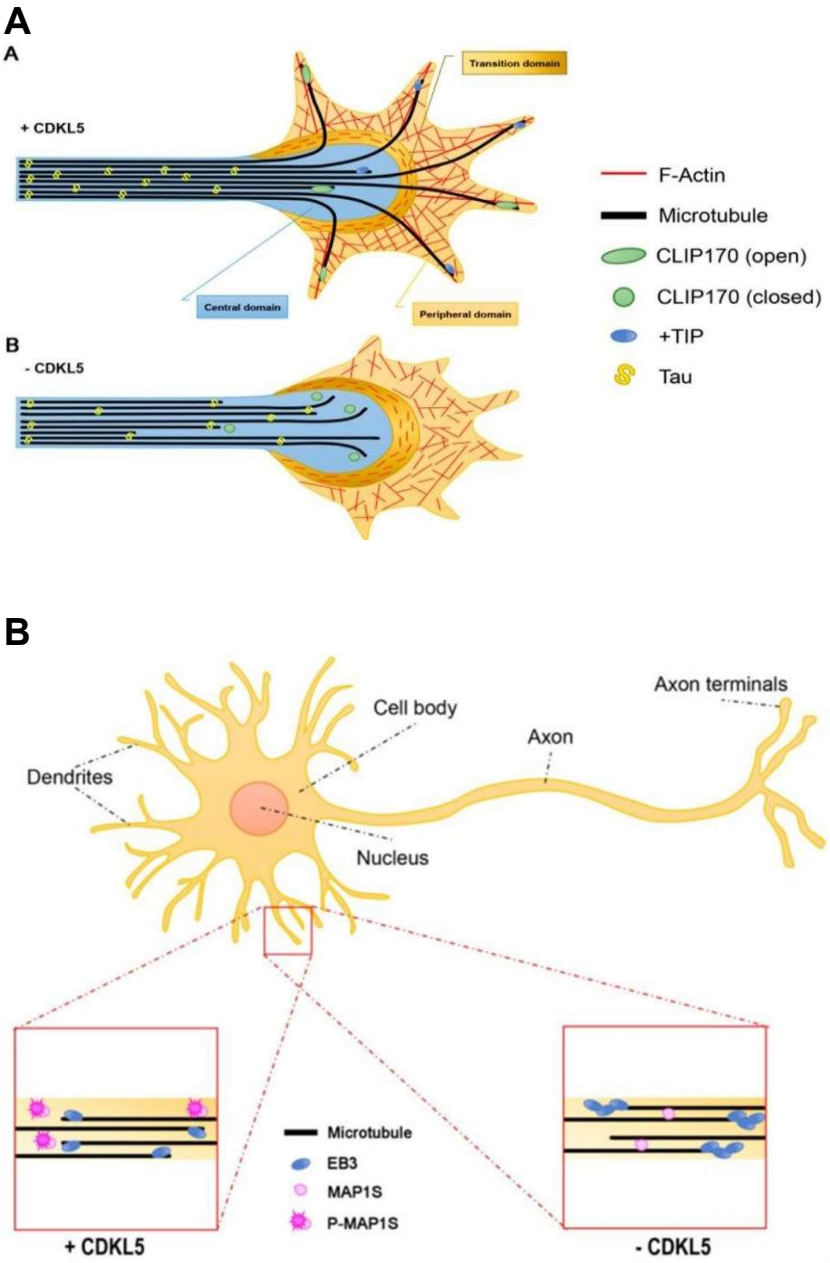
The CDKL5 localization within the cell is dynamic and strictly regulated since its amount in nucleus and cytoplasm compartments is tissues-specific and varies according to the developmental stage. For example, in the murine brain, CDKL5 is almost completely cytoplasmatic in the first post-natal days (E18-p5), while almost half is located in the nucleus as the animal grows (p14-p120) (Rusconi et al., 2008). Moreover, analysis of different samples from p120 mouse brains showed that the cytoplasm/nucleus ratio is almost 1 in the cortex, while only 20% is localized in the nucleus in the cerebellum (Rusconi et al., 2008). Therefore, CDKL5 expression regulation is tissue and stage-specific, leading to different quantities of the protein (Hector et al., 2016a; Hector et al., 2016b), as well as a differential distribution within the cell at different developmental stages (Rusconi et al., 2008).

The discovery of targets of CDKL5-mediated phosphorylation is a topic of interest because it can clarify the relationship between the lack of CDKL5 function and the development of CDD. Due to the overlapping symptoms, CDKL5 deficiency was initially considered a cause of Rett syndrome (RTT), usually linked to MeCP2 loss of function. This led to wonder whether MeCP2 could be a CDKL5 target. *In vitro* kinase assay initially supported this thesis (Mari et al., 2005), but subsequent publications

showed that, in cell cultures, CDKL5 dependent MeCP2 phosphorylation level is very low compared with other established CDKL5 targets (Sekiguchi et al., 2013). The possibility that MeCP2 is not a direct target of CDKL5, but, rather, modulated by a CDKL5 target has been put forward (Lin et al., 2005). This intermediate could be the DNA methyltransferase 1 (Dnmt1), a protein that is phosphorylated by CDKL5, and, as that is also known to interact with MeCP2 to maintain the right degree of DNA-methylation (Kimura et al., 2003; Kameshita et al., 2008). Based on the current findings, CDKL5 could impact the DNA-binding efficacy of Dnmt1, acting as an epigenetic transcriptional regulator inside the nucleus, like casein kinase 1 (CK1) is doing. A phosphoproteomic screening identified other nuclear targets of CDKL5 (Khanam et al., 2021), including: the E1A-binding protein p400 (EP400); the chromatin-remodeling transcriptional activator (Pradhan et al., 2016); Elongin A (ELOA), a transcriptional elongation factor (Wang et al., 2021); the YLP motif-containing protein 1 (YLPM1, also known as ZAP3) (Chemudupati et al., 2019). The identification of these substrates defines new possible CDKL5 nuclear function in the context of the DNA damage response (DDR), where the kinase senses anomalies and, in response, triggers the transcriptional silencing of regions adjacent to the DNA breaks (Khanam et al., 2021). Among the nuclear targets, also SMAD3, a transcriptional factor involved in the neural functions regulated by TGFβ (transforming growth factor-β) signaling, has been identified. The action of CDKL5 on SMAD3 is likely to modulate its expression level (Fuchs et al., 2019), and, furthermore, to regulate a SMAD3-mediated and CDKL5-dependent apoptosis, in line with data showing how the loss of CDKL5 function results in neuronal apoptosis (Fuchs et al., 2014). In this context also histone deacetylase 4 (HDAC4) has been identified as CDKL5 direct target. HDAC4 is a transcriptional regulator involved in neuronal survival (Trazzi et al., 2016), and the loss of CDKL5 changes the HDAC4 localization, disturbing its physiological function. In fact, HDAC4 localization is regulated by phosphorylation (Li et al., 2012) and the CDKL5 absence is expected to lead to abnormal HDCA4 nuclear accumulation, triggering cellular apoptosis (Fuchs et al., 2014).

On the other hand, most of the putative cytoplasmic CDKL5 targets are involved in the organization of microtubules. For example, MAP1S (microtubule-associated protein 1S) is a stability regulator of microtubules during the cell cycle, and its binding efficiency is modulated by CDKL5-mediated phosphorylation (Bolger et al., 2005, Tegha-Dunghu et al., 2014). In addition to MAP1S, other proteins involved in the

biology of microtubules identified as potential CDKL5 substrates are: the 131-kDa centrosomal protein (CEP131) (Muñoz et al., 2018); the disc large membrane-associated guanylate kinases scaffold protein 5 (DLG5) (Muñoz et al., 2018); the Rho/Rac guanine nucleotide exchange factor 2 (ARHGEF); the microtubule-associated protein RP/EB family member 2 (EB2) (Tegha-Dunghu et al., 2014); and CLIP170 (Barbiero et al., 2017). The importance of cytoplasmic CDKL5 in neurons is sustained by the fact that its loss leads to dysfunction in the synaptic network and abnormal neural morphology, coherently with its proposed role in regulating microtubules in both axon and dendrites (Amendola et al., 2014; Barbiero et al., 2019) (**Fig. 3.3.1.3**).



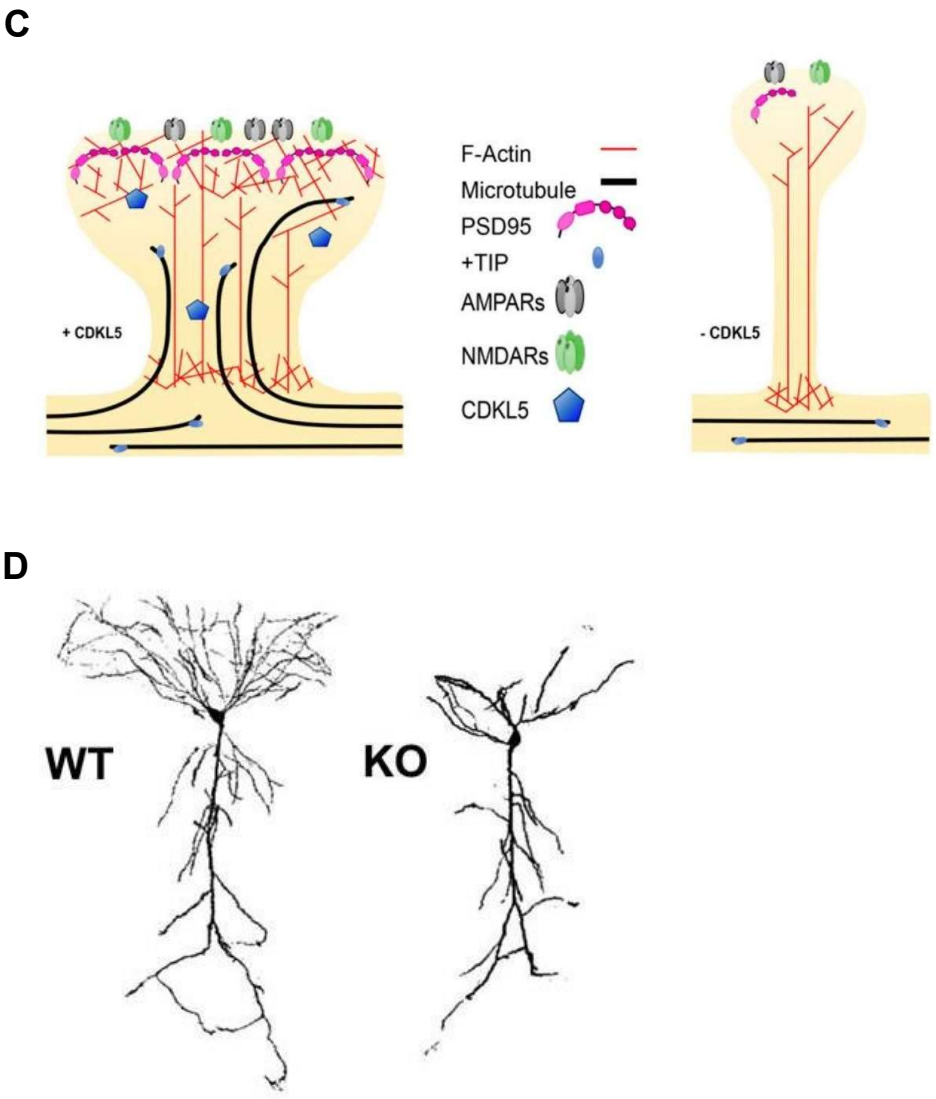


Fig. 3.3.1 The Effects of CDKL5 Deficiency on neuronal cytoskeleton. *(A) In the context of axonal growth cones, the loss of CDKL5 leads to cessation of axonal outgrowth. In healthy axons, microtubules are well organized in a polar manner, with well-spaced plus-ends. The axonal central domain contains numerous microtubules bound with plus-end tracking proteins (+TIPs), including CLIP170, whereas the peripheral domain is rich in actin extending in filopodia. The transition area between these two domains contains actin arcs impeding the complete invasion of microtubules in the peripheral region. In shortage of CDKL5, CLIP170 and other TIPs are less associated with microtubules, which are less numerous than in +CDKL5 condition. This lack of +TIPs impacts the microtubules capacity to extend in the actin rich region, resulting in an abnormal axonal morphology. (B) In dendrites, CDKL5 phosphorylates MAPS1, leading to the inhibition to its binding to microtubules. MAPS1 binding is in turn associated with EB3 binding to the microtubules plus-ends. The loss of CDKL5 reduces these processes, leading to microtubule instability and a global impairment of dendritic arborization. (C) CDKL5 regulates dendritic spine morphology. In physiological condition, actin enriched spines are rich in microtubules upon neuronal activation. +TIPs bind*

microtubules plus-ends, facilitating microtubules dynamism. In CDKL5 absence, microtubules are less dynamic and fail in invading spines, strongly reducing the activity-dependent structural changes, and leading to the loss of mushroom-shaped spines. CDKl5 influences directly also various proteins involved in excitatory neurotransmission, such as PSD95 and Glu2A subunit of AMPA receptors. Therefore, its loss also impairs the availability of these types of proteins. (D) Comparison between the representative images of neurons from adult wild type (WT) and CDKL5-KO (KO) mice, in which it is appreciable the difference in morphologies. Panels A, B, C of this figure are from Barbiero et al., 2019; panel D is modified from Amendola et al., 2014.

Another potential target of CDKL5 is netrin-G1 ligand protein (NGL-1, also known as LRRC4C) (Ricciardi et al., 2012), a postsynaptic membrane protein that binds to the presynaptic netrin-G1 protein (also known as NTNG1), helping to shape the formation of a correct neural network (Nishimura-Akiyoshi et al., 2007). Interestingly, NTG1 is also a causative gene of atypical RTT, strengthening the hypothesis that neuronal anomalies occurring in RTT could be linked to abnormal CDKL5 regulation on downstream targets (Borg et al., 2005; Archer et al., 2006; Nectoux et al., 2007). Another interesting CDKL5 target could be amphiphysin-1 (Amph1) (Senga et al., 2011), a protein involved in clathrin-mediated endocytosis by binding to endophilin (Nogueras-Ortiz et al 2014). By phosphorylation, CDKL5 prevents the Amph1 ability to bind its partner, inhibiting endocytosis (Senga et al., 2011). Endocytosis plays a crucial role in the neural context, and particularly in neurodevelopment, as it acts on mechanisms such as synaptic vesicle recycling, spine formation and axon growth (Heuser et al., 1973). The description of all these pathways involving CDKL5 further indicates how important is the role of CDKL5 in neurons (Katayama et al., 2020). Therefore, it is no surprise that serious cellular dysfunctions take place when CDKL5 loses its function because of mutations or altered expression, resulting in the onset of a severe neurodevelopmental condition such as CDD.

3.3.2 CDKL5 Deficiency Disorder

CDD is an X-linked neurodevelopmental disorder caused by the deficiency of functional CDKL5 in the context of brain development. Most patients are female, with a ratio of 4:1 compared to males, who experience more severe symptoms and intrauterine lethality due to the localization of the gene on the X chromosome. In

women, instead, the X chromosome inactivation (XCI) and the somatic mosaicism can alleviate the symptoms (Liang et al., 2019; Liang et al., 2020). CDD is usually caused by *de novo* mutations that occur spontaneously in the population. It has been calculated that approximately 1 child out of 40,000-60,000 carries pathogenic mutations in the CDKL5 gene (Lindy et al., 2018; Kothur et al., 2018). Currently there are more than 265 known pathogenic variants of the CDKL5 gene, the majority of which (50%) are point mutations, while missense mutations represent roughly 38% of the mutations found in CDD patients. However, only approximately 27% are confirmed to be pathogenic, and mainly involves the catalytic domain of the protein, creating a loss of function of CDKL5 by disrupting the phosphorylation activity (Bahi-Buisson et al., 2011; Krishnaraj et al., 2017; Liang et al., 2019; Olson et al., 2019). Frameshift mutations, which involve the insertion or deletion of a single nucleotide, are - as one might expect - the mutations with the most serious pathological consequences, affecting approximately 13% of patients (Bahi-Buisson et al., 2011; Hagebeuk et al., 2015; Jakimiec et al., 2020). Splicing mutations resulting in exon skipping (Krishnaraj et al., 2017) and nonsense mutations often result in truncated isoforms unable to have a correct intracellular localization and, thus, to correctly fulfill the role of CDKL5 in the right compartment (Nectoux et al., 2006; Rusconi et al., 2008; Fazzari et al., 2019). Additionally, nonsense mutations can activate the nonsense-mediated decay (NMD) of the transcript reducing CDKL5 expression levels (Yennawar et al., 2019).

The symptoms of CDD vary based on the type of mutation and the region of the gene involved. The main ones are early onset refractory epilepsy, delayed development and gross motor impairment, serious cognitive disability, autistic-like features, cortical vision impairment and hypotonia (Olson et al., 2019).

Currently, there is no cure for this neurodevelopmental disorder. The different constitutive Cdkl5 knockout (KO) murine models provide a useful tool to test new drugs, and, moreover, they allowed a direct observation of the consequences of the loss of CDKL5 in the context of brain development (Wang et al., 2012; Amendola et al., 2014; Okuda et al., 2017). Nevertheless, it should be noted that the mouse phenotype differs from the one observed in patients as, for example, does not reproduce the drug-resistant spontaneous seizures that are one of the diagnostic features of CDD. The symptoms observed in mice, such as hindlimb clasping, hypoactivity, impaired learning and memory, and visual attention/acuity deficits have been associated to the abnormal development of the neural morphology and the impaired synaptic communication

(Amendola et al., 2014; Fuchs et al., 2014). The neurons of the CDKL5-KO mice display lower branching and spine density, developing abnormal dendritic arborization and altered the number of neural connections (Wang et al., 2012; Amendola et al., 2014; Okuda et al., 2017).

Like CDKL5-KO mice, cortical neurons in culture in which endogenous CDKL5 is silenced through small interfering RNA (siRNA) show similar morphological defects. As expected, this negative impact on the neural morphology can be partly rescued through the expression of exogenous CDKL5, suggesting that the lack of the correct amount of CDKL5 might be, by itself, causative of CDD or, at least, of the abnormal neural development of CDD (Ricciardi et al., 2012). Therefore, it could be plausible that also dysfunctions in the pathways leading to CDKL5 protein expression can produce the morphological abnormalities and the clinical symptoms. In support to this hypothesis, it has been reported that MeCP2 could itself epigenetically regulate the expression of CDKL5 (Carouge et al., 2010). Therefore, deepening the knowledge about the regulation of CDKL5 protein expression at multiple levels may be not only interesting *per se* but also required to envisage possible therapeutic strategies.

3.3.3 Regulation of CDKL5 expression

Despite there is a good number of publications exploring the downstream pathways regulated by CDKL5, the identification of the regulatory processes that modulate CDKL5 expression is an emerging field of study concerning CDD. Investigation of these mechanisms focuses mainly to two classes of processes: transcriptional, which modulates the amount of mRNA transcribed from the gene; and post-transcriptional, which includes translational control (**par. 3.1.2**). Regarding transcription, there are at least two known mechanisms regulating CDKL5. First, MycN inhibits CDKL5 transcription, in line with the competing role between MycN, which enhances cell proliferation during brain development, and CDKL5, which triggers neuronal differentiation (Valli and et., 2012). Secondly, MeCP2 regulates CDKL5 gene expression inhibiting its transcription by binding with two 5' flanking regions of the gene. In fact, the MeCP2 transcriptional inhibition consists in its binding to two CpG islands (between 893 bp upstream and 670 bp downstream to the transcriptional start site) that are subsequently methylated. The effect has been observed in rats, after the overexpression of MeCP2, as well as after treatment with cocaine or serotonin, two drugs known to have an impact on these two proteins and on long-term brain plasticity

(Carouge et al., 2010). There is also the possibility that MeCP2 and MycN modulate CDKL5 transcription through a coordinated action. In fact, these proteins are often co-localized in the nucleus, close to the active transcription sites, and affect the transcription of other important genes that are involved in neural development, such as the brain-derived neurotrophic factor (BDNF) (Murphy et al 2011).

On the other hand, the translational regulation of CDKL5 is a scarcely explored field, even though there is at least one publication that anticipated the interest that is likely to emerge about this topic. In 2015 La Montanara and colleagues proposed a novel paradigm for an activity-dependent neuronal control of CDKL5 expression (La Montanara et al., 2015). This research explored an interesting aspect of the regulation of CDKL5 expression in neurons. In fact, it demonstrated that CDKL5 protein level rapidly increases following the stimulation of hippocampal neurons in culture (DIV17). This effect has been attributed to translation by pretreating the neurons with the protein synthesis inhibitor cycloheximide (40 μM, 30 min). This work is currently the only one that investigated the translational control of CDKL5 in neurons in a stimulus-dependent manner. Since CDKL5 is translated locally in the dendrites (La Montanara et al., 2015), this regulation possibly links it in the development of the synaptic network and synaptic plasticity.

3.3.4 Pathogenic variants in the promoter and the 5'UTR of CDKL5

Even if almost all the mutations found in CDD patients do not lie in non-coding regions of the CDKL5 gene, it is important to start to consider the mutations in the 5'UTR and the promoter as potentially linked with an unbalance of CDKL5 expression, and thus, causative of the pathology. Therefore, a better understanding of the roles of these mutations in CDKL5 expression would improve our knowledge about CDD etiopathology. Most of these mutations lead to the total loss of the protein, since they are gene deletions causing the complete loss of the TSS and thus the lack of the transcript (Bahi-Buisson et al., 2010). Similarly, the loss of the exon 2 caused by exon skipping or deletion, results in the absence of the start codon and, therefore, of the protein. (Córdova-Fletes et al., 2010; Bartnik et al., 2011; Van Esch et al., 2007; Nemos et al., 2009). However, in 2005, the screening of the CDKL5 gene in 94 patients suffering with RTT or RTT-like pathological phenotypes - previously tested negative for MeCP2 mutations – discovered, in the non-coding regions, the following two mutations that were not present in the 69 control patients:

- A mutation in the promoter of a male patient, c.-440G>T (transversion).
- A mutation in first untranslated exon of CDKL5 TV NM_003159.3 of a femalepatient suffering with Atypical RTT, c.-189C>T (transition).

The meaning of these mutations is still not clear and there are no publications exploring their role in CDKL5 regulation (Evans et al., 2005). However, these SNPs are reported on ClinVar portal and on dbSNP database, with rs777401314 and rs786204994 identification codes respectively. The frequency of the rs777401314 (promoter SNP) was calculated to be 0.00039% in the Genome Aggregation Database (GnomAD, Koch et al., 2020) and 0.00053% in the TOPMED database (Burgess et al., 2021), suggesting the benignant nature of the SNP. On the other hand, rs786204994 (5'UTR SNP) does not appear in any of these databases, being identified uniquely in the patient presented in the Evans' study. This indicates that it could be a malignant SNP responsible for the onset of CDD, probably altering a regulatory 5'UTR cis-acting element. Therefore, investigating the composition and the functionality of the CDKL5 5'UTR can aid to clarify this hypothesis, leading to a better comprehension of CDKL5 regulation and paving the way to new therapeutic strategies, such as ASOs, in CDD treatment.

4 AIM OF THE WORK

Despite CDKL5 being a topic of great interest, the study of the regulation of its expression is poorly described in the scientific literature. CDKL5 is involved in key processes in neuronal cells, as witnessed by the consequences of its deficiency in CDD, and its expression is expected to be under strict regulation, in agreement with the complex morphology and activity of the neurons. Indeed, it has been demonstrated how insufficient levels of CDKL5 can cause morphologic alterations in neurons in vitro, similar to the ones typically observed in neurons of CDKL5-KO mice. If little is known about transcriptional regulation of CDKL5, even less has been investigated in relation to its translational regulation. In fact, only one publication (La Montanara et al., 2015) has reported that CDKL5, upon NMDA receptor stimulation, is subject to a rapid, activity-dependent upregulation, compatible with a translational effect. The transcript leader, or 5'UTR, of the mRNA has been for a long time overlooked as an element of regulation. CDKL5 has a long, complex and extremely conserved 5'UTR, but nothing is known about the mechanisms of a possible translational regulation. For this reason, my work was aimed at:

- collecting evidence of the presence of a translational control mechanism in modulating CDKL5 expression level;
- analyzing in detail the sequence of the CDKL5 transcript leader, considering the number of first alternative exons in 5'UTR variants, by using bioinformatics tools and experiments with reporter genes;
- providing experimental evidence of a possible pathogenic role of SNPs in the 5'UTR of CDKL5, with particular attention to patients with a diagnosis of CDD.

We considered essential the elucidation of these points because it would allow the expansion of the knowledge of translational control of CDKL5, especially in the context of neuronal development and synaptic network homeostasis. Moreover, this would provide a key step in the development of new therapeutic strategies for CDD, based on the manipulation of translational efficiency.

5 RESULTS

5.1 CDKL5 Expression pattern analysis

5.1.1 CDKL5 Protein and mRNA levels in mouse tissues

Since the levels of transcript are not always proportional to the amount of the produced protein, the comparison of the levels of protein and RNA from a gene is not a superfluous measurement (Wang et al., 2013). In particular, these measurements enable us to identify possible mechanisms of post-transcriptional regulation. For this reason, we collected tissues by dissecting three wildtype p30 mice and performed mRNA and protein quantification through Real-Time qRT-PCR and WB. The selected mice were all littermate males, to minimize discrepancies in protein expression levels due to differences in sex or environmental conditions. Three cerebral tissues were selected for the analysis - cortex, hippocampus, cerebellum - together with heart, lungs, liver, spleen, and kidneys. We performed WB to quantify CDKL5 protein levels. The signal for CDKL5 obtained from each tissue was normalized on protein loading through ponceau staining and compared against lung. We averaged the measurements obtained from three individual experiments - namely, in tissues coming from three distinct animals - using SEM as a dispersion index (**Fig. 5.1.1.1**). The result from this analysis showed, in accordance with the literature, that. CDKL5 protein levels are high in cerebral tissues, with the highest levels recorded in samples obtained from the cortex and hippocampus.

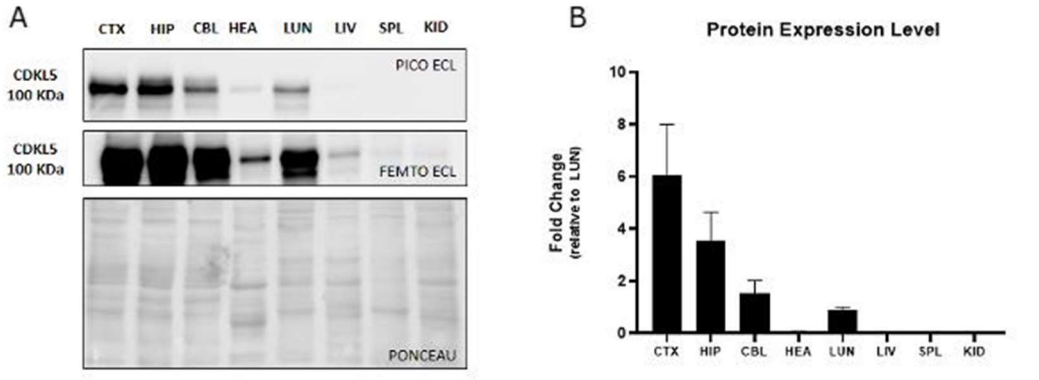


Fig. 5.1.1.1 CDKL5 protein levels in various p30 mouse tissues. *(A) Endogenous CDKL5 protein levels measured in cortex (CTX), hippocampus (HIP), cerebellum (CBL), heart (HEA), lung (LUN), liver (LIV), spleen (SPL) and kidney (KID). High expression levels were detected in brain tissues, while the signal of the antibody was weak in other tissues, needing as revealed by the more sensitive Femto ECL. Ponceau was used for total protein normalization of CDKL5*

signal. (B) Three distinct experiments were performed and CDKL5 signals were presented as fold change of the protein level, using lung as reference tissue. Obtained data was plotted as mean; error bars represent ±SEM.

Appreciable amounts of the protein have also been found in the lungs, even though it is still unclear what the meaning is for such a high CDKL5 expression level in this tissue compared to the others. However, the result is noteworthy since some patients affected by CDD show respiratory disorders (Hagebeuk et al., 2013) and recent work on CDKL5-KO mice also show how the loss of CDKL5 disrupts respiratory function in mice (Lee at al., 2017). Such phenomena have been interpreted as a failure on the part of the respiratory centers of the brainstem, but the higher CDKL5 protein level in the lung could be linked to a still unknown function in this tissue, and thus give a better explanation for the respiratory symptoms. CDKL5 protein levels in heart, liver, spleen, and kidney tissues all resulted barely detectable, with protein bands obtained only by using the highly sensitive "Femto" kit.

In parallel, we extracted RNA from each tissue and proceeded to prepare samples for Real-Time qRT-PCR, to quantify CDKL5 mRNA levels. RPL13 and GAPDH were selected as reference genes and were used to normalize the CDKL5 signal detected in each sample. However, due to variations of uncertain significance related to the use of GAPDH (Kozera et al., 2015), only RPL13 normalization was presented as a preliminary evaluation. mRNA expression levels of CDKL5 obtained in various tissues were normalized against lung and reported as fold changes. Data obtained from three different experiments were reported as mean, using the SEM index as measurement of statistical dispersion of data(**Fig. 5.1.1.2**).

In this experiment we evaluated the CDKL5 levels in male mice, since they were considered a simpler model since they lack the more complex mechanisms regulating CDKL5 expression in females (i.e X-inactivation and mosaicism). Given the results obtained, it will be worth to perform the same evaluation in females, also to verify a possible difference.

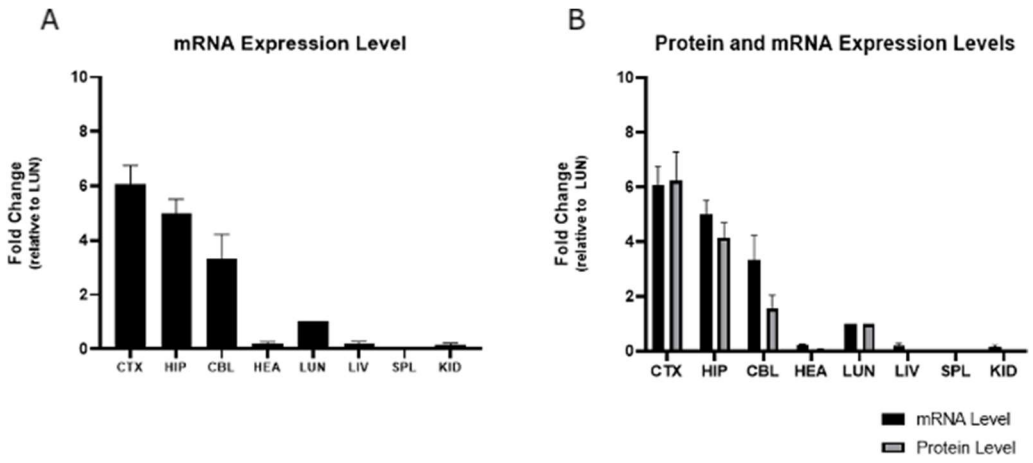


Fig. 5.1.1.2 CDKL5 mRNA level in various p30 mouse tissues. *(A) CDKL5 mRNA expression levels measured in cortex (CTX), hippocampus (HIP), cerebellum (CBL), heart (HEA), lung (LUN), liver (LIV), spleen (SPL) and kidney (KID), using Real-Time qRT-PCR. Higher expression of the transcript was detected in brain tissues. CDKL5 measurement is normalized on RPL13, used as reference gene. mRNA levels are presented as fold change using lung as reference tissue. Data obtained from three distinct experiment was represented as mean; error bars represent ± SEM. (B) Histogram that compares the results obtained from the quantification of CDKL5 protein level and CDKL5 mRNA expression in all the tissues taken into consideration, presented as mean. Error bars represent ± SEM.*

As expected, the higher levels of transcript have been detected in cerebral samples. More specifically, we identified the cortex as the tissue having the highest levels of mRNA, followed by hippocampus, cerebellum, and lung. This follows the same trend observed from the WB in the case of proteins. Conversely, the tissues that showed the lowest levels of CDKL5 protein (heart, liver, spleen, and kidney), exhibited extremely low levels of transcript as well. The comparison between the quantification of protein and mRNA levels of CDKL5, made with the reported procedures, do not show an appreciable discrepancy, highlighting that the expression of CDKL5 within the tissue is mainly regulated at a transcriptional level. However, we still cannot exclude the possibility of a post-transcriptional spatial or temporal regulation of the protein levels based on the current analysis.

5.1.2 CDKL5 Protein and mRNA levels in various postnatal developmental stages

We conducted the same CDKL5 expression analysis for mRNA and protein levels on

cortex obtained from wild-type mice at different postnatal developmental stages. The postnatal days taken into consideration were p2, p7, p14, p30, p60 and p120.

We analyzed the cortices of three animals for each reported postnatal day, using one cortex to quantify CDKL5 protein level through WB and the other to measure mRNA levels through Real-Time qRT-PCR.

The post-natal day chosen as reference was p2. The quantification of CDKL5 mRNA levels obtained at the selected postnatal days is expressed as fold change, obtained from the comparison with the mRNA level detected in p2. The data collected from three separate experiments, thus considering three different samples at the same postnatal stages, are presented as mean, using SEM index to measure the dispersion of data.

From the analysis of the protein, it emerges that CDKL5 level in the cortex during the animal development, increases up until the 30th day of life. At this point, the expression level of CDKL5 reaches its maximum. The following stages analyzed, p60 and p120, show a protein abundance that remains stably high but lower than that observed at p20. This pattern follows the dendritic arbors development in mouse cortex (Richards et al., 2020), in accordance with the function of CDKL5 in regulating the experience-dependent growth of synapses (Chen at al., 2010; Zhu et al., 2019). In fact, dendritic features such as the degree of arborization, the segment lengths and the segment number, peak in between p16-p30 during the development of the mouse brain, correlating with the increase recorded in CDKL5 protein levels.

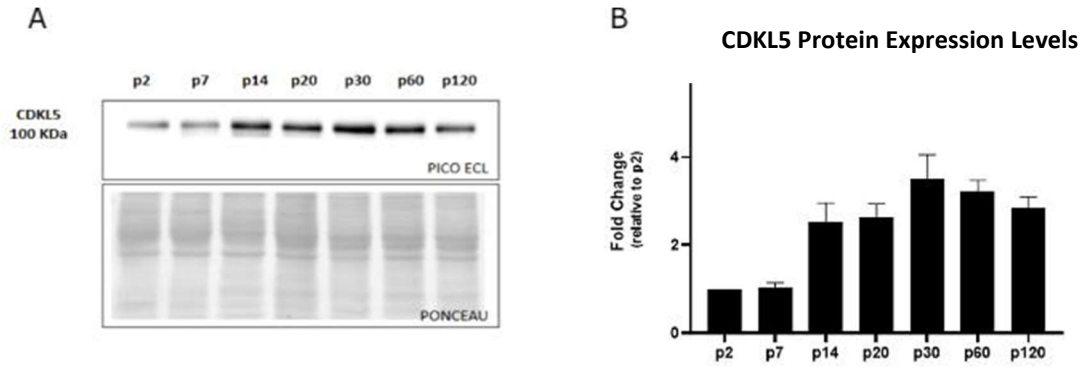


Fig. 5.1.2.1 CDKL5 protein levels in various mouse developmental stages. *(A) Endogenous CDKL5 protein levels measured in mouse cortex at various postnatal days: p2, p7, p14, p20, p30, p60, p120. Protein level increases until p30, with a moderate decrease at p60 and p120. Ponceau was used for total protein normalization of CDKL5 signal. (B) Three distinct experiments wereperformed and in all of them CDKL5 signals were represented as fold change of the protein level, using p2 as reference. The obtained data from three distinct experiments was*

representedas mean; error bars represent ±SEM.

Regarding mRNA levels, the measurements show the same trend seen with the protein levels, with peaks at p30, preceded by a gradual increase and followed by a decreasing trend (**Fig. 5.1.2.2 A**).

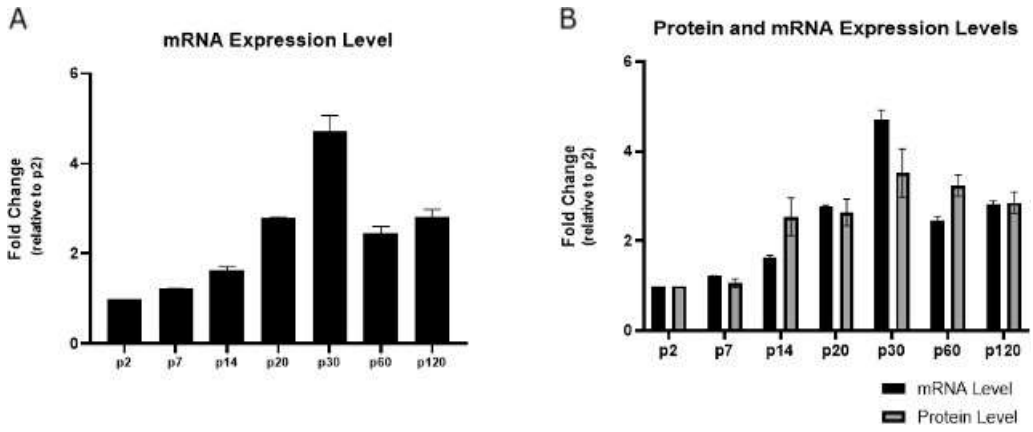


Fig. 5.1.2.2 CDKL5 mRNA levels in various mouse developmental stages*. (A) CDKL5 mRNA expression levels measured in mouse cortex at postnatal days (p) 2,7,14, 20,30, 60, 120, using RT-PCR. GAPDH and RPL13 were used as reference genes to quantify CDKL5 mRNA level in the samples, with identical outcomes. In this figure we reported the results obtained through RPL13 normalization. p2 was chosen as reference developmental stage, to express the variationsin CDKL5 mRNA level as fold change. The experiment was repeated three times, with three distinct set of cortices from the various postnatal days. The data were represented as mean; error bars represent ± SEM. (B) Histogram shows the comparison between the level of CDKL5protein and mRNA detected in the two set of experiments, presented as mean.; error bars represent ±SEM.*

We can conclude that, similarly to what we observed in the context of various tissues, the transcriptional program plays a major role in determining the expression levels of the protein in the context of murine cortex development (**Fig. 5.1.2.2 B**). As stated before, also this experiment should be performed in female individuals.

5.2 Rapid changes in CDKL5 levels in response to neuronal stimulation

5.2.1 Neuronal Stimulation in Vitro

To investigate the mechanisms of the rapid increase of CDKL5 protein levels in response to neuronal stimulation (La Montanara et al., 2015), we chose cortical neurons

as cellular model. We prepared neuronal cultures from wild-type mice cortices taken at E15, following the protocol reported in the Materials and Methods section. Treatments were administered at DIV14.

The experiment was repeated three times, starting from neuronal cultures obtained from three separate dissections. Two technical replicates were performed for each experiment.

The aim of the experiment was to reproduce the increase in CDKL5 protein levels in cortical neurons, following a drug-approach stimulation (La Montanara et al., 2015) to establish a model that would allow us to investigate possible post-transcriptional mechanisms compatible with the rapid changes reported in the literature. We used cortical neurons treated with vehicle as negative control for the experiments. The stimulation duration was set to ten minutes, according to the literature and to avoid any transcriptional-dependent effect.

The drugs chosen as treatments were the following:

- NMDA, used to stimulate the NMDA receptors. Evidence from La Montanara et al., 2015 has previously shown how this stimulation was successful in causing a rapid increase in CDKL5 protein levels that were attributed to a possible translational mechanism. Thus, this condition represents a positive control for the experiment, as well as allowing us to compare the effect of the other similartreatments.
- Ionomycin (IONO), an ionophore which triggers a Ca^{2+} increase within the cytosol by stimulation of both entry from the plasma membrane and release from intracellular stores, without the involvement of voltage-operated Ca^{2+} channels (Morgan et Jacob, 1994).

  Bicuculline (Bic), a competitive allosteric antagonist of the GABA-A receptor (Johnston, 2013). Bic was administered in the form of a quaternary salt (bicuculline methobromide) due to its major stability and solubility. The actionof bicuculline as an antagonist of GABAergic neurons causes the absence of inhibition within the neuronal network, thus leading to a sustained neuronal activation. In this manner, Bic plays a role in the enhancement of neuronal activity, promoting neuronal firing and maintaining sustained Ca^{2+} dynamics inside the neurons (Glass et al., 1980). For all these reasons, in accordance withthe possibility that Ca^{2+} influx is associated with CDKL5 translation (LaMontanara et al., 2015), Bic was considered a good stimulating agent for this kind of experiments.

After the incubation period with the individual treatments, neurons were lysed, and protein contents were harvested and prepared for WB analysis to assess CDKL5 protein levels. The duration of the treatment incubation was chosen in accordance with the literature (La Montanara et al., 2015) and taking into account the time frame of the translation mechanism ruling out transcriptional events (Sonenberg et al. 2000). Of course, having identified the drug treatments able to increase the level of CDKL5, the confirmation of the translational mechanism will require a final confirmation by using transcriptional (and translational) inhibitors.

Due to the sensibility of α-tubulin to the stimulations, which interfered with the detection of its protein levels, the normalization of the of CDKL5 protein signal was performed using β-Actin as internal reference.

As resulted from the analysis, we recorded a positive trend in the increase of the CDKL5 protein level in the samples in which neurons were treated with the pharmacological stimuli, that, as confirmed by the increase in the phosphorylation status of ERK1/2, caused a rapid neuronal activation. Noteworthy, we found that there was a strong increase in samples treated with bicuculline, an enhancer of the neuronal network activity, strengthening the working hypothesis that synaptic activity is linked to rapid expression of CDKL5 (**Fig. 5.2.1**).

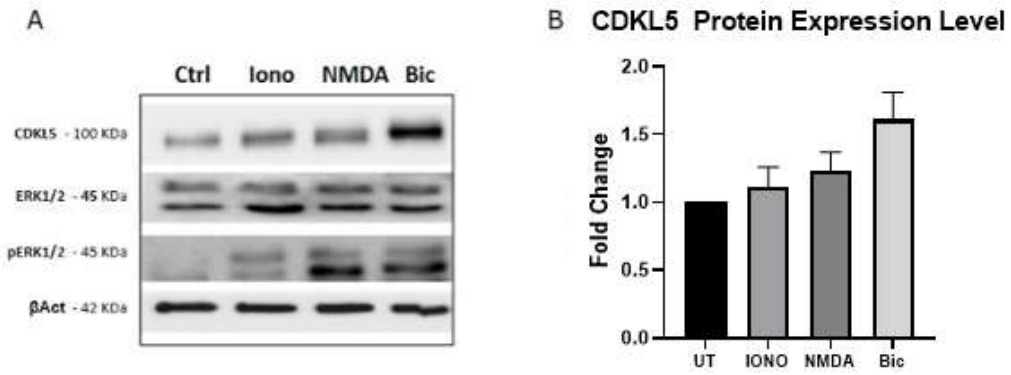


Fig. 5.2.1 Acute treatments in neurons cause an increase in CDKL5 protein levels. *(A) CDKL5 protein levels in cortical neurons E15 DIV14. UT = untreated sample; IONO = ionomycin 5uM; NMDA = NMDA 50uM; Bic = Bicuculline methobromide 10uM. Treatments were incubated for 10 minutes. pERK1/2 (45 kDa) was used to assess neuronal activation. B-Actin was used as the internal reference protein. All the stimulations increase the CDKL5 protein levels after 10 minutes of incubation. (B) Results of each experiment were normalized to the untreated sample and expressed as fold change of CDKL5 protein level. Data obtained from three distinct experiments was expressed as mean value and SEM was used as index of*

dispersion of the data. A one-way ANOVA was performed to compare the effect of the four different conditions. The analysis revealed that there was a statistically significant difference in means between at least two groups (F (10.41), p = 0.039). Tukey's multiple comparison test found that the mean value of Bic was significantly different to the other values, pointing towards this as the best acute treatment to increase CDKL5.

5.2.2 Neuronal Stimulations in Vivo

Building upon the results obtained from stimulations in vitro, we planned a new experiment to investigate rapid changes in CDKL5 expression in activated neurons in an in vivo model. In these experiments we administered bicuculline intraperitoneally (IP). After 30 minutes of incubation – a reasonable period to observe an effect on protein levels without the involvement of transcription (Sonenberg et al., 2000) - mice were sacrificed, and cortices were taken for protein extraction. CDKL5 protein levels were assessed by WB, using β-actin as internal reference. First, we examined two different concentrations of administrated bicuculline: 1 mg/kg and 3 mg/kg. Three animals for each group were compared with controls treated with saline. Both treatments caused an increase of CDKL5 that was close to the threshold of p = 0.05 (**Fig. 5.2.2 A, B**). Waiting for the availability of additional mice, and assuming that both treatments were producing a ceiling effect (indeed there was no significance between the two treated group), we evaluated also the two treated groups together – Bic 1mg/kg and Bic 3mg/kg – and compared them with the pooled untreated groups in order to obtain a first preliminary, but statistically significant, result about the effect of bicuculline on the CDKL5 level in mice cortices. So, the result confirmed as expected that CDKL5 expression levels were higher in the Bic-treated animals and the observed two-fold increase is interesting evidence, obtained *in vivo*, possibly linking CDKL5 translation to the enhancement of neuronal activation although artificially induced by Bic (**Fig 5.2.2 C, D**). In spite of the short time frame of the stimulation, a final confirmation of the translational nature of the observed effect, will require a RT-qPCR assessment of mRNA levels, together with a WB analysis of the phosphorylation status of ERK1/2 to confirm the neuronal activation upon Bic treatment.

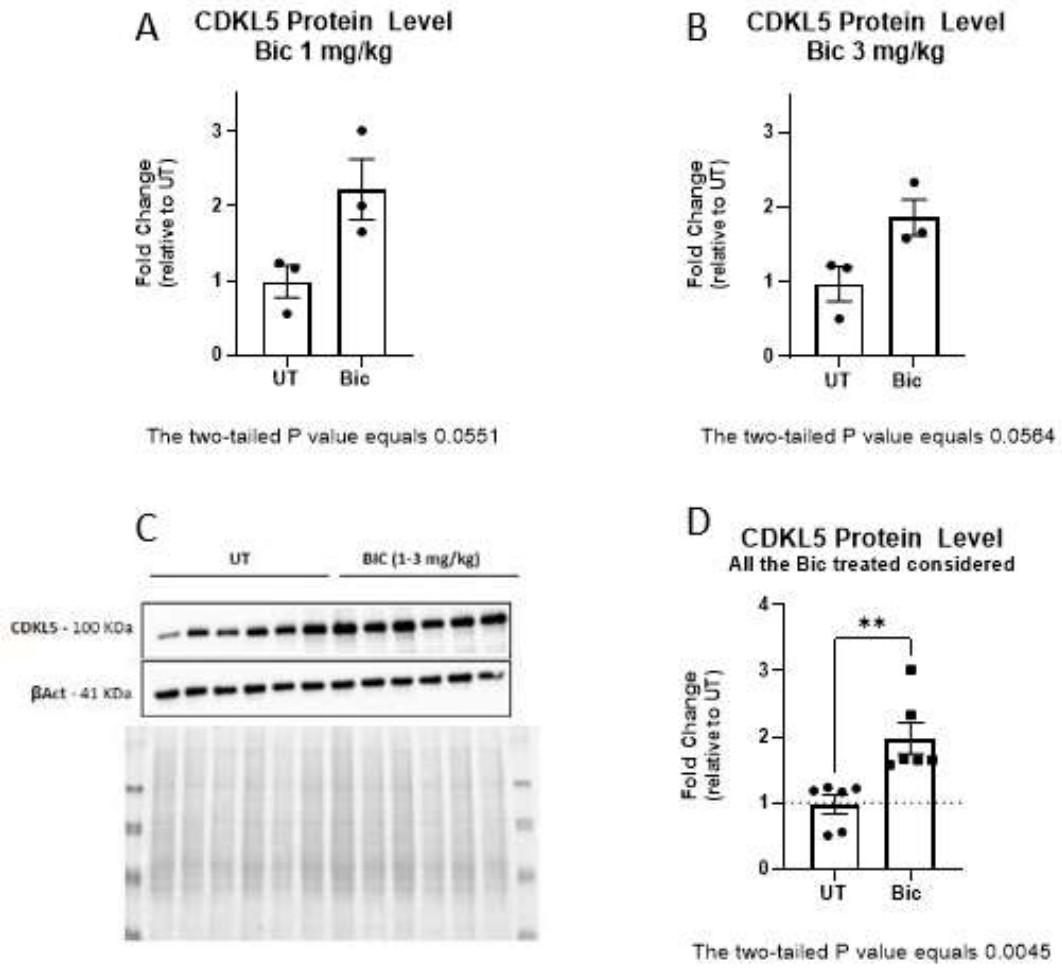


Fig. 5.2.2 Bicuculline treatment increases CDKL5 protein level in the mouse cortex. *CDKL5 protein levels in the mouse cortex of mice treated with saline solution (UT) or bicuculline at two different doses: 1mg/kg (A) or 3 mg/kg (B). Mice were sacrificed after 30 minutes from the injection and analyzed by Western blotting (C). CDKL5 always shows an increase in protein abundance in Bic treated group. βAct is the negative control and the internal reference. (D) Pooled data from the previous experiments evaluated with Student t-test (**p<0.01).*

5.3 Gene Silencing of eIF4B causes a reduction in CDKL5 protein level

In order to obtain further evidence of CDKL5 translational regulation, we performed gene silencing of the translational initiator factor eIF4B, known to be an adjuvant of the dead-box helicase eIF4A by acting in the initiation step of cap-dependent translation. Because eIF4B, although not essential in the translation process, it is involved in the

modulation of translation initiation, possibly acting as a helper in the unwinding of specific high-structured 5'UTRs, it was chosen as the target for gene silencing, given the 5'UTR structure of CDKL5. Moreover, it is known that among the preferential targets of its action there are proteins involved in modulating cellular apoptosis (Shahbazian et al., 2010), which is one of the cellular pathways in wherein CDKL5 might act as a regulator (Loi et al., 2020; Kim et al., 2020). In addition to these established functions, it was recently demonstrated that eIF4B can act as a key player in the activity-dependent translation in neurons (Bettegazzi et al., 2017; Bettegazzi et al., 2021).

Therefore, our first aim was to test whether the downregulation of eIF4B using small interference RNA can cause a decrease in CDKL5 protein level in resting conditions. The result of this experiment is interesting, because for the first time it would show the importance of initiator factors involved in modulating translational efficiency of CDKL5 transcript and, consequently, its expression levels. Moreover, eIF4B is sensitive to neuronal activation (Bettegazzi et al., 2017), and positive results could lead to experiments aimed to test its involvement in CDKL5 local translation.

The experiment was conducted on SHSH-5Y cells. Gene silencing, performed using eIF4B siRNA incubated on the cells for 72h, was conducted following the protocol reported by Shahbazian et al., 2010. α-tubulin and GAPDH were used as negative controls, as reported in the paper, while a positive control, XIAP, was chosen from the list of proteins that were previously reported to be influenced by eIF4B gene silencing (Shahbazian et al., 2010) (**Fig. 5.3.1 A, B**). We also evaluated another interesting gene, MeCP2, involved in Rett syndrome, which shares several phenotypic manifestations of CDD (as discussed in **par. 3.2**). We performed a preliminary bioinformatic analysis of the transcript leader of MeCP2 and found features that strictly resemble the ones of CDKL5. This analogy prompted us to investigate in parallel also MeCP2 and, even though this goes beyond the aims of this project, it was included in the analysis.

WB analysis, performed on cells treated with eIF4B specific siRNA, showed the success of siRNA transfection, resulting in a decrease of eIF4B by 80% (**Fig. 5.3.1 A, B**). Indeed, when the eIF4B gene was silenced, CDKL5 protein levels were significantly decreased. The effect on CDKL5 - a decrease of about 50%, links for the first time the expression level of CDKL5 with the availability of a translational initiator factor that works in the unwinding of cis-acting regulatory motifs on the 5'UTRs of the

transcripts (**Fig. 5.3.1 A, D**).

Interestingly, gene silencing of eIF4B had an effect also on MeCP2 protein level, reducing its expression level by about 50% (**Fig. 5.3.1 A, E**). This result, in line with the bioinformatic analysis of the MeCP2 transcript, could suggest a new link between CDKL5 and MeCP2. Interestingly, since the general features of their 5'UTRs are similar, further studies are required to clarify the relationships among these proteins.

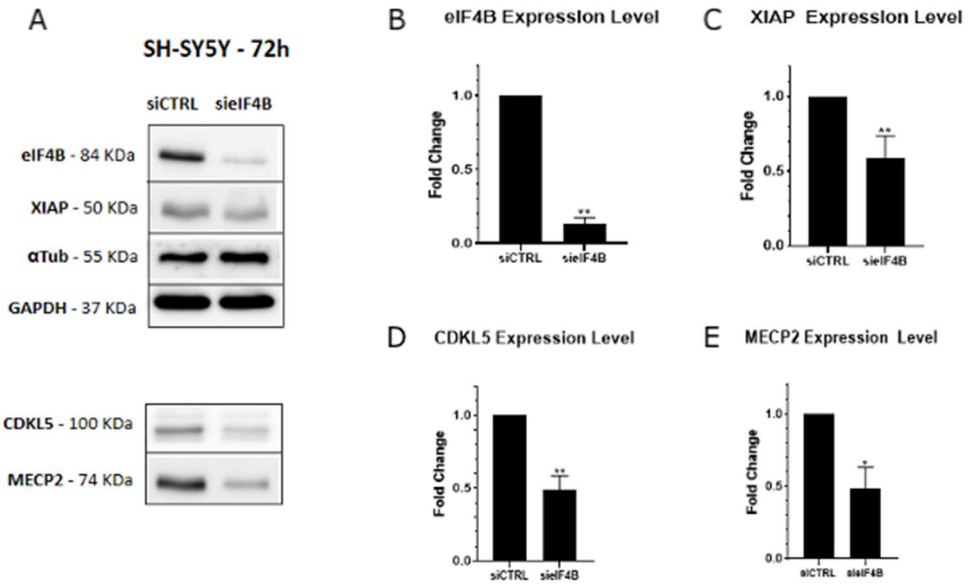


Fig. 5.3.1 A decrease in eIF4B protein results in lower CDKL5 and MECP2 levels. *(A) siCtrl = sample obtained from cells transfected with a random siRNA, used as negative control; sieIF4B = sample obtained from cell properly transfected with sieIF4B. The incubation period was 72h. XIAP was used as positive control for the effect of eIF4B gene silencing with siRNA; αTub was used as internal reference; GAPDH and αTub were considered as negative controls. (B) eIF4B shows a decrease in its protein level of about 80%. Consequently, as expected, XIAP expression levels were reduced by 50% (C). CDKL5 and MeCP2 showed a decrease in their protein expression of about 50% (D, E). The decrease in expression levels were statistically significant, as determined through Student-t test analysis (*p<0.05; **p<0.01; n=3).*

5.4 Analysis of CDKL5 5'UTR Transcript Variants

5.4.1 CDKL5 5'UTR Variants

In order to have a better description of the various 5'UTR variants, CDKL5 leader sequences obtained from different sources as have been taken in consideration, i.e. the RefSeq Transcript Variants of CDKL5 present in GenBank (NCBI), Ensembl, and from the unique paper in which different 5'UTRs of CDKL5 have been reported to date

(Hector et al., 2016). All the reported leader sequences share a common part in exon 2, which contains the unique start codon of CDKL5 protein isoforms. However, although the leader sequences of CDKL5 in NCBI presents only two different variants (one in NM_003159.3/NM_003159.3 and the other in NM_001323289.2), the number of alternative first exons in the 5'UTRs reported in Ensembl exceed ten. The high variability in alternative untranslated first exon usage is confirmed by Hector's work, which presents additional alternative first exons. These were experimentally confirmed only by non-quantitative amplification approach and did not mention some of the ones reported by Ensembl. On the other hand, some of the untranslated exons determined by Hector's group are not included in the 5'UTR variants verified by Ensembl, and the ones that are in common in the two lists are often different at the 5' and 3' terminations. This discrepancy can be explained in part by the limitation of the experimental approach used by Hector and colleagues in exploring the 5' terminations of the CDKL5 transcript. In fact, the authors themselves mentioned that their 5'RACE approach does not allow a proper amplification of GC-rich sequences, causing a lack of precision in determining 5'UTRs sequences.

For this reason, in a preliminary analysis of 5'UTR variants of CDKL5 we chose to consider all the different first exons found in these three sources as possible players in the composition of the leader sequences (**Tab. 5.4.1**).

	ID	Annotation	Start	End
1	Ex1_202	Ups.NM_003159.3/NM_003159.3	18425583	18425695
2	Ex1_205:Ex1_207:Hector_a	NM_003159.3/NM_003159.3	18425608	18425695
3	Hector_a1		18426069	18426401
4	Ex1_209		18426198	18426401
5	Hector_b1		18426691	18426919
6	Ex1_211		18426713	18426919
7	Hector_b		18426876	18426919
8	Ex1_201		18442188	18442261
9	Ex1_208_a		18442213	18442423
10	Ex1_204_a:Hector_c	NM_001323289.2	18442224	18442261
11	Ex1_203		18442275	18442390
12	Ex1_204_b:Hector_d	NM_001323289.2	18442375	18442423
13	Ex1_208_b:Hector_e1		18443819	18443959
14	Hector_e		18443839	18443959
15	Ex1_204_c		18457483	18457534

Tab. 5.4.1 List of all the alternative first exons of human CDKL5 5'UTR reported to date. *Sequences were taken from RefSeq TVs, Ensembl, and from the paper by Hector and colleagues (Hector et al., 2016). For each sequence are reported: ID = identification code; Start: the start position of the sequence, thus the TSS; End: the position in which the sequence s ended. Annotations are reported for identify first exons of RefSeq TVs. A total of 15 alternative first exons are classified and ordered based on their start site. Assembly GRCh38.*

The three characteristic features of functional 5'UTRs we focused on were: high GC content, the presence or absence of uAUGs/uORFs, and, most importantly, the extremely high conservation of the sequence among species (Davuluri et al., 2000;

Araujo et al., 2012).

5.4.2 GC-content Analysis

GC Content, expressed as GC%, is reported in Table 5.4.1.2. We also distinguished the G% and C%, considering the established role of G-rich sequences in determining structural motifs. Alternative first exons were analyzed individually and compared to the common untranslated region of exon 2 (UTex2) in order to understand the specific contribution of the first exon to the GC content of the entire 5'UTR (**Tab. 5.4.2.1**).

The result of the analysis showed that the GC Content of alternative first exons is generally higher than the content of the UTex2, with the highest contribution given by G. We chose a GC% of 60 as threshold to estimate if the GC Content of the sequences analyzed could be considered high enough to be comparable to the GC content of other functional 5'UTRs. This threshold was chosen from the indications reported in the CART classification of human 5'UTR (Davuluri et al., 2000). From the analysis, only three first exons exceed the threshold: the Ex1_205 (RefSeq NM_003159.3/NM_003159.3), the Ex1_202 reported in Ensembl and the third was Hector-a1. Ex1_202 includes entirely Ex1_205, as it has the TSS 25nts upstream the TV 205's TSS. The first exon 202 is not reported in any papers concerning CDKL5 to date.

Taking in consideration the single contribution of G and C in calculation of GC%, exons with a high G over C ratio were considered particularly interesting, because of the principle that this feature is associated with the possible presence of G4 structures. Interestingly, the higher G% is calculated for Ex1_205 that is indeed the most represented first exon in the CDKL5 transcripts found in brain (Hector et al., 2016) (**Fig. 5.4.2.2**).

	ID	Length	GC%	G%	C%	G/C
1	Ex1_202	113	67.3	31.71	24.39	1.30
2	Ex1_205:Ex1_207:Hector_a	88	67	43.18	23.86	1.81
3	Hector_a1	333	68	33.00	35.00	0.94
4	Ex1_209	204	39.5	29.90	29.41	1.02
5	Hector_b1	229	35.40	21.40	13.97	1.53
6	Ex1_211	207	35.3	21.26	14.01	1.52
7	Hector_b	44	38.60	18.18	20.45	0.89
8	Ex1_201	74	56.1	43.18	23.86	1.81
9	Ex1_208_a	211	50.2	24.17	26.07	0.93
10	Ex1_204_a:Hector_c	38	57.9	47.37	10.53	4.50
11	Ex1_203	116	44.8	15.52	29.31	0.53
12	Ex1_204_b:Hector_d	49	57.1	18.37	38.78	0.47
13	Ex1_208_b:Hector_e1	141	46.1	22.70	23.40	0.97
14	Hector_e	121	43.00	20.66	22.31	0.93
15	Ex1_204_c	52	46.2	23.08	23.08	1.00
	UTEx2	162	39.5	20.99	18.52	1.13

Tab. 5.4.2.1 GC Content Analysis of exons of CDKL5 5'UTR. *For all the sequences taken in consideration, length (number of nucleotides of the sequence); GC% (GC content expressed as percentage on the entire composition of the sequence); G% (guanine content expressed as percentage) and C% (cytosine content expressed as percentage) as reporter. Moreover, GC ratio, indicated as G/C is considered.*

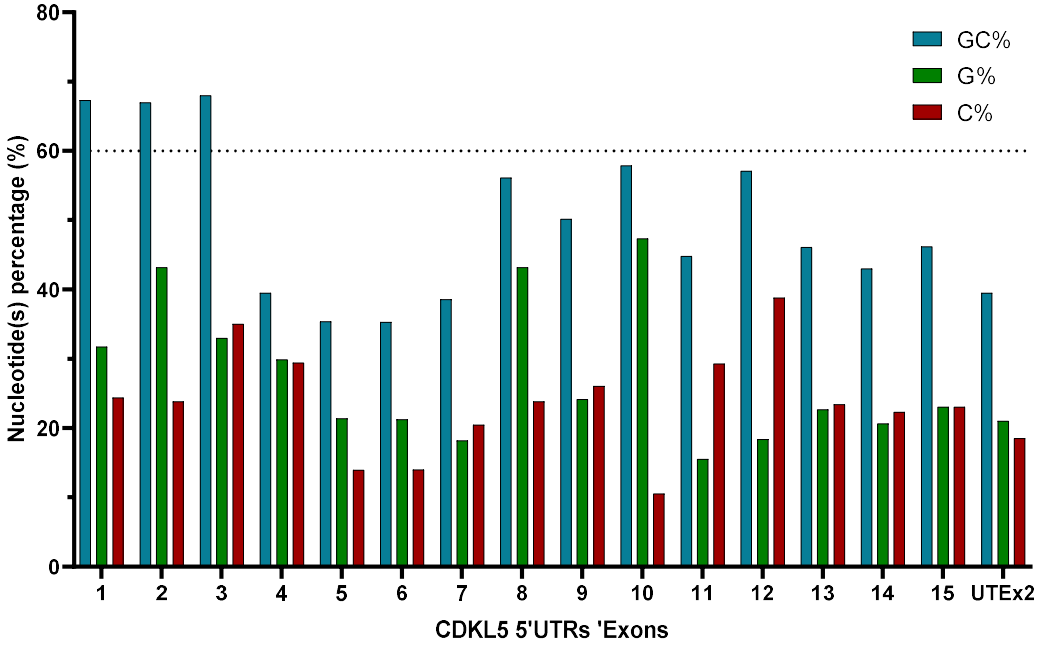


Fig. 5.4.2.2 GC%, G% and C% of CDKL5 5'UTR exons. *GC% of all the sequences of alternative first exons and UTex2 was plotted. A threshold of significance of 60% was set, in accordance with the literature (Davuluri et al., 2000). Ex1_202, Ex1_205 and Hector_a1 have a significative GC content, exceeding the threshold. Moreover, Ex1_205 and Ex1_202 have a G content twice as high as C. UTex2 is not particularly GC rich when compared to first exons.*

5.4.3 uAUG and uORF prediction

Another feature pointing towards a possible function in the 5'UTR sequence is the presence of uORFs since they are known to modulate the translation efficiency of the main ORF by competing in the formation of the TC (Calvo et al., 2008; Lee et al., 2012; Hinnebusch et al., 2016). The uORFs prediction is mainly based on the evaluation of the translation initiation start sites (TISs) considering the Kozak consensus sequence. Indeed, the various transcript leaders of CDKL5 include several AUGs that might act as start codon of putative uORFs (Diaz de Arce et al., 2017). We did not consider at this stage of the analysis non-AUG TSSs, and we used NetStart (DTU Health Tech), an open-source tool that can predict potential uAUGs in a favorable Kozak context for the

translation initiation. The algorithm of NetStart is developed on artificial neuronal networks that takes into consideration the local context of the AUG and the global sequence tested. It has reported a success rate of 85% for NetStart predictions, with an output score value ranging between 0 and 1, where score >0.5 is the threshold for a good prediction of functionality (Pedersen et al., 1997; Ozretic at al., 2015). The sequences subjected to the analysis were the UTex2 and all the alternative first regions of the CDKL5 5'UTR variants – that can be represented by a single first exon, as the case of Ex1_205, or be composed by two first exons, as the case of Ex1_204. The result of the NetStart analysis is reported in the **Tab. 5.4.3**, where for every sequence are reported the position of a detected uAUG, its score and the prediction as functioning TIS. It emerges that even if several uAUGs are found in most of the sequences – with exceptions of Ex1_205 and Ex1_202 – only two uAUGs are predicted to be TISs, showing a score value > 0.5. More specifically, a predicted functioning uAUG was found at position 187 of the first exon of TV 208, while another with a score >0.5 was found at position 89 of the first exon of TV 211.

To evaluate the goodness of the predictions obtained with NetStart, we compared its results with those obtained from TIS Miner (Liu et al., 2004), another tool based on different prediction algorithm. TIS Miner also allows to detect AUGs as TIS by evaluating the Kozak context. The tool was trained on 3312 vertebrate transcript sequences obtained by GenBank, with an accuracy of the classification model of 92,45%. TIS Miner analysis returns a score between 0 and 1, where higher scores are indicative of a good predicted uAUG in a favorable Kozak context. The considered threshold for the present analysis was set to 0.5, taking into consideration the information provided by the creators of the software (Liu et al., 2004). The results obtained by TIS Miner confirmed that the AUG codons detected in all the sequences analyzed show a low score for the TIS prediction. Concerning the two uAUG predicted to be TIS by NetStart, TIS Miner did not confirm the prediction for the uAUGs found in Ex1_211 and Ex1_208, even though the AUG in Ex1_208 was better than the one in Ex1_211, and slightly below the threshold (**Tab. 5.4.3**).

A **B**

	NetStart-1.0 Analysis			TIS Miner	
Name	**Pos.**	**Score**	**Pred.**	**Score**	**Identity to Kozak consensus [AG]XXATGG**
UTex2	19	0.16	No	0.379	CXXATGA
	34	0.216	No	0.235	TXXATGT
	125	0.073	No	0.39	CXXATGT
202	No	No	No	No	
205	No	No	No	No	
201	99	0.112	No	0.03	CXXATGG
203	No	No	No	0.03	CXXATGG
204	63	0.094	No	0.216	CXXATGG
	112	0.26	No	0.02	CXXATGG
207	172	0.16	No	0.034	CXXATGG
208	**187**	**0.62**	**Yes**	**0.484**	CXXATGG
	312	0.051	No	0	AXXATGA
209	79	0.246	No	0.209	GXXATGC
	120	0.159	No	0.101	AXXATGG
211	14	0.159	No	0.011	AXXATGG
	29	0.18	No	0.017	AXXATGT
	89	**0.589**	**Yes**	0.061	GXXATGT
Hector_a1	208	0.17	No	0.241	GXXATGC
	249	0.73	No	0.011	AXXATGG
Hector_b1	36	0.173	No	0.019	AXXATGG
	51	0.161	No	0.017	AXXATGT
	111	0.595	Yes	0.061	GXXATGT
Hector_b	No	No	No	No	
Hector_cd	63	0.081	No	0	AXXATGA
Hector_e	81	0.053	No	0.009	CXXATGG
Hector_e1	101	0.049	No	0	AXXATGA

Tab. 5.4.3 Predictions of uAUGs/uORFs. *(A) Predictions performed with NetStart 1.0. While a great number of uAUGs are detected in almost every sequence, only Ex1_211 and Ex1_208 show score values > 0.5, and thus are considered to potentially contain functional uAUGs. (B) Predictions performed by TIS Miner. The analysis confirmed the results obtained from NetStart, with the exception for the prediction in the Ex1_211, which resulted to have a very low score probably due to the unfavorable Kozak context. However, none of the sequences analyzed are predicted to contain significant uAUGs. Name (identification code); Score (score of the predictions; NetStart and TIS Miner thresholds are both set to 0.5); Pos. (position in which there is predicted the uAUG); Pred (result from the NetStart analysis).*

In the end, the two uAUGs predictions returned the absence of significative AUG TSS prediction in the UTex2 and in the first exons of 5'UTRs, with an exception for an

uAUG in the Ex1_208, that was detected by both the software and thus is more probable a TSS of a functioning uORF.

5.4.4 Conservation of First Exons and UTex2

From the point of view of the conservation of the sequence, we used alignments as the primary method for the investigation of the evolutionary conservation of UTex2 and the alternative first exons, based on the assumption that 5'UTR regions with functional features based on cis-acting regulatory elements would show a certain degree of conservation that is uncommon in untranslated sequences.

First, we analyzed the sequence of UTex2, i.e. the common part of all the 5'UTR variants. We used nucleotide BLAST to search human UTex2 as query sequence, with the BlastN algorithm. The choice of BlastN was made considering that it allows a word-size comparison down to seven bases, which is useful to find short sequences with a percentage of identity that is not expected to be high, like within the untranslated regions of the transcripts. In fact, the E value was gradually set higher up to 0.1 for each iteration in order to obtain more sequences associated with the query sequence. In addition, the scoring parameters of match/mismatch and gap costs were set to promote the inclusion of more homologous sequences. The database selected for the search was a nucleotide collection, consisting of sequences present in various databases such as GenBank, EMBL, PDB. The aim of the analysis was to assess first the conservation of UTex2 in comparison with the adjacent translated region of exon two (Tex2), as well as to analyze the conservation of the first exons of CDKL5 gene. The UTex2 search returned 490 sequences from 168 vertebrates, most of which are mammals (including placentals and marsupials), but, interestingly, there were also three partial sequences of birds from two genus, with 25% of query cover. - *Dromaius novaehllandiae* (Emu), *Apteryx rowi* and *Apteryx mantelli* (Brown kiwis) (**Fig. 5.4.4.1 A**). From the alignment of 27 selected sequences chosen from the mentioned hits, it appears that UTex2 could be divided in three regions based on the conservation:

- the more conserved part, common to all the sequences found by BLAST, that is the more proximal region to the main AUG (118nts to 162nts);
- the middle region (approximately from 60 to 118 nts) that is the less conserved and is absent in sequences of the group of chiroptera, rodentia, marsupialia, peryssodactila and aves;
- the distal region (approximately from 1 to 60 nts), surprisingly is found to be conserved in marsupials and some members of the chiroptera order, suggesting that the

length of the exon 2 might have also been conserved in more ancient species of mammals (**Fig. 5.4.4.1 B**)

On the contrary, the translated region of CDKL5 exon 2 (Tex2) in the same species is highly conserved, with a query cover of 100% and with the maximum percentage of identity (100%)- except for the sequences of *Mus musculus* and *Rattus norvegicus* (**Fig. 5.4.1.1 A**). We must consider that the library in which we performed this analysis – the nucleotide collection - is full of predicted sequences, that represent most of the sequences for our selection - even if the sequences of *Mus musculus* and *Apteryx mantelli* are experimentally verified. For this reason, an additional BLAST analysis to improve the reliability of the result is necessary.

A

Accession Number	Scientific Name	UTex2			Tex2		
		Query Cover	E value	Per. ident	Query Cover	E value	Per. ident
XM_016943184.2	Pan troglodytes	100%	1.00E-80	100	100%	1.00E-24	100
XM_028842232.1	Macaca mulatta	95%	6.00E-53	92.26	100%	1.00E-24	100
XM_012784691.2	Microcebus murinus	100%	6.00E-42	87.5	100%	1.00E-24	100
XM_004269638.2	Orcinus orca	100%	9.00E-37	86.83	100%	1.00E-24	100
XM_024988497.1	Bos taurus	97%	3.00E-29	84.66	100%	1.00E-24	100
XM_032840837.1	Lontra canadensis	100%	1.00E-28	85.37	100%	1.00E-24	100
XM_036305037.1	Myotis myotis	100%	1.00E-28	83.93	100%	1.00E-24	100
XM_045051234.1	Felis catus	100%	6.00E-27	87.04	100%	1.00E-24	100
XM_013509898.1	Chinchilla lanigera	97%	2.00E-23	81.6	100%	1.00E-24	100
XM_038449549.1	Canis lupus familiaris	100%	2.00E-23	83.54	100%	1.00E-24	100
XR_002340830.1	Sus scrofa	97%	2.00E-19	82.21	100%	1.00E-24	100
XM_033119851.1	Rhinolophus ferrumequinum	76%	1.00E-13	91.3	100%	1.00E-24	100
XM_015488064.1	Marmota marmota marmota	29%	3.00E-11	95.83	100%	1.00E-24	100
XM_017602292.2	Rattus norvegicus	29%	3.00E-11	95.83	100%	8.00E-23	98.44
XM_023633572.1	Equus caballus	52%	1.00E-09	97.62	100%	1.00E-24	100
NM_001024624.2	Mus musculus	29%	1.00E-09	93.75	100%	8.00E-23	98.44
XM_006911629.3	Pteropus alecto	41%	6.00E-09	88.06	100%	1.00E-24	100
XM_039852327.1	Pteropus giganteus	41%	6.00E-09	88.06	100%	1.00E-24	100
XM_031958923.1	Sarcophilus harrisii	72%	2.00E-08	95.35	100%	1.00E-24	100
XM_027835407.1	Vombatus ursinus	72%	2.00E-08	95.35	100%	1.00E-24	100
XM_021007398.1	Phascolarctos cinereus	69%	1.00E-06	82.86	100%	1.00E-24	100
XM_043992522.1	Dromiciops gliroides	72%	1.00E-06	93.02	100%	1.00E-24	100
XM_026120748.1	Dromaius novaehollandiae	25%	0.013	88.1	100%	1.00E-24	100
XR_003252799.1	Apteryx rowi	25%	0.013	88.1	100%	1.00E-24	100
LK064688.1	Apteryx mantelli mantelli	25%	0.013	88.1	100%	1.00E-24	100

B

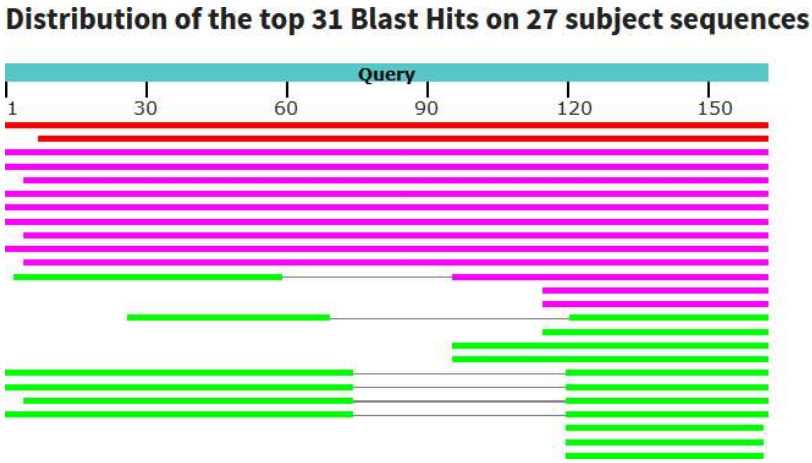


Fig. 5.4.4.1 Conservation of UTex2. *(A) The table reported 25 selected hits returned from the nBLAST analysis in the database nucleotide collection using as query the human UTex2 sequence, taken from Ensembl. nBLAST search was performed using the BLASTN algorithm, with a threshold E value of 0.1 and the scoring parameters of match/mismatch and gap costs respectively at 1/-2 and 1/2, to promote the search of more dissimilar sequences. For each selected sequence the following information is reported: accession number, scientific name of the specie in which the sequence was found, query cover, E value and percentage of identity for the UTex2 and for the Tex2 (translated region of the exon 2). The selected 25 sequences were chosen from the 490 sequences returned from the nBLAST search based on the taxonomicgroup which they belong to, in order to reproduce a representative sample of the results and, also, to highlight the more interesting hits returned. Table rows are colored to indicate the phylogenetic group of belonging: light green for primates, green for artiodactyla, pink for carnivora; light blue for chiroptera; grey for rodentia; dark gray for perissodactyla; orange for marsupials and yellow for aves. (B) Graphical representation of the distribution of the 25 selected hits on the query sequences UTex2, showing that the sequence can be divided in threesubregions with different conservation, at 60 nts intervals. The most conserved subregion wasthe proximal subregion (120-162), detected in aves, the more distant phylogenetic group to human in which UTex2 was detected in the analysis. The middle subregion (60-120 nts) is the less conserved part. The distal subregion (1-60 nts) is more conserved that the middlesubregion, as showed by the fact that species of more ancient taxa, as marsupials and some chiroptera, return the conservation from the distal and the proximal subregions but did not recognize the middle subregion in the alignment with the query, as showed by the uncolored line.*

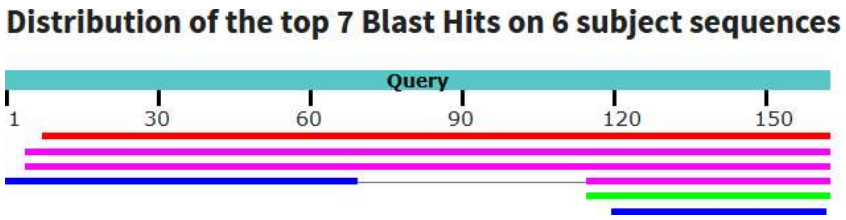
Therefore, a second BLAST search was performed in dbEST, a different database of

sequences. In dbEST each sequence is a retrotranscribed short fragment that was experimentally certified, but the number of species included in the database is exiguous compared to the number of species reported in the nucleotide collection. For this reason, both BLAST searches are important and necessary: nucleotide collection search provides a better overview of the evolution of the sequences, whereas the dbEST search acts as a control of the nucleotide collection search, providing experimental information that can validate or not the first result. In fact, in this case the BLAST search in dbEST - performed using the same parameters of the first search - returned sequences found in only 6 different organisms, two of which are from the group of primates (*Macaca fascicularis* and *Callithrix jacchus*), two from rodentia (*Mus musculus* and *Rattus norvegicus*), one from another mammals (*Bos taurus*) and one from a bird (*Lonchura striata domestica*). It is noteworthy that the alignment of the sequence found in *Rattus norvegicus* confirms the length of the UTex2 and the major conservation of the distal region (1-60 nts) of the exon 2 rather than its middle part (60-120 nts), as observed in the search in the nucleotide collection. In addition, the sequence from the bird *Lonchura striata* - verified to be adjacent to the coding region of CDKL5 as a real positive result - strengthens the presence of a part of UTex2 in the class aves identifying it as the first phylogenetic group in which UTex2 is detected (**Tab. 5.4.4.2**).

A

Scientific Name	Total Score	Query Cover	E value	Per. ident
Macaca fascicularis	219.000	0.950	0.000	92.260
Callithrix jacchus	177.000	0.970	0.000	88.050
Bos taurus	141.000	0.970	0.000	84.660
Rattus norvegicus	128.000	0.720	0.000	95.830
Mus musculus	75.700	0.290	0.000	93.750
Lonchura striata domestica	46.800	0.250	0.045	85.710

B



Tab. 5.4.4.2 UTex2 nBLAST analysis in dbEST. *(A) The analysis returns hits in six dbEST libraries from different species, using the same parameters of the first nBLAST analysis in nucleotide database. (B) Diagram of distribution of the hits on the query sequence, where they are reported with the color code described in Fig.5.4.4.1 A. The presence of the proximal*

subregion of UTex2 in birds is confirmed, since the analysis returns a hit in dbEST library of Lonchura striata domestica. The difference in the degree of conservation between the three subregions of the UTex2 is confirmed as appreciable from the distribution of the Rattus norvegicus sequences on the query.

Concerning the evaluation of the degree of conservation of the alternative first exons, we performed the BLAST search in the nucleotide collection with the same parameters used for the analysis of UTex2. The result showed that only the sequence Ex1_202 - containing the sequence Ex1_205 and Hector_a - found homologous sequences in 66 mammalian species (placentals and marsupials both) (**Fig. 5.4.4.3**). On the contrary, all the other sequences used as queries returned few hits and mainly in primates and other few species of placentals, in which the query cover was not greater than 50%. Moreover, when BLAST search was performed in dbEST, the majority of the other alternative first exons returned at least a result in human libraries, but only the Ex1_202/205 returned hits also for non-primate species, such as *Mus musculus* and *Bos taurus*, strengthening the evidence of the presence of this sequence in CDKL5 transcripts of other species.

	Ex1_ 202/205			
Scientific Name	Query Cover	E value	Per. ident	Accession Number
Pan troglodytes	100%	5.00E-51	100	XM_016943184.2
Macaca mulatta	100%	5.00E-51	100	XM_028842232.1
Microcebus murinus	69%	4.00E-25	94.94	XM_012784695.1
Orcinus orca	69%	8.00E-14	87.18	XM_033427878.1
Bos taurus	100%	1.00E-30	89.38	XM_024988497.1
Lontra canadensis	86%	6.00E-24	88.78	XM_032840837.1
Myotis myotis	89%	1.00E-20	86.14	XM_036305036.1
Felis catus	100%	8.00E-36	92.04	XM_045051234.1
Chinchilla lanigera	59%	2.00E-11	88.06	XM_013509898.1
Canis lupus familiaris	40%	3.00E-04	89.13	XM_038449544.1
Sus scrofa	100%	2.00E-27	87.61	XR_002340830.1
Rhinolophus ferrumequinum	66%	2.00E-19	92	XM_033119851.1
Marmota sp.	100%	7.00E-32	90.27	XM_027927442.2
Rattus norvegicus	94%	3.00E-18	84.4	XM_017602292.2
Equus caballus	100%	1.00E-33	91.15	XM_023633572.1
Mus musculus	97%	1.00E-21	85.45	AL670462.8
Pteropus alecto	93%	6.00E-24	87.85	XM_006911629.3
Pteropus giganteus	93%	2.00E-19	85.45	XM_039852327.1
Sarcophilus harrisii	56%	4.00E-08	85.94	XM_031958923.1
Phascolarctos cinereus	36%	2.00E-05	92.68	XR_002328379.1
Dromiciops gliroides	56%	7.00E-10	87.5	XM_043992522.1

Fig 5.4.4.3 nBLAST analysis of CDKL5 Ex1_202/205. *Blast search performed with the same parameters used in the previous analysis reported in Nucleotide collection for the query Ex1_202/205. The selection of 25 species presented in the Fig. 5.4.4.1 are presented here to compare Ex1_202/205 hits features, such as Per Id, with the ones returned using UTex2 as query. Percentage of query coverage (query cover), E value and the percentage of identity (Per Id).*

Our analysis confirms that the RefSeq Ex1_205 and Ex1_202 can be considered as highly conserved untranslated regions with a good degree of reliability. The control through the BLAST search in dbEST confirmed its presence in four different libraries from placentals. Nevertheless, the difficulty in obtaining the 5' termination of transcripts caused a lack of completed records in all the returned hits and resulted in a partial query cover for some hits not allowing the quantitative evaluation of these results. As a matter of fact, the evaluation of all the first exons of CDKL5 indicates the Ex1_202

- and thus Ex1_205 – as the primary sequence of interest in the composition of CDKL5 5'UTR. The comparison of the sequence with UTex2 showed how Ex1_202 is the more conserved part of the 5'UTR, measured as percentage of identities and percentage of gaps between the two query sequences and aligned sequences from the common species (**Fig. 5.4.4.4**). This finding highlighted how the Ex1_202/205 resulted to be highly conserved and thus potentially very important in the 5'UTR function, even though the UTex2 is the oldest region of CDKL5 5'UTR, being present in marsupials and even birds. Other alternative first exons which included the other RefSeq sequence, here reported as Ex1s_204, were evaluated to be either not highly conserved or poorly assessable due to the scarcity of hits returned.

In conclusion, our evaluation of possible cis-acting regulatory motifs on the 5'UTRs of CDKL5 pointed out that, among the alternative first exons, Ex1_202/205 seems to retain a relevant function according to the level of conservation. Moreover, the abundance of GC content and the absence of predicted uAUGs focused our attention on structural motifs as most probable cis-acting regulatory elements in this sequence.

A

	Ex1_202			UTex2		
Species	**% Id.**	**E value**	**Gaps %**	**% Id.**	**E value**	**Gaps %**
Pan troglodytes	100.00%	5.00E-52	0	100	5.00E-79	0
Mus musculus	85.09%	1.00E-23	2	77.89	2.00E-18	4
Sus scrofa	87.61%	1.00E-28	0	82.4	1.00E-29	4
Lutra lutra	89.47%	5.00E-32	0	85.98	4.00E-40	3

B

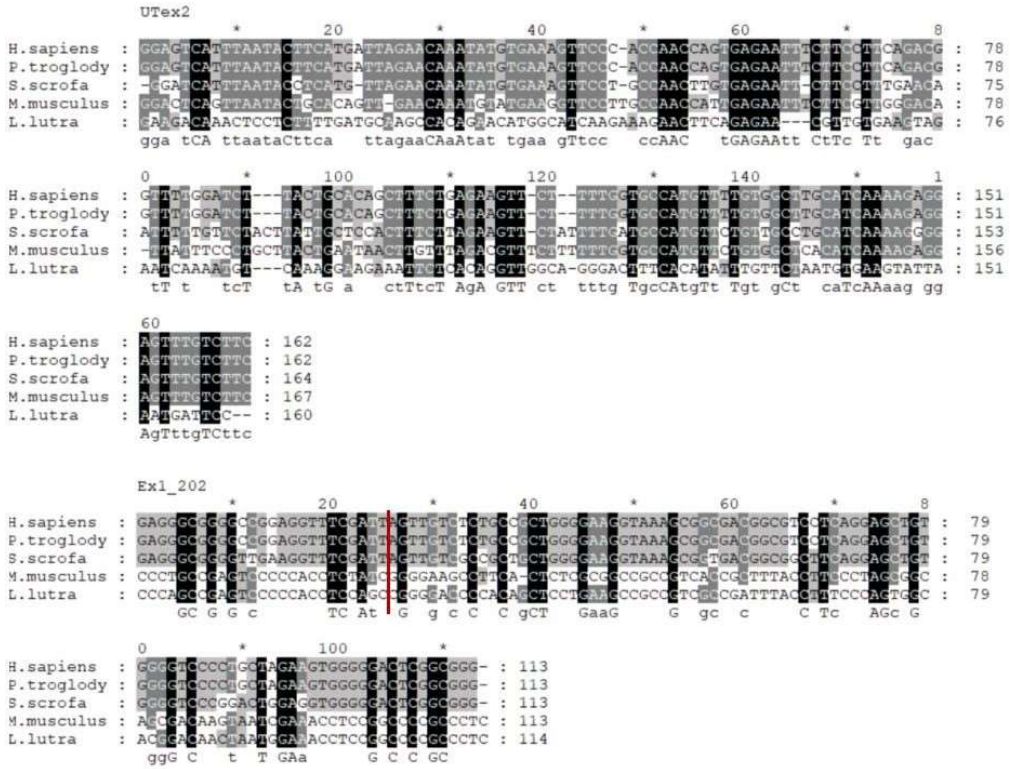


Fig. 5.4.4.4 Comparative Analysis of Ex1_202 and UTex2 conservation. *(A) The analysis takes in consideration the percentage of identity (% Id.), the E valued (measure of the reliability of the homology between the query and the hits returned by nBLAST) and the percentage of gaps in the aligned sequence (Gap %) as parameters for determining the degree of conservation. The human queries are shown in comparison with the sequences of Pan troglodytes, Mus musculus, Sus scrofa and Lontra canadiensis homologous. As a result, Ex1_202 (containing Ex1_205, the TSS of which is indicated by an upstream red line, at position 26) is found to be more conserved than UTex2, showing less gaps percentage and higher percentage of identity with the human query than the UTex2. (B) Grafical output of the alignments of Ex1_202 and*

UTex2, obtained using ClustalX and GeneDoc. Scale of grey is used to indicate the conservation status of each position of the queries.

5.5 Experimental Confirmation of the Existence of First Exon 202

Following the results obtained from the preliminary analysis of alternative first exons of the CDKL5 5'UTR variants, we decided to proceed with an in-depth characterization of the Ex1_202. The specific features of Ex1_205, and consequently of the Ex1_202, suggested to us that among all the alternative first exons of the 5'UTRs of CDKL5, the study of the Ex1_202 could be an interesting field to explore, in addition to the RefSeq Ex1_205.

In order to include the Ex1_202 in the sequences of interest for further experiments, the first step we made was to obtain first-hand evidence of the existence of Ex_202 in biological samples. We have already evaluated the sequence conservation of the 25nts upstream the putative TSS of 205 5'UTR, since it was reported as a possible alternative first exon in Ensembl (**Fig. 5.4.4.3**), but the lack of evidence of its existence in experimental works and the confusing information about the origin of the Ensembl record required a further first-hand experimental verification.

We performed BLAST analysis in the Expressed Sequence Tags database (dbEST) to verify if the 25nts upstream sequence of Ex1_202 has ever been found among the sequences read from mRNA library experimentally obtained from biological samples. We submitted the complete Ex1_202 as input sequence, as reported from Ensembl, to be sure that the length of the input sequence was long enough to avoid false positive identifications. As a result, we found that BLAST was able to find 26 sequences in human dbEST that aligned with the input sequence. However, only two of these presented 100% identity with the 25 upstream nucleotides of interest:

- HY013706.1, from RIKEN full-length enriched human cDNA library(unpublished), organ: testis.
- BF679133.1, from the NIH-MGC EST Sequencing Project (unpublished), organ: prostate.

By analyzing the returned sequences of these records, it was possible to confirm without any doubt their belonging to CDKL5 TV 205 transcripts, since the tags contains the downstream sequence belonging to RefSeq CDKL TV 205 5'UTR.

We concluded that Ex1_202 exists in at least two human tissues and was sequenced with success, even though it has been not associated officially to CDKL5 transcripts in

any publication.

Seeking further evidence of the presence of additional nucleotides upstream the known sequence of the CDKL5 mRNA, we performed the same BLAST in dbEST for all the species available. We found that the sequence under analysis has also been found in a cDNA library obtained from a visual cortex sample from mouse and in a sample obtained from the testis of the common marmoset – a primate often used as experimental model:

- BY281423.1, from RIKEN full-length enriched, visual cortex Mus musculus (Okazaki et al., 2002)
- HX595850.1, from full-length enriched common marmoset testis cDNA library (Tatsumoto et al., 2013)

Interestingly, we noted that the sequence from the common marmoset exceed the 5' limit of the 25nts input sequence, showing how the cDNA fragment that contains the 25nts of the Ex1_202 - and the rest of the downstream sequence – has an upstream region belonging to a longer mRNA of CDKL5, with a 5'UTR that – based on this result – seems to be longer than the variants previously reported for *Homo sapiens*.

To explore whether the Ex1_202 is itself a shorter form of a longer alternative first exon of CDKL5 still unknown in human, and to collect the first experimental evidence of the existence of Ex1_202, we performed a RT-PCR experiment with the following experimental design. We designed a common reverse primer on the UTex2 – that we called Rev - and three different forward primers on the sequence upstream to the TV 205 TSS. The first forward primer – called Top1 – was set on the 25 nts of Ex1_202; a second forward primer – Top2 – was set approximately 10nts upstream the Ex1_202 TSS; the last – Top3 - was approximately 20 nts upstream Top2 (**Fig. 5.5.1 A**). The experiment was performed on RNA samples obtained from the SHSY-5Y cell line, to have a confirmation of the presence of the sequence of our interest in human transcriptome. Moreover, it was repeated on RNA samples collected from mouse cortex, in order to confirm the result in an actual brain model. Amplification products were visualized on 2% agarose gel.

While we expect the amplification product from the Top1-Rev couple of primers, evidence of additional amplification products with the other primers could suggest that Ex1_202 is longer than previously reported remapping the possible TSS of CDKL5. A not retrotranscribed sample was used as control for the DNA contamination of the

samples. GAPDH primers were used as internal control.

In cDNA obtained from the retro transcription of the SHSY-5Y RNA, Top1-Rev amplicon was detected, as expected. We also detected an amplification product of the Top2-Rev couple of primers, while the Top3-Rev combination did not produce any amplification. The experiment was repeated three times (**Fig. 5.5.1 B**) and the same result was confirmed in cDNA obtained from RNA extracted from three distinct p30 mouse cortex (**Fig. 5.5.1 C**).

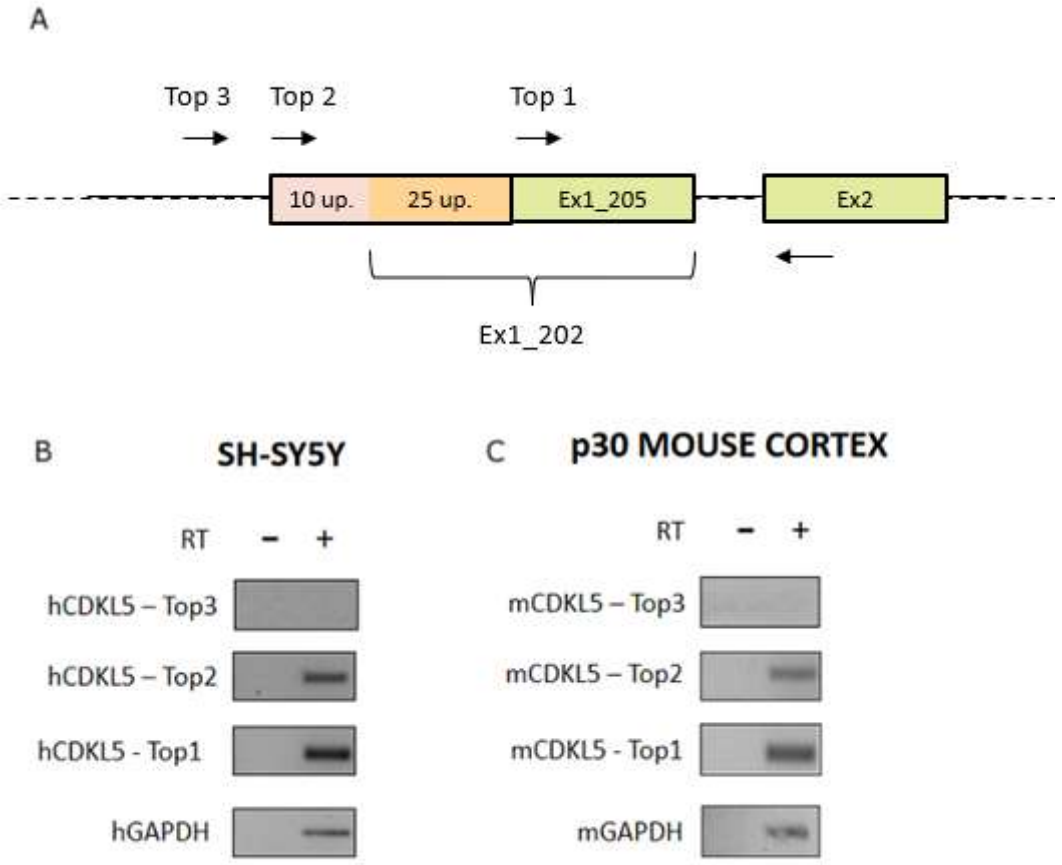


Fig. 5.5.1 RT-PCR analysis assessing Ex1_202. *(A) Experimental design diagram in which the positions of the used primers were schematically reported. Three couple of primers were employed in the analysis, having the reverse primer on the UTex2 in common. Top1 was set on*

Ex1_202, Top2 was set 10nts upstream the Ex1_202 TSS, and Top3 was set 20 nts upstream Top2. (B) Agarose gel electrophoresis of the PCR products obtained using the couples of primers described on SHSY-5Y cDNA. As expected, Top1-Rev was visible. Top2-Rev amplification product confirmed the presence of the upstream 10nts from the Ex1_202 TSS in the transcript. Not retrotranscribed sample was used as control of the absence of DNA contamination in the RNA used to produce the cDNA. GAPDH was a technical control to assess the good quality of the technique. (B) Agarose gel electrophoresis of RT-PCR products obtained starting for RNA extracted from SHSY-5Y cells (C) The results were confirmed by repeating the experiment also on p30 mouse cortex samples.

Our analysis confirms the presence of the Ex1_202, reported in Ensembl, in the SHSY-5Y human cell line and in p30 murine cortex. However, we found that a sequence of 10nts upstream the reported Ex1_202 is also transcribed in CDKL5 mRNA. The presence of the additional nucleotides could mean that either the TSS reported for TV_202 is wrong, and thus the Ex1_202 would be a truncated version -as Ex1_205 could be a truncated version of Ex1_202 - or there is another, previously unreported, CDKL5 TSS belonging to a new alternative first exon. We decided to name Ex1_Elongated202 (Ex1_202up) the sequence derived from Ex1_202 with the additional upstream 10nts suggested by our experiments.

5.6 Structural Prediction of CDKL5 5'UTRs

5.6.1 Watson & Crick Structural Prediction of 5'UTR Variants

To evaluate the possible presence of structural elements relevant for translation initiation, we performed a bioinformatics analysis that predicts the folding of RNA sequences. We performed a structural analysis using RNAfold (Gruber et al., 2008), which uses a single stranded nucleic acid as input for the prediction of Watson&Crick (WC) secondary structures. We decided to analyze all the CDKL5 5'UTR variants composed by all the alternative first exons reported in **Tab. 5.4.1.** We considered the 5'UTR of TV 205 – and thus TV 202 – as the most promising sequence in this kind of analysis, according to the GC Content analysis (**par. 5.4.2**) and, above all, its evolutionary conservation (**par. 5.4.4**)

For each analyzed sequence, we obtained the following results from RNAfold:
Minimum Free Energy (MFE) and Centroid (Ce) secondary structures and their ΔGs. The MFE method is the most popular structure prediction algorithm, which allows to fold the query sequence in the most energetically stable manner following the principle of

minimum free energy (Gruber at al., 2008; Gardner et al., 2004; Sato et al., 2009). The centroid structure is the structure with the minimal average distance to all the other structures obtained from theuse of the partition function algorithm (pf, McCaskill 1990). Pf algorithm usedby RNAfold allows to obtain an ensemble of secondary structures with different probabilities of various sub-structures. As such, the Ce structure is thebest representation of this ensemble of structures (Ding et al., 2005). The degree of reliability of the structural prediction of the input sequence is given by the similarity between the MFE and the Ce structures. This indicates that the MFE structure is not a predicted unrealistic structural optimization of the sequence analyzed, thus strengthening its reliability by not being at an extremeof the ensemble of the predicted structures (Gruber et al., 2008). The similarities and the discrepancies between the two structures can be evaluatedin a more schematic manner through the Mountain Plot of the folding.

- Ensemble diversity (ED, pf algorithm), which is the average base-pair distancein the ensemble of predicted structures. If ED is high, it represents the wide diversity of the possible folding, which is interpreted as a minor reliability of the predicted MFE structure that is always included in the ensemble. On the other hand, if ED is low, it represents the similarity of all the possible folding of the structure that is associated with a stronger prediction reliability (Gruberet al., 2008).
- The positional entropy of the folded sequence, which provides a local (per base) measure of the predicted folded MFE structure, allows to estimate which region of the structure is more reliable and which is, on the contrary, less probable. Positional entropy can be visualized as a plot of entropy versus position and through color annotation on the folded structure. Here we showedboth the representation.

  ΔG_{MFE}, ΔG_{Ce} and ED for all the 5'UTR variants are reported in the **Tab. 5.6.1.1.**

ID	ΔG_{MFE} (kcal/mol)	ΔG_{Ce} (kcal/mol)	ED
202	-101.4	101.2	38.69
205/207/Hector_a	-91.6	-90.11	33.41
Hector_a1	-170.7	-141.59	146.08
209	-115.3	-79.05	106.55
Hector_b1	-99.2	-69.7	120.3
211	-93.4	-68.8	100.08
Hector_b	-55.1	-55.1	17.98
201	-88.8	-86	56.17
208_long	-148.9	-113.3	144.9
204_short:Hector_cd	-74.5	-71.4	37.91
203	-68.4	-57.9	50.78
208_short/Hector_e1	-77.5	-54.1	88.73
Hector_e	-71.1	-60.9	72.88
204_long	-94.2	-91.6	44.78

Tab. 5.6.1.1 ΔG_{MFE}, ΔG_{Ce} and Ensemble Diversity of CDKL5 5'UTRs*. The table reports the ΔG_{MFE}, ΔG_{Ce} and ED (Ensemble Diversity) for all the 5'UTR variants analyzed through RNAfold. ΔG_{MFE} and ΔG_{Ce} were expressed in kcal/mol. Sequences that show a difference of maximum 5 kcal/mol between ΔG_{MFE} and ΔG_{Ce}, and an ED not exceeding the value of 50 were considered to have a reliable MFE structural prediction. ID = identification code.*

There is no threshold to evaluate the absolute values of the parameters resulted from RNAfold. Therefore, in accordance with the principles mentioned before, we considered reliable the MFE predictions of the following 5'UTRs: 202, 205, Hector_b and 204 short (RefSeq NM_001323289.2) and long variant. These sequences show a ED lower than 50 and the ΔG_{MFE} value which is very near to the ΔG_{Ce} – often ΔGs presents the same values. Moreover, one of the parameters to evaluate the stability of a structure is how negative the ΔG_{MFE} is. However, deciding when the ΔG_{MFE} could be considered significantly negative is a subjective matte. This represents a difficult problem of interpretation, since returned ΔG_{MFE} is a parameter influenced by the length of the RNA sequence analyzed and, therefore, the absolute value is not a valid option to compare sequences with different length . For this reason, it is necessary to normalize the resulted ΔG_{MFE} to obtain a useful information from this type of prediction.

One of the first ΔG_{MFE} normalization, AMFE (Trotta, 2014), takes into consideration the impact of the length of the RNA molecule dividing the ΔG_{MFE} by the number of nucleotides. However, the length normalized AMFE index tends to overestimate the significance of returned ΔG_{MFE}, resulting in false positives (Trotta 2014). Thus, in order to give a useful meaning to the numerical data obtained by RNAfold, we decided to

employ the normalization of the ΔG_{MFE} by sequence length formulated by Trotta (Trotta 2014). We considered this method to be reliable because it is the only one taking into consideration the fact that the ΔG_{MFE} variation regarding the sequence length is not perfectly linear and needs some corrections and it has already proven to be valid in different population of human RNAs (Trotta, 2014). The length normalized MFE index, called MFEden, introduced by Trotta, is calculated from the ΔG_{MFE} of the sequence – using RNAfold – and obtained following the present function, here reported:

$$MFEden = 100 * (MFE - MFE_{Ref}^{L})/(L - L_0)$$

where L is the sequence length expressed as number of nucleotides; is the MFE_{Ref}^{L} average ΔG_{MFE} calculated by Trotta starting from an ensemble of 100 random sequences (with an equimolar ratio of the four basis) with the same length of the sequence of interest; L_0 is a length constant of 8 nucleotides (empirically introduced by Trotta to optimize the calculation). Once MFEden is calculated, it is possible to compare it with a reference plot built by Trotta, in which MFEden of group of sequences randomly taken from the human genomes are reported, taking in consideration the GC content of the sequence. If the MFEden of the analyzed sequence is more negative than the values reported in the mentioned plot, the ΔG_{MFE} from which is calculated is considered significantly negative, and thus, the prediction is assumed reliable (**Fig. 5.6.1.2**).

As a result, it appears that only the MFEden of the 5'UTRs of the variant 205 – and thus 202 – is consistently more negative than the reference index MFE_{Ref}^{L} reported on the graph, showing a value of -9 (Tab. 5.6.1.3) versus its reference of 0 for GC% = 50. Also the other sequences, selected by using ED and the similarities between ΔG_{MFE} and $\Delta G_{Ce,}$ show a significantly low MFEden index. However, the MFEden of 205/202 is the more negative one and the more distant from the reference MFEden (as reported in the **Tab. 5.6.1.3**.).

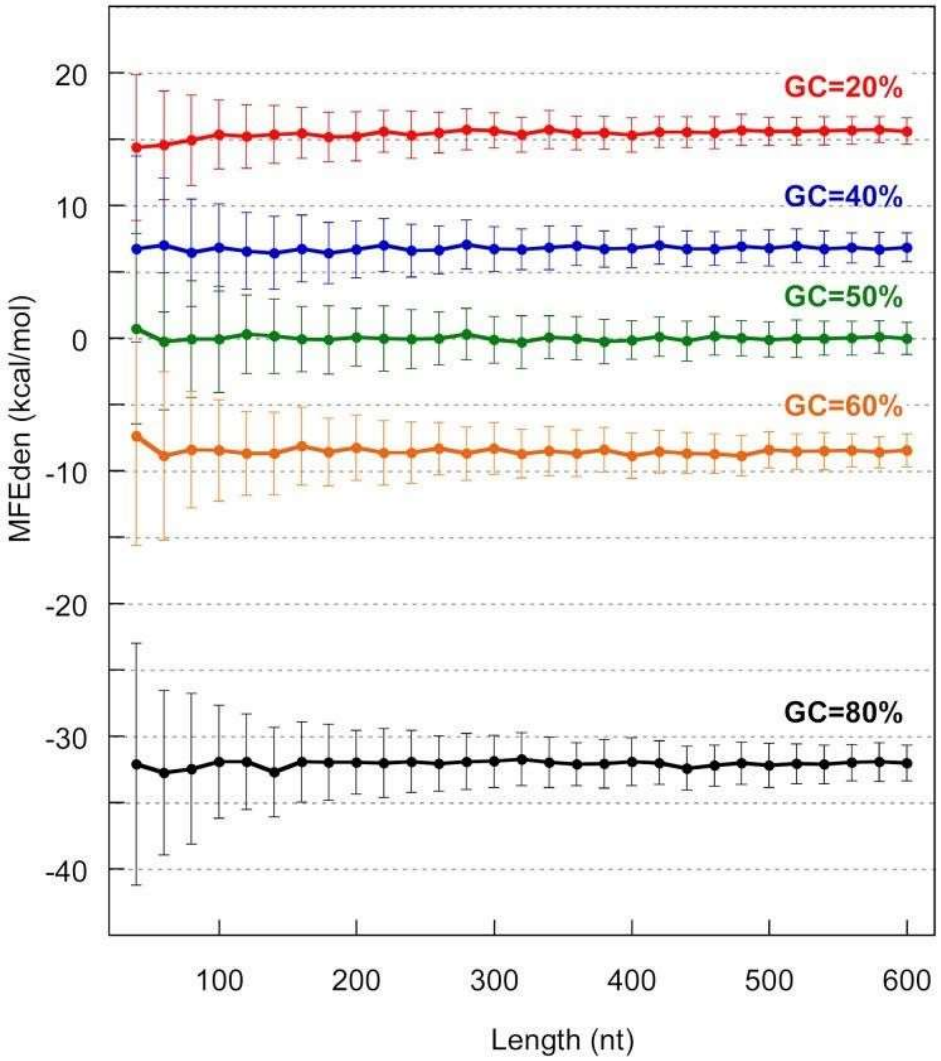


Fig. 5.6.1.2 Reference MFEden versus length plot. *The plot shows the mean of MFEden values obtained from 1000 shuffled sequences with various GC Content (20%, 40%, 50%, 60% and 80%) and with various length. Each point corresponds to the mean MFEden of 100 sequences and error bars indicate standard deviation (n=100). The reference plot was built by Trotta and used in this analysis to properly interpretate the* ΔG_{MFE} *of CDKL5 5'UTRs considered to be reliable (Trotta 2014).*

ID	Length	ΔG_{MFE} (kcal/mol)	MFE^{L}_{Ref}	MFEden	GC%
202	275	-101.4	-77.3888	-8.99	50.9
205/207/Hector_a	250	-91.6	-69.631	-9.08	49.2
Hector_a1	462	-170.7	-136.301	-7.58	58
209	366	-115.3	-105.8965	-2.63	51.1
Hector_b1	391	-99.2	-113.5883	3.76	37.1
211	369	-93.4	-106.787	3.71	37.1
Hector_b	206	-55.1	-55.8605	0.38	39.3
201	285	-88.8	-80.5848	-2.97	46.7
208_long	514	-148.9	-152.6685	0.74	45.7
204_short:Hector_cd	249	-74.5	-69.3105	-2.15	45.8
203	278	-68.4	-78.4295	3.71	41.7
208_short/Hector_e1	303	-77.5	-85.9813	2.88	42.6
Hector_e	283	-71.1	-80.0445	3.25	41
204_long	301	-94.2	-85.4578	-2.98	45.8

Tab. 5.6.1.3 MFEden index of CDKL5 5'UTRs*. The table reports length, ΔG_{MFE} obtained from RNAfold, the reference ΔG_{MFE} for the sequences of the same length of the sequence of interest and the MFEden index for each CDKL5 5'UTR variants. MFEden was calculated following the indication reported in the paper of Trotta, 2014. GC% was reported to allow to comparisonof the calculated MFEden with the reference MFEden plot.*

To confirm the results obtained by the Trotta normalization we decided to perform an additional bioinformatic analysis. We used a Bioconductor package, XNAString (Górska et al., 2022), on RStudio. The aim was to compare the ΔG_{MFE} of the 5'UTR of CDKL5 with the ΔG_{MFE} of a set of sequences generated from the original one with the first exon, with the same length and nucleotide composition, modified in a shuffled, random manner. The UTex2, instead, was maintained unvaried. We evaluated the frequency of the cases in which partially random sequences return a ΔG_{MFE} that is more negative than the original 5'UTR. Therefore, variations in the resulted ΔG_{MFE} were caused uniquely by the position of the nucleotides. If the original 5'UTR ΔG_{MFE} ranked negative within 2 sigma from the mean of 1000 partially random 5'UTRs, we considered it significant (p-value < 0.05) and, thus, the structural prediction worth of consideration. By this approach we confirmed the results obtained by the Trotta method and the importance of the ΔG_{MFE} of the CDKL5_205 5'UTR – and thus 202 - in the context of structural predictions (**Fig. 5.6.1.4**).

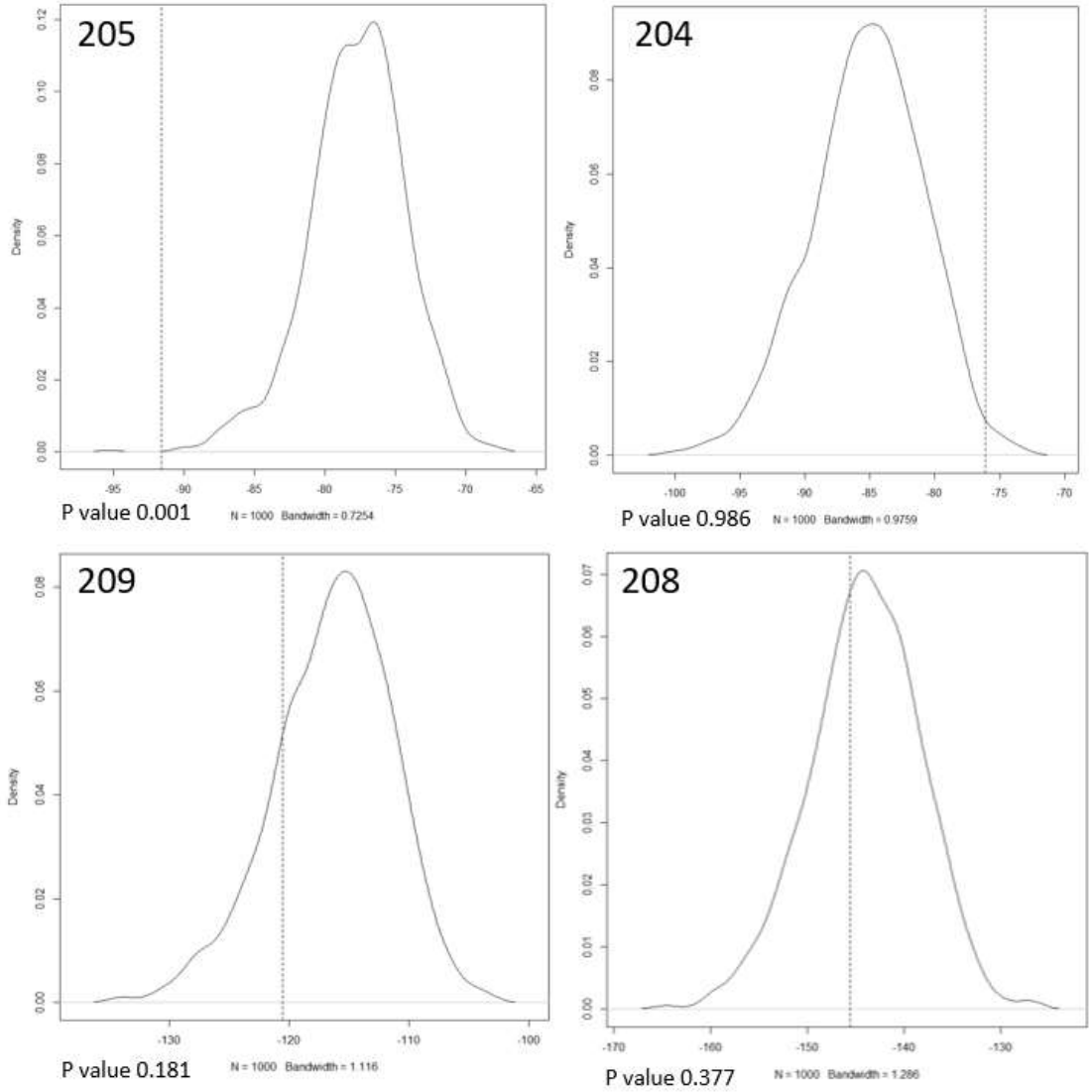
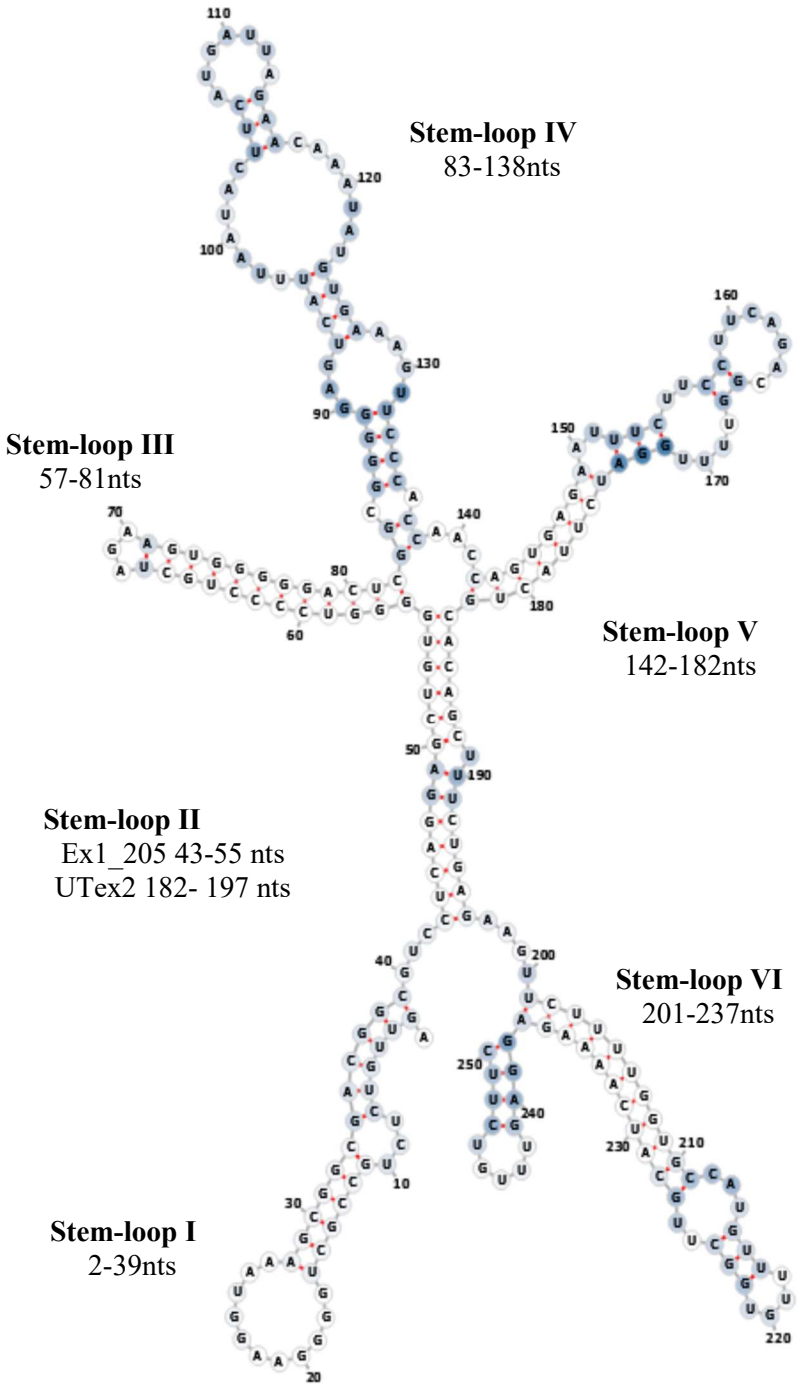


Fig. 5.6.1.4 MFE Distribution Plots. *Examples of MFE distribution (x axis) of 1000 partially random 5'UTRs obtained shuffling the nucleotides of the first exon and maintaining unvaried the UTex2 sequence. In each plot the position of the MFE of the original CDKL5 5'UTR is highlighted (vertical bar) . P-value was calculated as the frequency of the observation in which partially random UTRs return a more negative MFE than the original one. The 205,204, 208 and 209 CDKL5 5'UTR variants are shown. Of note the 205 5'UTR is the one to rank with negative values outside the distribution (p-value 0.001) of the partially random 5'UTRs. This confirmed that the 205 the 205 MFE cam can be considered significantly negative.*

The MFE structure of 205 5'UTR shows a stable folding, according to the color indication of positional entropy displayed on the structure. The base-paired structure is strictly folded, excluding the presence of unpaired region in the sequence. We reported the presence of six different stem loops (canonical helices), involving both Ex1_205 and UTex2. The two helices in Ex1_205 showed a very low positional entropy,

resulting in a very stable Ex1_205 folding landscape. On the other hand, UTex2 is showed to be less stable in comparison of the folding of EX1_205, that stabilize the level of positional entropy by forming a stem-loop with the UTex2 region between 183-196nts. The last four predicted stem-loops in UTex2 have higher positional entropy level - in particular stem-loop III - as indicated in the colored structure and in the positional entropy plot (**Fig. 5.6.1.5**).

A



B

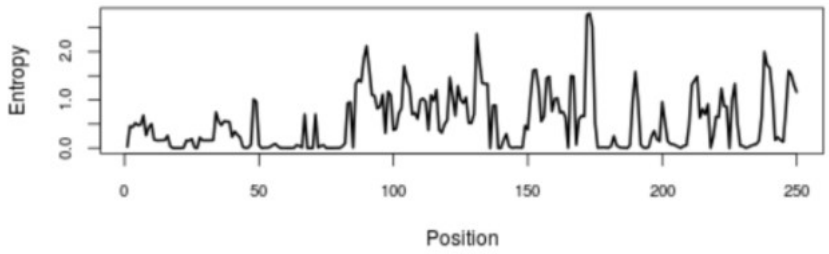
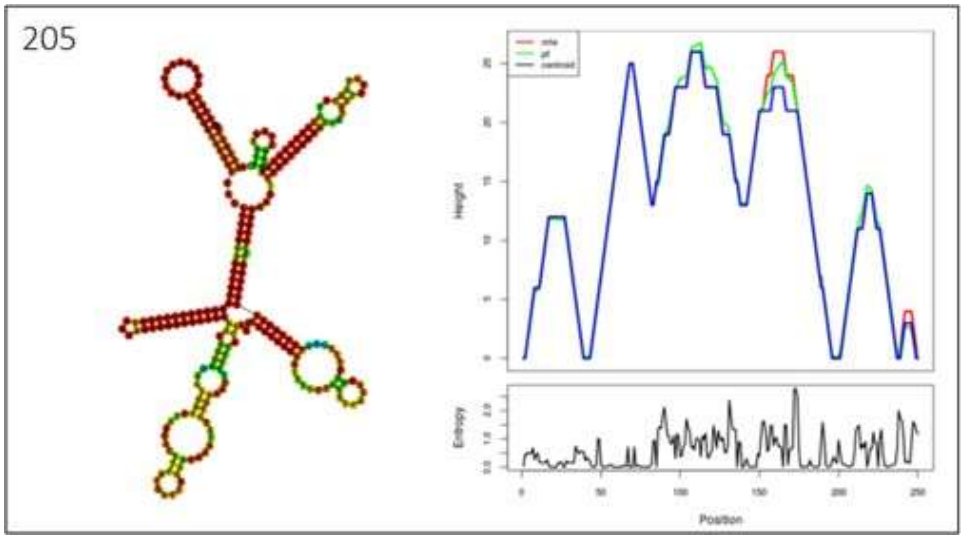


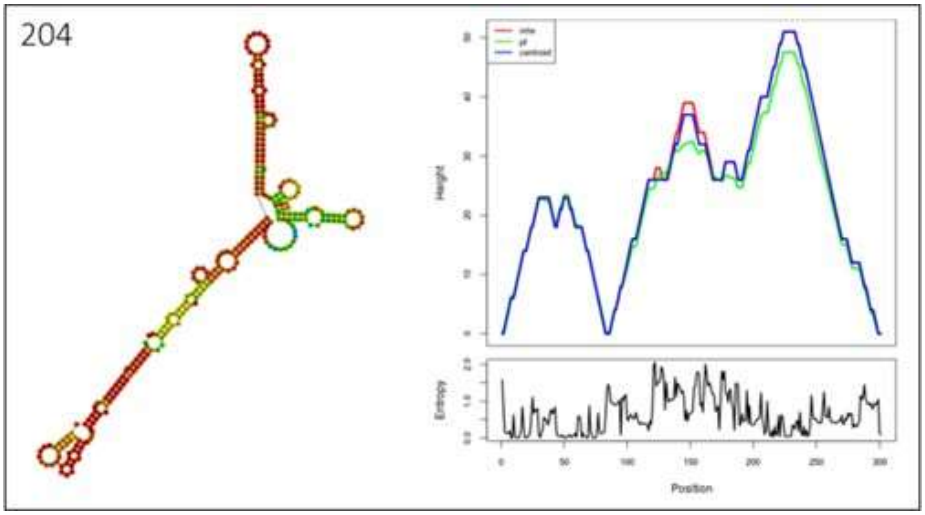
Fig. 5.6.1.4 CDKL5 205 5'UTR MFE structure*. (A) The structure shows a coloration that indicates the positional entropy calculated by RNAfold for each base-pair (white=low entropy; blue= high entropy). Stem-loops found in the structure were named based on its position, as I, II, III, IV and V, with the indication of their positions (nts). (B) Plot of positional entropy versus position obtained with RNAfold. The 88nts of Ex1_205 have a very low positional entropy, contributing to the stable folding of the 5'UTR.*

The contribution of Ex1_205 in the folding of the structure and in setting the ΔG_{MFE} is confirmed by comparing its contribution in the global 5'UTR folding with those brought by other alternative first exons, which do not strongly have interactions with the UTex2 sequence, resulting in a more disordered and unstable 5'UTR structures, such as the ones in which Ex1_208 and Ex1_211 are present, reported here as an example (**Fig. 5.6.1.6**).

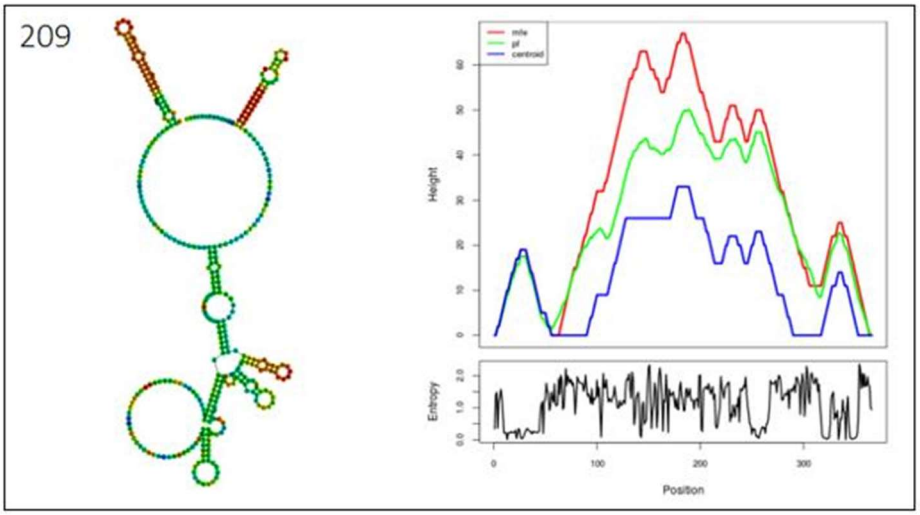
A



B



C



D

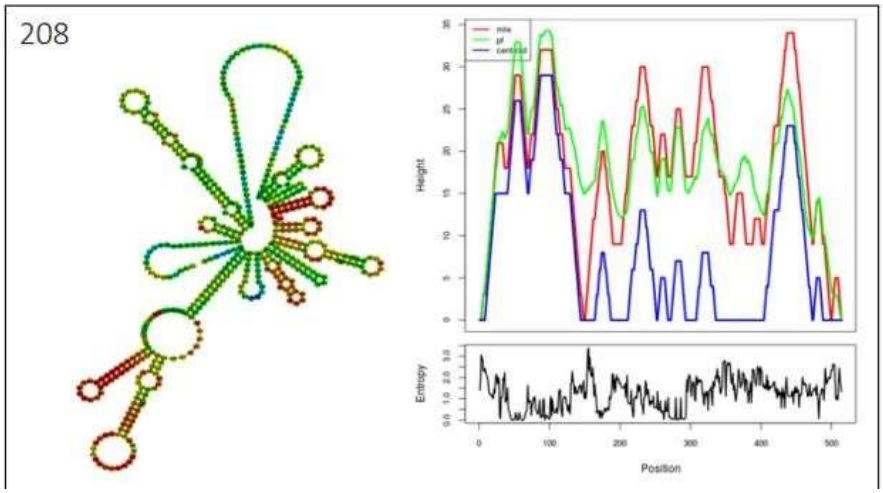


Fig. 5.6.1.6 Structure Representation, Mountain plot and Positional Entropy Plot of 205, 204, 208, 209 5'UTR variants. *Comparison between the mountain plots and the entropy plots of the CDKL5 5'UTRs of the TV 205, 204_short, 208 and 209. (A) 205 5'UTR has the most reliable structural prediction, taking into consideration the similarity of the MFE, Ce and pf structures*

in Mountain plot and the lower positional entropy values – especially in the region of Ex1. (B) 204_short has shown higher level of positional entropy than 205, even if the reliability of the predicted structure can be considered good enough and the entropy level is moderate low. The main difference is due to the absence of a strong structural motifs between the Ex1 and the UTex2, such as the stem-loop II, that lower the positional entropy of UTex2. (C and D) 5'UTRs of 208 and 209 are reported as examples of low reliable structural prediction. Mountain plots show the greatest discrepancy between the model structures and the positional entropy plot of their MFE structures show the highest peaks in comparison to the reported ones for 205 and 204. The higher positional entropy profiles are due to the presence of wide unpaired region in the sequences. All the data was obtained through RNAfold. For each variant it is present a structural representation based on MFE structure, colored in accordance with the positional entropy value of each base-pair.

The comparison of TV 205 5'UTR with the TV 202 5'UTR, differing from the first for 25nts upstream Ex1_205, shows how the presence of the few additional nucleotides can profoundly disturb the folding of the very stable stem-loop I, increasing the positional entropy at the beginning of the sequence (**Fig. 5.6.1.7**).

A possible reason of the weak WC base-pairing and high positional entropy in WC folded structure is the presence of an alternative and competitive base-pairing, such as the Hoogsteen base-pairing involved in the folding of G4 structural motifs (Takahashi et al., 2021). Due to the high G content of the upstream 25nts of Ex1_202 - equal to more than the 50% - we also took in consideration also the G4 prediction as a possible tool to understand the possible function within the Ex1_202.

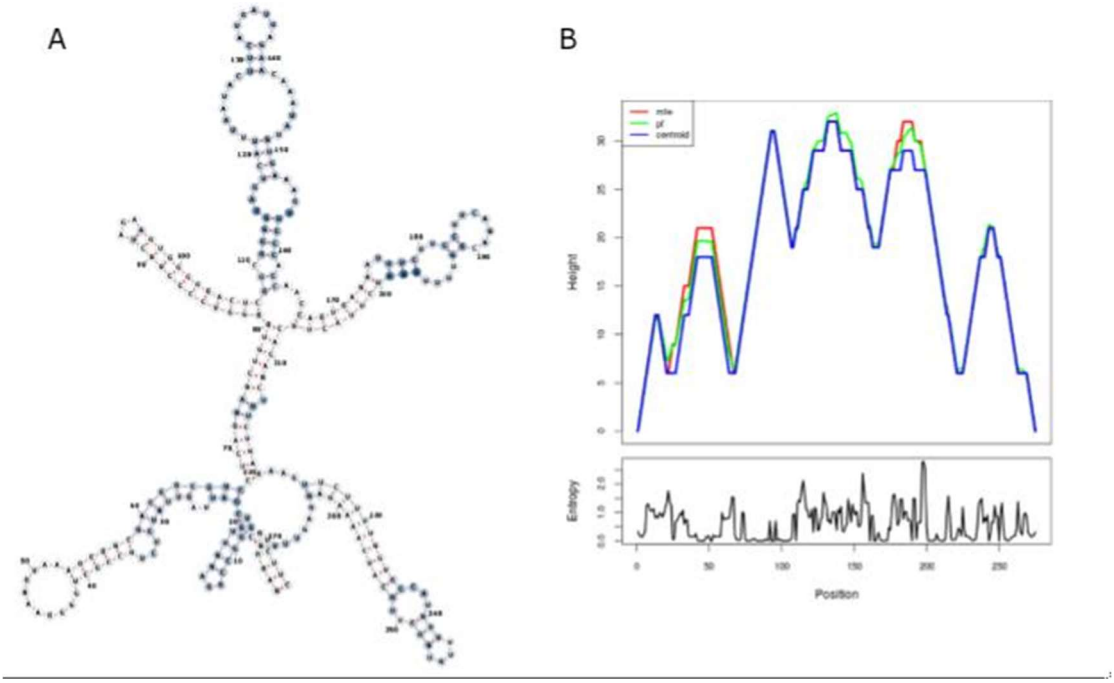


***Fig. 5.6.1.7 MFE structure of CDKL5 202 5'UTR**. (A) The structure shows a coloration that indicates the positional entropy calculated by RNAfold for each base-pair (white=low entropy; blue= high entropy). It is notable how the presence of the 25nts at the beginning of the sequence disturbs the stem-loop I's fold. (B) Plot of positional entropy versus position obtained with RNAfold. First Exon show a lower-level positional entropy in comparison to the UTex2 and difference in the common sequence with Ex1_205. Moreover, Mountain plot is reported.*

5.6.2 GQuadruplex Prediction in Ex1_205 and Ex1_202 sequences

We performed G4 analysis using the QGRS Mapper software, which generates predictions on the composition and the distribution of putative quadruplex forming G-rich sequences (QGRS) in nucleotide sequences (Kikin et al., 2006).

For each predicted G4, QGRS Mapper returns a G-score that depends on the length of the G-rich sequence, with features compatible with a possible G4 motif. Initially, we decided to not consider the G-score as a parameter to exclude predicted G4s due to a lack of a real threshold in interpreting it. As mentioned before, we analyzed the sequence of the 205 and 202 5'UTRs. QGRS Mapper predicted two G4 on Ex1_205 and an additional G4 in 25nt (13nts of which are G) of the Ex1_202 (**Fig. 5.6.2.1**). We named $G4_1$, $G4_2$ and $G4_3$ the three G-rich motifs starting from the 5' end and noted their G-score.

B

Position	Length (nts)	QGRS	G-Score
3 - 14	12	GGGCGGGGCCGG	19
42 - 63	22	GGGGAAGGTAAAGCGGCGACGG	19
99 - 115	17	GGGGGACTCGGCGGGGG	18

Fig. 5.6.2.1 Predicted G4s in Ex1_202 and Ex1_205. *Schematic representation of the positions of the predicted G4s in Ex1_202 and Ex1_205. The prediction was performed using QGRS Mapper and the position, length and G-score are reported for each G4 found. Figure was built through Benchling (https://benchling.com).*

Thus, we assumed that a difference between Ex1_205 and Ex1_202 might reside in the presence of an additional G4. This could strongly affect the modulation of the translation efficiency of the transcript since it has been suggested that the presence of closely spaced G4s can exponentially increase their effects. In fact, it has been reported how adjacent G4s can interact in modulating gene expression in a complex manner in HRAS and c-KIT promoters (Cohoi et al., 2014; Ducani et al., 2019). Moreover, the effect of two adjacent G4s in the 5'UTR of TGFβ2 in increasing, in a "combinatorial" manner, the translational efficiency of the transcript was recently proven (Agarwala et al., 2019). The G-scores of the G4s presented in these studies are comparable with the ones reported for the predicted $G4_1$, $G4_2$ and $G4_3$ in the CDKL5 transcript. Furthermore, even if QGRS Mapper cannot distinguish between DNA and RNA when calculating G-scores, RNA G4s are assumed to be biologically more stable than DNA G4s (Agarwala et al., 2015).

To evaluate the reliability of the G4 prediction, we checked the positional entropy plot obtained through RNAfold for Ex1_202 and Ex1_205. The predicted $G4_1$ motif is placed in the high entropy region of the 25nts, and, interestingly, $G4_2$ is placed nearby the second high peak of positional entropy in Ex1_202. It is worth of notice that the mentioned peak in the positional entropy plot is present in Ex1_202 where its intensity is notable, whereas in Ex1_205 is more restricted both in amplitude and heigh. Thus, it appears that the presence of $G4_1$ not only affects the WC base-pairing, but also favors

the folding of $G4_2$, strengthening the idea of an interplay between the two G4s in CDKL5 Ex1_202. In addition, the predicted $G4_3$ sequence also lies in a high positional entropy region of Ex1_205/202, even though, based on the prediction, it seems less relevant compared with $G4_1$ and $G4_2$.

Another approach to strengthen the reliability of G4s prediction is to evaluate the local conservation of the sequence in which the structural motifs where predicted. To achieve this aim, we employed Consurf (Ashkenazy et al., 2010). Consurf allows to calculate the conservation of each position of the input sequence from a sequence alignment built with ClustalX. The conservation of each base is expressed by a numerical value called color score that ranges from 1 to 9. A color score of 1 means that the nucleotide at the position is very variable and extremely less conserved in the given alignment, while a color score of 9 means that the nucleotide at the position is very conserved and mainly shows an identity in the alignment.

We used the Ex1_202up sequence as query sequence and provided Consurf with an alignment of 66 ortholog sequences obtained from BLAST analysis (E=0.1, to increase the number of sequences and thus help the tool in computing a more accurate calculation) performed with Clustal X. Color scores obtained for all the 275nts were moved to Prism and used to build a positional conservation plot, in order to assess the conservation of the predicted G4s motifs. From the plot it emerges that the regions in which $G4_1$ and $G4_2$ are placed showed peaks in color score - in which the major contribution is given by G positions (**Fig. 5.6.2.2**), further strengthening the hypothesis that functional G4s are indeed present.

Therefore, the results obtained from the G4 prediction through GQRS Mapper are supported by both positional entropy and positional conservation analysis, suggesting that a functional difference between Ex1_205 and Ex1_202 could be due to the additional G4 motif.

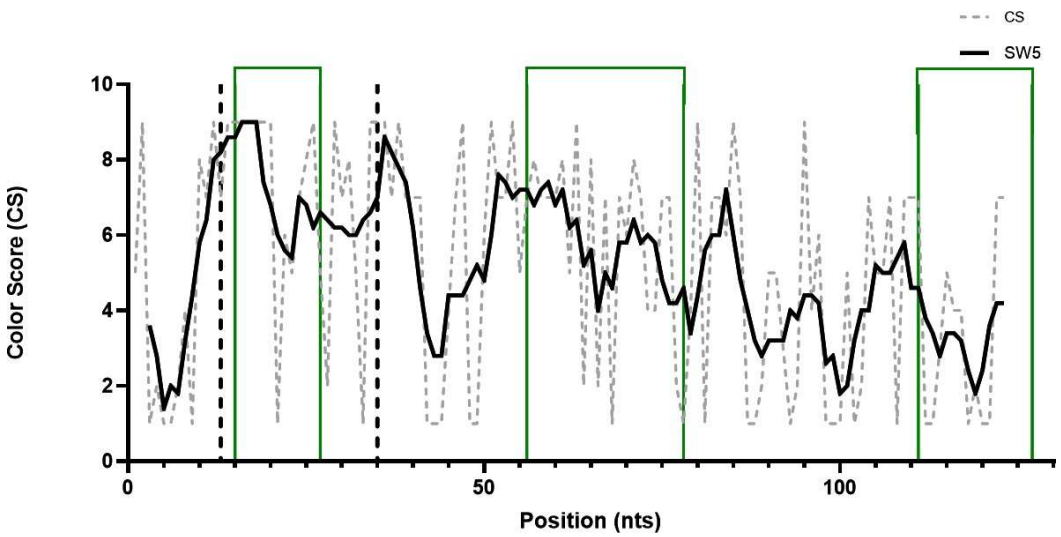


Fig. 5.6.2.2 Positional conservation plot of Ex1_202up and Ex1_205. *Color scores (CSs) – obtained from Consurf – are reported for each position, based on the alignment of 66 ortholog sequences found by nBLAST search (E=0.1). Sequence conservation is visualized as sliding window mean of 5 positions. From the plot it emerges how the peak of conservation in Ex1 are in proximity of the position of predicted $G4_1$ and $G4_2$, and how the 35nts upstream the TSS of Ex1_205 particularly conserved – the 25nts of the beginning of Ex1_202 and the additional 10nts upstream the Ex1_202 TSS both.*

5.6.3 Circular Dichroism for assessing $G4_1$ in 202up

We decided to use Circular Dichroism (CD) spectroscopy to validate the prediction of $G4_1$, the G-quadruplex that, according to our analysis, could be highly relevant in determining the translational efficiency of the CDKL5 transcript. Traditionally considered the golden standard in assessing the capability of folding of predicted G4 (Randazzo et al., 2013) CD was employed to: 1) define the topology of the $G4_1$, considering the characteristic topology of RNA G4s (Simone at al., 2015); 2) assess the proper folding of G4 in response to different buffers that vary in salt composition.

We performed the analysis of the RNA oligonucleotide with the additional 35 nts using a Jasco J-100 spectropolarimeter. Buffer compositions were formulated taking into consideration the influence of different cations in G4 stability, which could be detected as a difference in peak amplitude of the spectra resulted from CD. Therefore, in accordance with the literature, we employed a:

- Salt-free Buffer (RNase-free water with 10 mM Tris/HCl, 0.1 mM EDTA, pH7.4) (Lammich et al., 2011).
- K^+ Buffer, containing 1mM of KCl, that is recognized to strengthen G4s folding,resulting

in higher positive peak at 260nm and lower negative peak around 240nm (Gray et al., 2019; Lammich et al., 2011; Fisette at el., 2012).

- Li^+ Buffer, containing 1mM of LiCl, that is reported to not cause significant change in CD spectra of G4 due to its inability in intercalating in G4 structure and, thus, in helping in stabilizing its folding (Fisette at el., 2012, Lammich et al., 2011):
- Na^+ Buffer, containing 1mM of NaCl. Na^+ cations are reported to be able to intercalate in G4s, but they have a weaker effect in stabilizing G4 folding than the K^+ salts, and, thus, they are considered negative control (Fisette at el., 2012,Lammich et al., 2011).

Each condition was analyzed through 9 consecutive measurements, used to obtain an averaged spectrum of the sample. The comparison between different spectra from different buffer-conditions was performed with SpectraGryph 1.2 (Menges 2016).

The results, shown in **Fig. 5.6.3**, confirm the proper folding of $G4_1$, as demonstrated by the positive and the negative peaks at the correct wavelength and the characteristic morphology of the spectrum. Moreover, the presence of K^+ in the buffer in which $G4_1$ was placed during the measurement results in the expected strengthening of positive and negative peaks. This effect was weakly present with other salts, demonstrating the proper G4 folding of the sequence in which the predicted $G4_1$ lies. Further confirmations should be performed by a mutagenesis approach on the RNA oligonucleotide, thought which guanines involved in $G4_1$ should be replaced with other nucleotides that disrupt the structure.

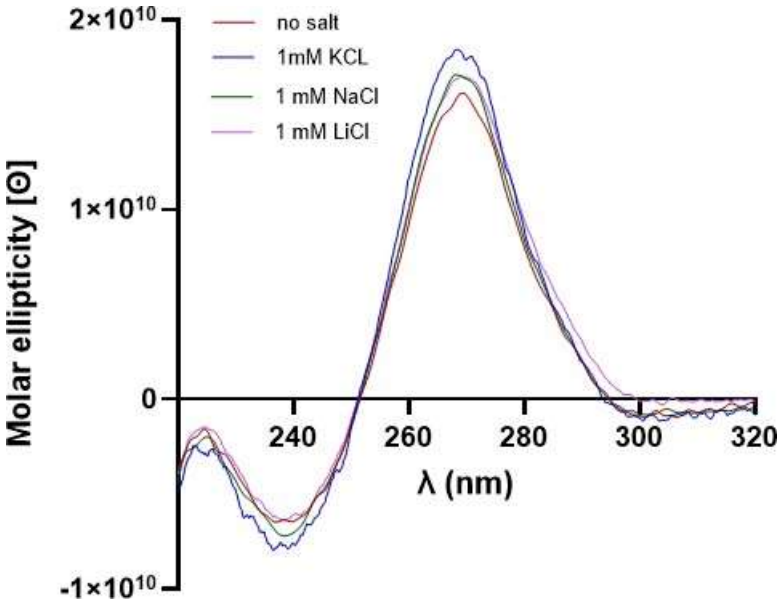


Fig. 5.6.3 ***$G4_1$ CD spectra.*** *CD spectra obtained from 5 μM of the oligonucleotide containing $G4_1$ (35 nts) in absence (red) or presence of the monovalent cations Li^+ (violet), Na^+ (green) and K^+ (blue). The basis of the buffers is RNase-free water with 10mM Tris/HCl, ap pH 7.4.*

The CD spectra return the characteristic topology of parallel G4s, with negative and positive peaks respectively at 239nm and 268 nm. The presence of K^+ strengthen the peaks amplitude as expected, while the presence of Na^+ and Li^+ show a minor influence on $G4_1$ spectra (salt concentration = 1 mM). Each spectrum is obtained averaging 6 CD measurements.

5.6.4 Evans Mutation in the context of Structural Predictions

In the context of the structural predictions, the position c.-189, in which the SNP of uncertain significance was found by Evans (Evans et al., 2005), appears to be involved in the folding of the stem-loop II in Ex1_205, at the position 61. Moreover, it should not affect any G4 sequences. The color score reported by the Consurf Analysis is 4, a value that is under the below the average of the color scores of Ex1_202 (that is 5.19). This value il still considered as not variable, even if it is indicative of a weakly conserved position in the considered alignment. We anyway tried to evaluate the potential effect of the transition C-T in the WC folding of 5'UTR 205 through RNAsnp (Sabarinathan et al., 2013), but the result of the analysis was that C61T is not predicted to cause relevant structural changes, as the ones reported for other predicted variation of single nucleotides inside the sequence – reported in the appendix **(par. 9.2**).

5.7 Quantification of alternative TSS usage of CDKL5 using CAGEr

One of the main problems in understanding the possible role of the CDKL5 5'UTRs,is the confusing and oftentimes contradictory presence of numerous first-exon sequences, either predicted or experimentally validated. Evidence of their actual existence or abundance is, so far, incomplete. This could be due to experimental limitations of the methodologies employed, often based upon amplification of transcript that might introduce a significant bias in the measurements. Moreover, the 5'RACE method, often used to identify the TSS and employed by Hector et al., 2016 todetermine the first exons of CDKL5's 5'UTR, can be limited by the high content of GCof the sequences, causing the loss of regions towards the 5' end (Hector et al., 2016). The high number of alternative first exons of CDKL5, reported in different sources as mentioned in par. 5.3, often shows inconsistencies regarding the compositions and confusing nomenclature. For these reasons, every prediction or result reached for one single 5'UTR variant of CDKL5 risks having an unclear interpretation of its relevance.In order to overcome this limitation, we performed a quantification of the abundance of every first exons of CDKL5 using a resource that is nowadays considered to be the most

precise source to analyze the real TSSs of mRNAs: Cap Analysis of Gene Expression (CAGE) libraries made available by the FANTOM 5 consortium (Noguchiet al., 2017).

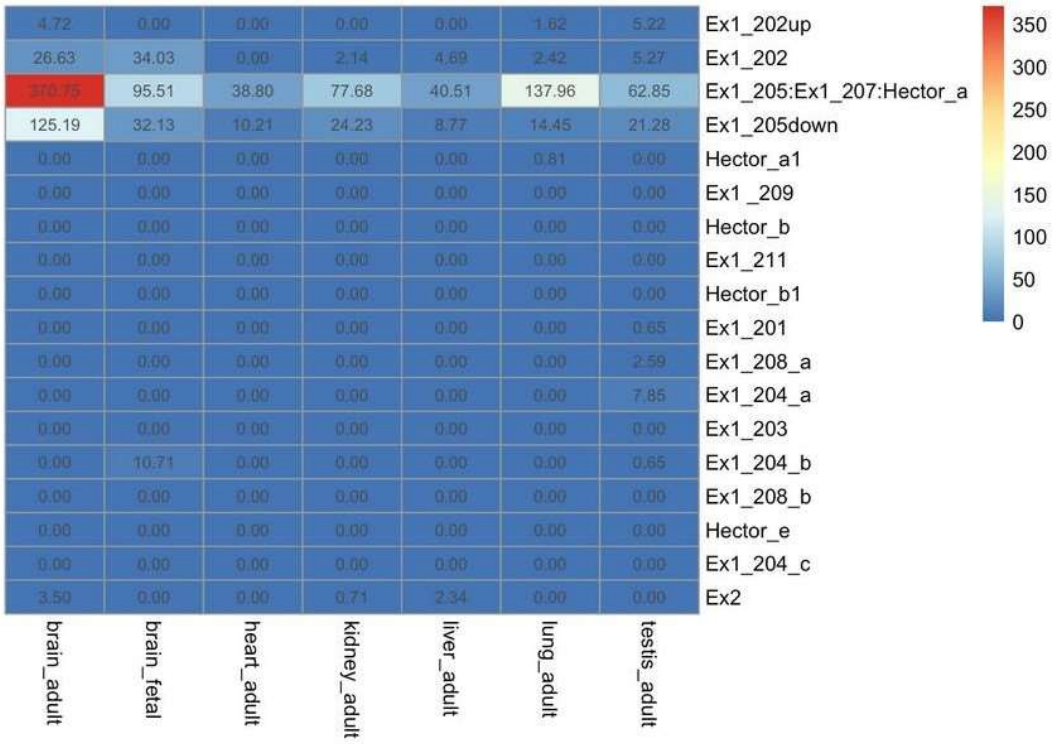
CAGE sequencing is a recent technique that allows to measure the expression of eukaryotic mRNA in a very precise manner. The greatest advantage is that it lacks the amplification step, avoiding bias in copy number estimation of target sequence. Quantification of gene expression is performed by counting the number of CAGE tags obtained from cap-trapping reaction based on biotinylation of the cap site of the RNA, efficiently retrotranscribed in cDNA and then sequenced (Takahashi et al., 2012; Takahashi et al., 2014).

CAGEr, a freely distributed R/Bioconductor package, was employed to quantify the relative frequency of TSS usage, based on the number of CAGE tags for each alternative TSS (Haberle et al., 2015). Input data were directly imported by the CAGE human dataset obtained from FANTOM 5 web resource (http://fantom.gsc.riken.jp/5/datafiles/latest/basic/human.tissue.hCAGE/). The selected tissues libraries for the analysis of CAGE sequencing data were the human adult tissues: brain, testis, lung, heart, liver, and kidney, as well as fetal brain, due to the importance of CDKL5 in brain development. The CAGEr workflow - accurately reported for our analysis in appendix, **par. 9.3** - allows the quantification of the abundance of TSSs used in the transcription of the gene. The accounting for alternative TSSs under the same shifting promoter obtained from this analysis, moreover, led to the identification of TSS clusters including TSSs normally not detectable trough amplification techniques, since they are masked by upstream TSS. Moreover, the package allows for quality filtering of the data and the remotion of G nucleotide addition bias, since during the CAGE experimental protocol often an addition G nucleotide is attached to the tags 5’ends (Harbers et al., 2005). To provide each individual TSS expression and to allow the comparison between the different tissue libraries, raw tag counts were normalized using a common referent power-law distribution (Balwierz et al., 2009).

A

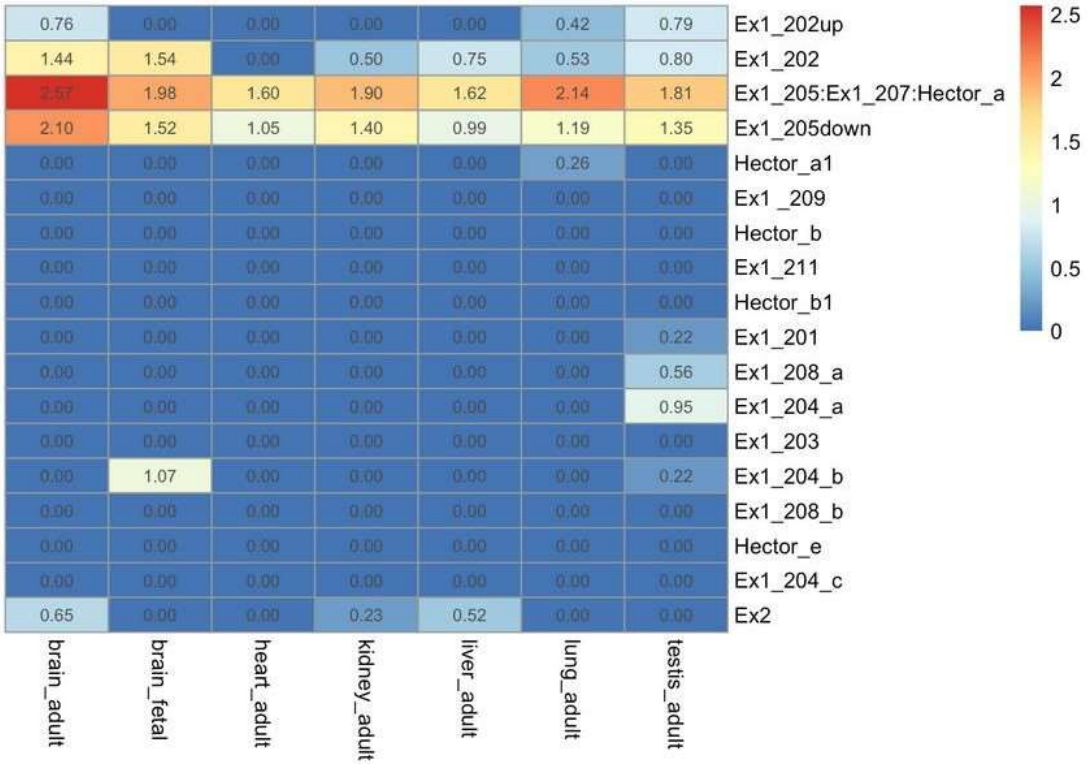
TSS usage in body tissues

abundance



B

log10 abundance



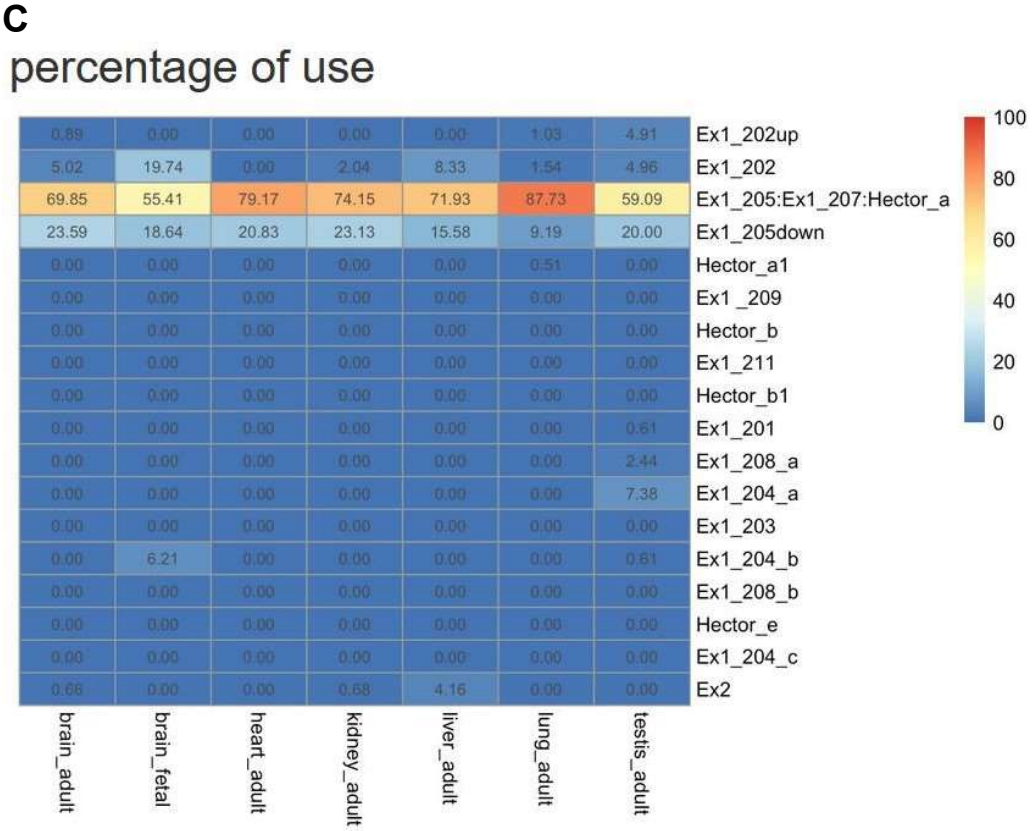


Fig. 5.7.1 TSS Abundance in Different Human Tissues*. (A) TSS abundance in different human tissues, expressed also as log10 abundance (B) and percentage of use (C) in adult brain,heart, kidney, liver, lung, and testis, and in fetal brain. Adult brain, adult lung, adult testis, andbrain fetal are the tissues in which CDKL5 is more expressed. The most abundant TSS belongsto Ex1_205 (RefSeq NM_001323289 and NM_003159.3), Ex1_205down, and Ex1_202, respectively. Other TSSs are weakly expressed and consequently the other alternative first exons are poorly represented in CDKL5 5'UTR. A TSS is detected also at the start of CDKL5 exon 2.*

The results of the analysis are shown in **Fig. 5.7.1**, where TSS usage is expressed as abundance (**A**), log10 abundance (**B**) and percentage of use (**C**). The analysis showed that most of the alternative first exons reported in either Ensembl or in the paper from Hector (Hector et al., 2016), are poorly expressed in adult tissues as well as in the fetal brain, confirming that their presence is detectable only with amplification procedures. Interestingly, even the TSS of Ex1_204a, relative to the 5'UTR of the RefSeq Transcript NM_001037343.2, is weakly expressed in most of the tissues except for testis, where it represents less than the 10% of the total TSSs used - for the rest represented by TSSs belonging to a cluster in the vicinity of the TSS of Ex1_205. This TSSs is, in fact, the main TSS from which the CDKL5 gene is transcribed, and thus, most of its transcripts contain Ex1_205 in their 5'UTR (**Fig. 5.7.1**). It is represented at

least in 70% in all the tissues considered, except for testis and fetal brain, where its percentage of use is around 50%. The other most used TSSs belong to the TSS cluster consisting in four TSSs - named with the name of the resulted exon: Ex1_202up, experimentally demonstrated to exist in **par. 5.5**; Ex1_202; Ex1_205; and the novel Ex1_205down, here identified for the first time (**Fig. 5.7.2**). While Ex1_202up is poorly expressed when compared to the other members of the cluster - achieving aboutthe 5% of usage only in testis - , the rest of the TSSs are the most used, following the trend of Ex1_205. Ex1_202 is the third most frequent TSS of CDKL5 in all the tissues taken in consideration - except for heart adult - with the highest percentage of use in fetal brain, where it represents the 20% of all the TSSs of CDKL5 against the 5% detected in adult brain. Ex1_205down is the second more expressed TSS, especially inboth adult and fetal brain, where it represents about the 20% of the used TSSs of CDKL5. It is noteworthy that the Ex1_down sequence does not include the predicted G42, thus resulting in the absence of a putative cis-acting regulatory element. Therefore, it is interesting that the three most used TSS - and thus first exons - of CDKL5 - Ex1_202, Ex1_205, Ex1_down - lie in the same cluster and differ from each other in the absence of one of two G4s, suggesting the possibility of a regulatory significance of the architectural choice of 5'UTR composition. This could lead to a differential translational efficiency of the CDKL5 protein depending on the choice between the three first exons, due to the addition of a further cis-acting element to the sequence.

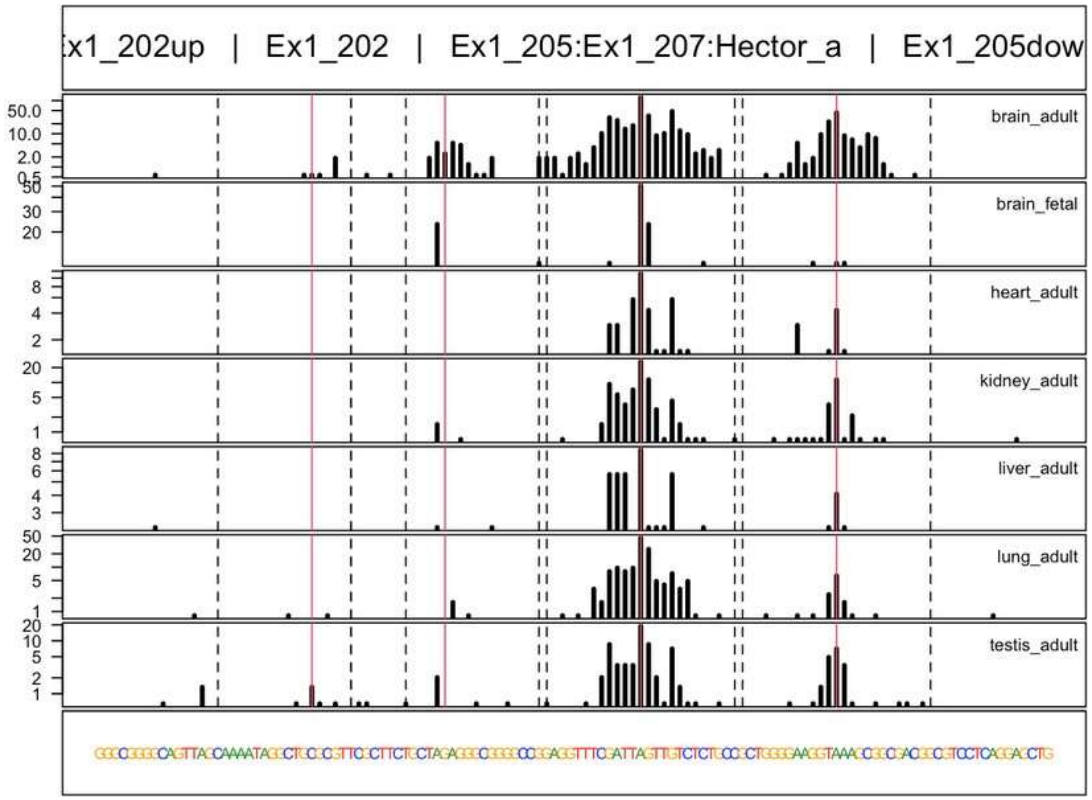


Fig. 5.7.2 Main TSS cluster of the CDKL5 gene. *TSS cluster present in the vicinity of the Ex1_205 TSS, consisting of four different TSSs, three of which resulted to be the more used in the transcription of CDKL5. The TSSs are named as Ex1_202up; Ex1_202; Ex1_205; and the novel Ex1_205down, referring to the name of the relative transcribed first exon. The tissues taken into consideration are adult tissues (brain, heart, kidney, liver, lung, and testis) with the addition of fetal brain. Red lines indicate the TSS quantified, the position of which in the sequence are here reported: previously unknown Ex1_202up: ChrX:184255573; Ex1_202: ChrX: 18425583; Ex1_205: ChrX: 18425608; previously unknown Ex1_205down: ChrX: 18425630.*

5.8 Functional Evaluation of CDKL5 5'UTR through Dual Luciferase Reporter Assay

5.8.1 Two-promoter vector

One of the goals of this thesis was to demonstrate, for the first time, the role of CDKL5 5'UTR in the modulation of the protein expression. To do this, we employed a Dual Luciferase Reporter Assay (DLRA) system, taking into consideration four different sequences as possible modulators of the translatability of CDKL5 (**Fig. 5.8.1 A**):

- A truncated variant of the 5'UTR, composed solely of the UTex2 sequence (162nts), to

assess the effect of the absence of the first exon in the global function of 5'UTR.

- The RefSeq 5'UTR 205 (88 nts + 162 nts), which we demonstrated to be the most represented first exon in the 5'UTR composition (88 nts) and which has the strongest structural prediction, with a significantly lower ΔG and thepresence of two predicted G4s as possible regulatory motifs of the sequence. Itsgreater sequence conservation, in comparison with the more ancient UTex2 sequence, led us to consider the CDKL5 5'UTR containing the Ex1_205 as themost promising functional sequence.
- The 5'UTR 202up (123nts + 162nts), including the Ex1_202, and the 10upstream nucleotides of the TSS of Ex1_202up. Notably, this sequence differs from Ex1_205 for the presence of $G4_1$, located in the region upstream the Ex1_205 TSS. We showed that the $G4_1$ can fold properly and is a very well- conserved motif among the orthologs of the sequence. Therefore, by comparingthe 5'UTR containing Ex1_202 with the one containing Ex1_205, we should beable to prove whether the $G4_1$ has an influence in the modulation of CDKL5 expression.
- The 5'UTR 202up containing the SNP of uncertain significance found by Evans(Evans et al., 2005) in an atypical RTT male patient. The reasoning behind the inclusion of this sequence into our analysis, was to determine whether the SNPis a mutation that can lead to a decrease in CDKL5 protein expression with pathological significance.

These leader sequences were cloned in a vector carrying the gene of the two luciferases used by the DLRA system - Renilla luciferase (RN) and Firefly luciferase (FLuc) – under two distinct T7 promoters. The vector used, pBRm2L (De Pietri Tonelli et al., 2004), was created in our laboratory to evaluate the efficiency of the cap-dependent translation driven by the 5'UTR variant of interest by using the expression of Fluc as a reporter gene (**Fig. 9.4 A**). The presence of the two independent cistrons on the same vector allows us to measure the influence of the 5'UTR cloned in the Fluc translation as a ratio between Fluc and RN signals, bypassing the problem of transfection efficiency variations. 30 minutes before the transfection, we infected SHSY-5Y cells with the MVA-T7 virus, a Modified Vaccinia Ankara strain expressing the T7 RNA polymerase that, in turn, triggers the production of the luciferases from the vectors. With this system, the transcription occurs directly in the cytoplasm, not only allowing the rapid production of the proteins due to the immediate availability of the mRNA to the ribosomes, but also bypassing any nuclear event that, especially with G4 sequence,

could alter the evaluation of the translation efficiency due to the presence of cryptic promoters. Cell lysed five hours post transfection, and the samples were analyzed for reporter protein expression with a Glomax 20/20 luminometer.
In addition to the four CDKL5 5'UTRs listed above, we included in the experiments the following controls:

- The empty pBRm2L vector, in which the sequences between the T7 promoter and the CDS of the RN and Fluc were almost identical and consisted of 45nts without relevant modulatory effects on translation efficiency.
- A pBRm2L vector in which the 5'UTR of the transcript of BACE1 was cloned (De Pietri Tonelli et al., 2004), and reported to have an inhibitory effect on BACE1 translation (Lammich et al., 2004; De Pietri Tonelli et al.,2004; Mihailovich et al., 2017).

The experiment was repeated six times and each sample was measured twice. These technical replicates were averaged to obtain the results relative to the 5'UTR analyzed.

Despite the length and structural complexity, CDKL5 5'UTR did not show an inhibitory effect in the translation of Fluc, such as the one performed by BACE1 5'UTR in the pBRm2L_BACE1 vector. Indeed, we found that its presence enhanced the translatability of the downstream CDS, displaying an opposite trend compared to the effect of the BACE1 5'UTR. In fact, whereas the BACE1 5'UTR inhibits the expression of the downstream Fluc cistron of about 90% compared to the empty pBRm2L, the presence of the UTex2 increases the Fluc expression of Fluc of 70%, while vectors carrying Ex1_205 and Ex1_202up showed an increase of about 50% and 30% respectively. Therefore, the data presented here implies that CDKL5 5'UTR has a modulatory effect in the transability of the downstream CDS, suggesting the existence of a stimulatory element, such as an IRES, rather than a complex structural inhibitory architecture. The suggested element probably involves part of the sequence of UTex2, whereas the addition of a first exon reduces its efficiency, maybe due to the presence of the predicted cis-acting regulatory elements, among which there are $G4_1$ and $G4_2$. In accordance with this interpretation of the results, the Fluc/RN activity decreases in the vector carrying Ex1_202up – containing both the predicted G4s – compared to the vector carrying Ex1_205 – containing uniquely $G4_2$. Notably, the results obtained from the analysis of pBR2mL_EVA did not show any significant difference when compared

with its control vector, pBR2mL_202up, returning a value very similar to Fluc/RN (**Fig. 5.8.1 B**).

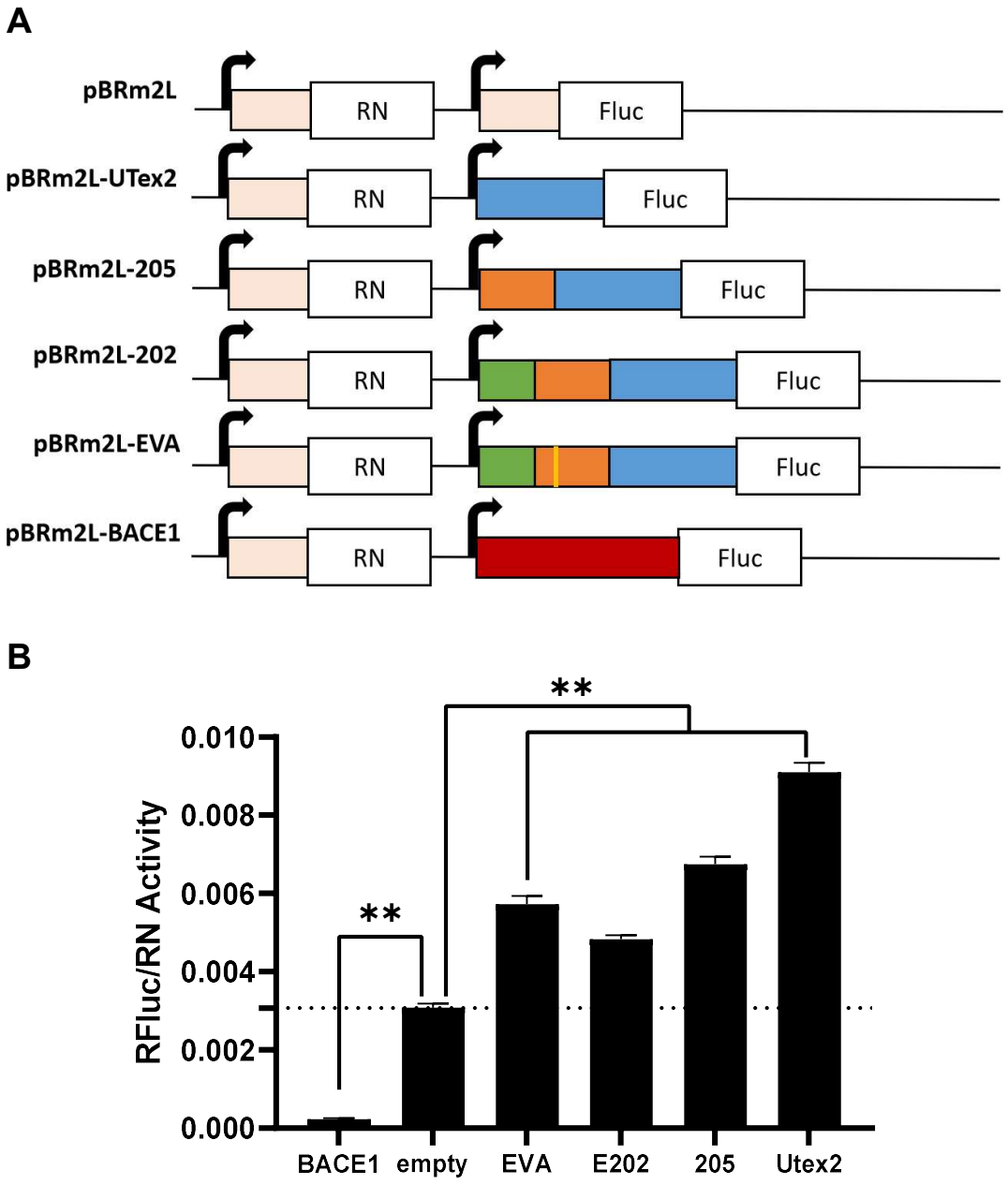


Fig 5.8.1 Dual Luciferase Reporter Assay with the two-promoter vector approach. *(A) Diagram of two-promoter vectors. The construct pBRm2L (empty) contains the two reporter luciferase genes, under control of two distinct T7 promoters. pBR2ml_UTex2 (UTex2) contains a truncated version of the 5'UTRs of CDKL5, composing by the 162nts of the untranslated region in the second exon of CDKL2; BRmL2_205 (Ex1_205) contains the RefSeq CDKL5 5'UTR between the end of the T7 promoter and the CDS of Fluc. pBRmL2_202up (202up)contains the TV 202's 5'UTR with the addition of 10nts upstream the presence of which was experimentally confirmed by RT-PCR and CAGEr analysis; pBRmL2_Eva (Eva) contains*

*the 5'UTR of pBRmL2_202up carrying the SNP of interest rs786204994. pBRmL2-BACE1 (BACE1) contains BACE1 5'UTR, employed as quality control of the experiment. (B) Transfection of SHSY-5Y cells with the mentioned vectors to verify the potential function of CDKL5 5'UTRs in modulating Fluc protein expression. Fluc/RN ratio were reported for the different vectors. UTex2 showed a significant effect in increasing the translation efficiency of Fluc when compared with all the other vectors. 205 and 202up showed an effect in decreasing the translational efficiency brough by UTex2. pBRmL2_Eva did not show an effect in disturbing cap-dependent translation rate of Fluc. BACE1 was used as quality control of the assay. Statistic test employed: Student's t-test (*p<0.05; **p<0.01). Error bars represents SEM index; n=6.*

In accordance with our results, we suggest that the 5'UTR of CDKL5 has a functional relevance in the context of translational regulation. However, the strong difference in translation efficiency with a 5'UTR with similar features, such as the one of BACE1, prompted us evaluate CDKL5 translation efficiency also in the context of cap-independent translation.

5.8.2 Dicistronic vector

Given the results described in the previous paragraph, we decided to evaluate the presence of a sequence with IRES properties within the CDKL5 5'UTR. To verify this hypothesis, we performed the DLRA assay using a dicistronic vector with the same experimental setting as previously described. Dicistronic vectors are vectors where the coding sequence of two genes are under the control of a unique promoter, and the terminator sequence is placed after the end of the second cistron (**Fig. 9.4 B**). This way, the transcript of the second gene lacks a leader sequence that can undergo cap-dependent translation. If the sequence inserted between the two luciferases contains an IRES, the translation of the Fluc will take place (**Fig. 5.8.2 A**).

The vector chosen for the experiment was pBATmod2 (De Pietri Tonelli et al., 2003), and MVA-T7 virus was again employed for the rapid cytoplasmatic transcription in order to obtain the rapid transient expression of the two reporter luciferases. pBATmod2 vector containing the CMV IRES upstream the Fluc CDS and was used as positive control of the quality/reliability of the technique.

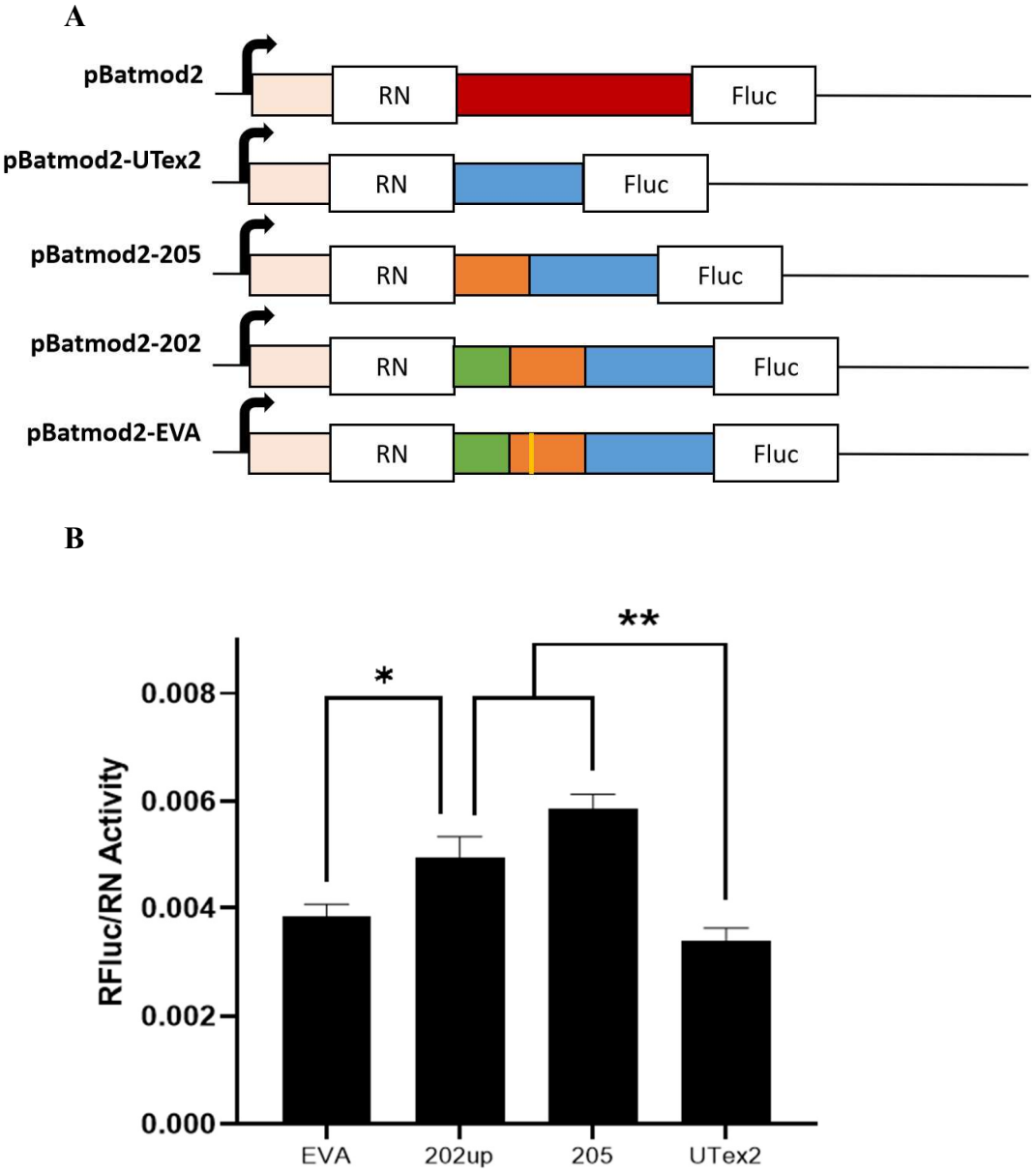


Fig. 5.8.2 Dual Luciferase Reporter Assay with the dicistronic vector approach to reveal a possible IRES activity. *(A) Schematic representation of the vectors employed for the experiments. 5'UTRs sequence was cloned upstream the Fluc CDS, using the restriction sites SalI and NcoI. The absence of a promoter upstream the inserted 5'UTRs allows uniquely cap-independent translation initiation if an IRES is present in the 5'UTR, contrary to RN, that, due to the presence of the T7 promoter and an optimized short 5'UTR, efficiently undergoes cap-dependent initiation. The vectors used in the experiments are: pBatmod2-205, pBatmod2-202up, pBatmod2-UTex2 and pBatmod2-Eva. (B) Result of the DLRA. Due to the Fluc/RN activity difference between the vectors carrying UTex2 and Ex1_205, it is suggested the presence of an IRES, the sequence of which lies in both the two exons of CDKL5 5'UTR, as explained in the paragraph. Even if also the vector carrying Ex1_202up shows an increase of the translation efficiency of Fluc, similarly to Ex1_205, the reduction of the effect but the presence of the total sequence of Ex1_205 suggests the existence of a cis-acting inhibitory*

*element on the additional 35nts upstream the Ex1_205 TSS, compatible with the presence of G41. Interestingly, the vector carrying the SNP at position -189 in Ex1_202up, shows a significative decrease in the translation efficiency of Fluc compared to what observed in the control Ex1_202up. The Fluc/RN ratio of the vector pBatmod2-Eva is comparable to the one recorded for pBatmod2-UTex2, suggesting a possible disruptive effect of the mutation in the IRES folding, presumably directly interested the stem-loop in which the original C lies in the unmutated Ex1_202up. Statistical test employed: Student's t-test (*p<0.05; **p<0.01). Error bars represents SEM index; n=6.*

The result of the DLRA (**Fig. 5.8.2 B**) suggests the presence of an IRES in the 5'UTR of TV 205 - and consequently TV 202up. The putative IRES sequence is included in the region composed by the Ex1_205 and the UTex2, as emerged by the comparison of Fluc/RN ratio resulted from the analysis of pBatmod2-UTex2 and pBatmod2-205 constructs. In fact, the difference in the protein levels of Fluc, showed in the absence of Ex1_205, resulted in a significant decrease in translational efficiency of about 50%. Also, the addition of the 35nts upstream (Ex1_202up) produced an appreciable variation in Fluc signal, even if the presence of Ex1_202up caused an increase of Fluc/RN activity of about 30%. These data suggest that when the Ex1_205 sequence is absent in the 5'UTR, the IRES sequence is probably truncated, and thus the expression of Fluc decreases due to the incompleteness of the IRES motif. The decrease in the RFluc/RN activity between the 5'UTRs carrying Ex1_205 and Ex1_202up could presumably be due to the influence of a cis-acting regulatory element presents in the upstream region of 35nts of Ex1_202up that slightly reduces the translational efficiency produced by the IRES. We believe that this cis-acting regulatory elements could be the $G4_1$ identified by us in **par. 5.6** and previously demonstrated to have an impact in translation efficiency within two-promoter vector.

An interesting observation comes from the analysis of the Fluc/RN activities recorded from the vector carrying the SNP described by Evans (Evans et al., 2005). The presence of the SNP in the Ex1_202up caused a convincing decrease in the expression of Fluc compared to its control vector (carrying the Ex1_202up), estimated to be about 25% - a value comparable to the effect obtained by the absence of the entire Ex1_202up within

the CDKL5 5'UTR. Thus, the presence of the transition C>T in the position -189 (when 1 is the first nucleotide of the main start codon), presumably alters the stability of the IRES, resulting in a reduction in CDKL5 translation efficiency. Of course, the result must be confirmed using the sequence of the CDKL5_205 5'UTR, since the absence of the decremental effect on the translation of Fluc given by the inhibitory $G4_1$ might alter the relevance of the SNP.

The effect of the SNP is probably due to the disruption of the folding of the second stem-loop of the CDKL5 205 5'UTR, in which the C interested by the mutation lies. Even though further experiments must be performed to strengthen the results derived from our data, this is the first (*in vitro*) experimental evidence of a possible pathological effect of the Evans' SNP on the translation efficiency, that potentially links a decrease of CDKL5 to the development of CDD.

6 DISCUSSION

The protein kinase CDKL5 is emerging as a very important player in a variety of neuronal functions, with regard to the development of an adequate neuronal morphology. In fact, its role in determining the correct axonal outgrowth, dendritic arborization and the overall homeostasis of the synaptic network points at CDKL5 as a central player when it comes to proper neurodevelopment (Barbiero et al., 2019). Moreover, its involvement in regulating the transcription of other genes, by modulating the action of transcriptional regulators such as Dnmt1 and MeCP2 (Kameshita et al., 2008; Sekiguchi et al., 2013), as well as in determining processes such as programmed cell death (Fuchs et al., 2014), strongly highlights the need of gaining a better understanding when it comes to its regulation. In addition, the existence of the CDKL5 Deficit Disorder (CDD), a monogenetic neurodevelopmental disorder, caused by the shortage of functional CDKL5, makes it imperative to deepen the knowledge regarding CDKL5 functions and the regulatory mechanisms that modulate its expression in neurons (Olson et al., 2019). Previous publications have reported that CDKL5 is highly enriched in the murine brain (Rusconi et al., 2008), displaying a great degree of variability both in terms of brain regions as well as developmental stages. This strongly suggests the involvement of a tight regulation program in the CDKL5 gene expression. In this context, our results highlight the role of the transcriptional process in modulating the CDKL5 protein expression. In fact, we observed that both the transcript and protein levels follow the same trend following quantification via RT-PCR and WB. This is true in both different murine tissues and developmental stages of the cerebral cortex (**par. 5.1**). The highest levels of CDKL5 expression were found in neurons, which is in accordance with the reported functions of the kinase within this type of cells (Rusconi et al., 2008). Specifically, a recent study has shown how CDKL5 protein translation can also be localized at synaptic sites, in response to stimuli mimicking neuronal activation (La Montanara et al., 2015). We confirmed these results *in vitro* on murine cortical neurons, by recording CDKL5 protein levels following acute treatments with drugs known to enhance neuronal translation (**par. 5.2, Fig.5.2.1**). Bicuculline was included in the experimental approach since it is known to enhance neuronal network activity also in vitro. Moreover, it has been shown to increase the translational efficiency of some neuronal transcripts, such as BACE1 (Bettegazzi et al., 2021). In particular, it is well documented that the targets ofbicuculline-mediated expression activation are a

subset of neuronal genes named immediate early genes (IEGs), i.e. genes whose expression is dynamically regulated inresponse to neuronal activation, often involved in synaptic plasticity. The regulation ofIEGs can be performed by both transcriptional and translational mechanisms (Bramhamet al., 2008; Minathohara et al. 2015).The definition of CDKL5 as IEG from atranscriptional point of view (Miranda et al., 2021), in addition to the first findings about the translational regulation of its expression after acute treatments *in vitro* (La Montanara et al., 2015), led us to wonder if Bic acute treatments could increase CDKL5expression levels in a similar way as it happens for BACE1. Our investigation showed that the acute treatment with Bic caused the upregulation of CDKL5 expression in neurons *in vitro* (**Fig. 5.2.1**). Therefore, in order to deepen and confirm the effect of Bic on CDKL5 protein expression levels, we administered bicuculline to an *in vivo* mouse model and observed a significant increase in CDKL5 expression levels after 30minutes from the intraperitoneal injection. To the best of our knowledge, this is the first*in vivo* experiment that demonstrated an effect on the CDKL5 regulation after acute treatment with a drug associated with enhanced neuronal activation. These findings indicate the presence of a translational control in setting CDKL5 expression, at least inneurons. This is in accordance with previous reports revealing how proteins with analogous function to CDKL5 often recur to mechanism modulating translational efficiency, especially if involved in activity-dependent local translation (**par. 3.1**).

Translational initiation is recognized as the main step of protein synthesis in which translational control occurs, often acting on cis-acting elements contained in 5'UTRs (**par. 3.2**). For this reason, we decided to analyze the CDKL5 leader sequence starting from studying the architecture of all the reported variants, in order to verify whether CDKL5 5'UTR retained the characteristic features associated with functional 5'UTRs reported in the literature (Davuluri et al., 2000). Surprisingly, we found that this task was more complicated than expected, since a lot of CDKL5 5'UTR variants reported in databanks often have discordant and confusing annotations, and information in the literature is scarce and incomplete (**par. 5.4.1**). Nevertheless, it appeared that all the analyzed 5'UTRs have the same architecture, consisting of a variable 5' region, composed of alternative first exons, and of a common region, composed of the 5' untranslated part of exon two (UTex2, 162 nts). Indeed, the unique CDKL5 start codon, common to all the protein isoforms reported for CDKL5, is in exon two, since all these

proteins share the same N-terminus but vary in C-terminus composition due to alternative splicing (**par. 3.3**). The reason behind this architecture and the abundance of first exons was unknown, and we decided that further work on this would be one of the aims of this research. To complicate the picture, most of the reported 5'UTR variants are predicted or poorly described from an experimental point of view. The only publication that experimentally explores the CDKL5 5'UTR composition is the one provided by Hector (Hector et al., 2016 a and b), but we did not find its results to be reliable. In this work, 5'RACE-PCR was performed on samples collected from human, mouse and rat, to determine qualitatively and quantitatively the CDKL5 5'UTR first exon usage. Unfortunately, the experimental limitations of the technique, as pointed out by the authors in the cited papers, do not allow to consider the work as definitive, but only as indicative. In fact, the amplification technology often causes the loss of some 5'UTR regions, masking the GC-rich sequences that have difficulty to be amplified.

We initially considered all the 5'UTR variants at our disposal to be equally important, in order not to erroneously exclude possible functional CDKL5 leader sequences from our analysis, due to the aforementioned limitations. Our aim was to analyze these sequences by using predictor bioinformatics approaches, in order to screen the most probable functional CDKL5 5'UTR variants and experimentally determine their significance in translational efficiency (**par. 5.4**). Our approaches followed the published guidelines to investigate the presence of cis-acting regulatory features of 5'UTRs, since a significant difference between the several alternative first exons composing the 5'UTR variants was almost certainly given by the different ensemble of regulatory elements within the sequences. On the other hand, the CDKL5 5'UTR architecture itself, composed by the common sharing of the 3' termination of the sequence leader, suggested a possible determinant role of UTex2 in the 5'UTRs, and thus, the presence of cis-acting elements. Thus, we hypothesized a model in which each existent CDKL5 transcript would be translationally regulated by a specific regulatory function embodied in the common UTex2, while the alternative composition at the 5' end of each 5'UTR variant integrates the translational control of CDKL5 protein levels, providing a specific ensemble of cis-acting elements. In this model, the choice of 5'UTR composition could be originally determined by a differential transcriptional program depending on cellular contexts or developmental stages. The integration of mechanisms of transcriptional and translational regulations would be central in setting CDKL5 protein levels in response to the needs of the cells. The first step to verify our

model was the aforementioned screening, in which we considered features of 5'UTR sequences, such as the calculation of GC content (**par. 5.4.2**) and the sequence conservation (**par. 5.4.4**), as well as the prediction of cis-acting regulatory elements, such as upstream AUGs (**par. 5.4.3**) and secondary structural motif (**par. 5.6**). Concerning the calculation of GC content and the sequence conservation, we evaluated distinctly the exon composition of 5'UTR variants, to investigate the difference between the alternative first exon and the common UTex2 regions. We found that the GC content of the total 5'UTR variants do not reach the threshold of 60% of the total nucleotide composition of the sequence, an indicative value consisting in the average of GC content of human 5'UTRs previously considered as threshold in other publications (Davuluri et al., 2000). Nevertheless, the distinct consideration of single exons, returned how most of the contribution on GC content was given by few alternative first exons, including Ex1_205/202. The GC content of these exon sequences reaches and exceeds the value of 60%, suggesting the possibility of a stable structural folding. Unfortunately, to the best of our knowledge there is currently no information present in the literature about exon distribution of the GC content of 5'UTRs in human transcriptome. Moreover, in our opinion any comparison with other specific 5'UTRs is uninformative, since each 5'UTR is an ensemble of integrative specific features and nowadays there is not a division of known 5'UTRs into categories sharing analogue architectures.

For the same reason, we could not compare our result regarding the conservation of the sequence of our interest with other individual 5'UTRs, since conservation is a specific feature of each sequence. Therefore, our evaluations in regard to these two parameters – GC Content and Sequence Conservation – can be interpreted only in reference to the analyzed sequences. Similar to GC content, it emerged that the conservation of a specific first exon, Ex1_205/Ex1_202 is greater than the sequence conservation of UTex2. This was unexpected, since Utex2 has a more ancient origin than Ex1_205/Ex1_202, but it suggests the possibility that this sequence may actually retain a function.

We have referred to Ex1_205/Ex1_202 up to this point, since Ex1_202 includes the complete sequenced of Ex1_205 and an upstream region of 25 nts. Ex1_202 was one of the CDKL5 alternative first exon found on Ensembl, the experimental details of which were poorly reported. Thus, since it emerged as the second first-exon of interest from our bioinformatics screening, we decided to verify its existence. To reach this aim,

we employed both an experimental analysis by using RT-PCR (**Fig. 5.5.2**) and a bioinformatics analysis performing a nBLAST search in dbEST. The integration of these two approaches confirmed the presence of the Ex1_202 sequence in human and mice transcriptome, as well as the presence of 10 additional nts upstream the reported beginning of Ex1_202 (**Fig. 5.5.1**). However, the lack of further information about Ex1_202 in the databanks, let us to wonder whether the reported Ex1_202 wastruncated or we found another novel CDKL5 first exon. Therefore, we decided to consider this elongated version of Ex1_202 for the further analysis, naming it 202up. Nevertheless, we confirmed the existence of Ex1_202 as a distinct alternative first exonof CDKL5 5'UTR. To understand what the possible difference in choosing Ex1_205 or Ex1_202(up), we continued with the prediction of cis-acting regulatory elements - to which we have subjected all the 5'UTR variants. uAUGs prediction did not return any significant result in any of the 5'UTR variants analyzed (**Tab. 5.4.3**) pointing towards a functional role of structural regulatory motifs within highly conserved sequences that are rich in GC content. Therefore, our structural prediction had to consider canonical Watson and Crick pairing, as well as non-canonical Hoogsteen secondary structures (**par. 5.6**). Even if most of the 5'UTR sequences do not contain strong Watson and Crick stable structured regions (**Tab 5.6.1.1** and **Tab 5.6.1.4**), it appeared evident how Ex1_205 and Ex1_202up strongly influenced the global folding of CDKL5 5'UTR in contrast with the other alternative first exons (**Fig. 5.6.1.6**). In fact, the best folded 5'UTR returned from our analysis has been undoubtedly the 205 5'UTR, characterized by the presence of six stem-loops that closely compact the sequence in a highly structured manner, where the first exon strongly contributes to stabilize the global folding (**Fig. 5.6.1.5**). Also, as expected, the 202up 5'UTR displays a very similar trend, but with the significant difference that the upstream 25 nts greatly disturb the foldingof the 5' region of the sequence, as shown by the higher positional entropy of the first 50 nucleotides in the two variants (**Fig. 5.6.1.7**). We have evaluated that this disturbance would be due to the presence of competitive non-canonical Hoogsteen base-pairing, in accordance with the high G content of the upstream 35 nts of Ex1_202up. Indeed, a G4, named $G4_1$, was predicted in the upstream 25 nucleotides ofEx1_202 (**Fig. 5.6.2.1**). Similar to the results obtained from the Watson and crick predictions, the only exons in which G4s have been predicted in addition to $G4_1$ have been found in Ex1_205, with a more reliable prediction for the G4 called $G4_2$, located at the beginning of the sequence (**Fig. 5.6.2.1** and **Fig. 5.6.2.2**). The presence of an additional G4 in Ex1_202up in

comparison with Ex1_205, suggested that the differencein the usage of the two variants could be the due to the presence of an additional structural cis-acting element, represented by $G4_1$. To strengthen the prediction and therefore this hypothesis, we performed circular dichroism spectrophotometry to assessthe $G4_1$ capability to properly fold itself. The analysis returned a positive result (**Fig. 5.6.3**). Thus, our predictions provided indications that were in line with our model, also in accordance with the literature. In fact, cases in which longer 5'UTRs bring additionalcis-acting regulatory elements, compared to shorter versions, are already described in the literature. The choice of alternative TSSs to include additional cis-acting elements in 5'UTRs is a well-described mechanism to model gene expression (**par. 3.2**). The involvement of 5'UTR structural regulatory elements in CDKL5 translational control has been strengthen by our findings showing that CDKL5 protein expression level decreases drastically when eIF4B is silenced using siRNA (**Fig. 5.3.1**). The action of eIF4B in modulating the unwinding of highly structured 5'UTRs has been recognized for a long time but has been recently shown how this factor acts in neurons as a convergent point of numerous signaling pathways modulating translationally the efficiency of transcripts (Merrick et al., 2015; Bettegazzi et al., 2017; Bettegazzi et al.,2021). The proposed link between eIF4B presence and CDKL5 expression, highlightedfor the first time in this thesis, could drive to future experiments to confirm the involvement of eIF4B in CDKL5 translational modulation. For example, it would be interesting to apply approaches in which the action of eIF4B in neurons is modulated by its phosphorylation status, and, in particular, at the neuronal specific Ser504 site. This site was previously suggested to be involved in local translation, sensing neuronal activation, and converting the information in an increased translational efficiency of some transcript. In fact, acute treatment with bicuculline - causing the phosphorylationof Ser504 - was found to increase the expression of BACE1, the GC rich 5'UTR of which is associated to close structural folding. (Bettegazzi et al., 2021). Therefore, since we also found that bicuculline increased CDKL5 expression levels – both *in vitro* and *in vivo*– and that the bic-mediated effect on translation of highly structured 5'UTRs is mediated by eIF4B, we proposed eIF4B as a regulator of CDKL5 translation in neurons.Interestingly, we found that also MeCP2, another X-linked gene linked to the neurodevelopmental disorder named Rett Syndrome (RTT), has a severe reduction of its expression level when we silenced the eIF4B gene (**Fig. 5.3.1**). Traditionally, the CDKL5-MeCP2 regulatory association has always been a topic of great interest in the context of

neurodevelopmental biology. Firstly, because in the past decades CDKL5 was a gene linked to the onset of atypical forms of RTT and, thus, there are many similarities between CDD and RTT (Pini et al., 2012; Chahil et al., 2021; Guerrini et al., 2021); and, on the other hand, because different publications have already found the mutual actions of these two proteins, since the MeCP2 is a transcriptional repressorof CDKL5 transcription (Carouge et al., 2010), and the presence or the absence of CDKL5 correlates with an increase or a decrease of the phosphorylation status of MeCP2 (Mari et al., 2005; Lin et al., 2005). Therefore, our findings provide novel evidence of the connection of these two proteins, indicating that they could probably share analogous translational control mechanisms in neurons. In accordance with that model, we found that the MeCP2 5'UTR is very similar to the CDKL5 one in terms offeatures and composition, since it contains predicted G4s, and it is very conserved in mammals (Bagga et al., 2013). Taking into consideration all this information, we strongly believe that MeCP2 5'UTR will also be a topic of great interest deserving attention.

Even more intriguingly, these results suggest how the role of eIF4B in neuronal biology might be more prominent that what it is thought nowadays, especially in the context of neuronal development. Unfortunately, we cannot strengthen this hypothesis by consulting the reported eIF4B mutations and their association with human pathologies, since the unique mutations reported in ClinVar are macrodeletions and macroduplications affecting numerous genes. Moreover, the absence of detailed information regarding the proteomics of eIF4B KO mouse models makes it impossible to date to further investigate the proposed link between eIF4B and abnormal neuronal development. However, we believe that it could be noteworthy to include the eIF4B gene in the diagnostic screening for neurodevelopmental disorders involving abnormal neuronal morphologies, especially when the diagnostic screening of recognized causative genes does not return any mutations.

Having collected new information about the translational regulation of CDKL5, in addition to the bioinformatics screening that helped us focus on two potentially interesting 5'UTR variants, we proceeded with the experimental verification of the functionality of the CDKL5 leader sequence (**par. 5.8**). The technique that we used was the traditional Dual-Luciferases Reporter Assay System, in which the measurement of relative activities of the firefly and the renilla luciferases in the same sample, expressed as Fluc/RN ratio, returns the influence on translation of the considered 5'UTR cloned

upstream the Fluc cistron. The two cistrons are cloned in two-promoter vector named pBRm2L, previously used to assess the effect on cap-dependent translation of the highly structured BACE1 5'UTR – that for the reasons mentioned above was included in our experiment as control (De Pietri Tonelli et al., 2004; Bettegazzi et al., 2021). The presence of distinct T7 promoters upstream the two cistrons allows the direct transcription of Fluc and RN transcripts in cytoplasm under the condition of infection by a vaccinia virus (MVA virus) carrying the T7 RNA polymerase. In this manner we avoided the occurrence of false positives because of cryptic promoters or splicing events that are often associated with this type of experiments when DNA vectors, transfected in a classical way, need to be transcribed in the nucleus as mandatory step (**Fig. 3.1.2.3**, Yang et al., 2019). The experiment successfully demonstrated the influence of the CDKL5 5'UTR on the translational efficiency of Fluc. Moreover, the result confirmed our hypothetical model for which the two exons constituting the entire 5'UTR cooperatively act in setting the Fluc protein expression. In fact, we recorded a sustained translational efficiency and even an increase in the Fluc expression when we considered uniquely the UTex2 as leader sequence inserted upstream the cistron. This result highlighted the possibility of the presence of an IRES, a structural cis-acting element heightening the translation of a certain eukaryotic transcripts, the activation of which is triggered by different kinds of stimuli, such as stress conditions. In these contexts, eukaryotic transcripts, which are usually translated trough cap-dependent initiation, switch to IRES-mediated translation to better fulfill cells requests (**par. 3.2.4**). It is well documented in the literature that one of these triggering stress conditions can be a viral infection (Spriggs et al., 2010; Yang et al., 2019), such as the MVA infection included in our experimental design. Therefore, we assumed that in this first experiment we observed a mixed effect due to the simultaneous occurrence of the cap-dependent translation and the IRES-mediated translation given by the probable presence of the cited element in UTex2. On the other hand, the vector carrying the complete 205 5'UTR variant displayed a decrease of translational efficiency of Fluc, probably due to the presence of the predicted structural cis-acting elements, among which we believe $G4_2$ could be present. Moreover, the vector carrying 202 5'UTR, possessing a further structural regulatory element represented by $G4_1$, showed a further reduction in Fluc translation, potentially confirming our assumption about the differential usage of Ex1_202 and Ex1_205 to integrate the 5'UTR functionality (**Fig. 5.8.1**). We also assessed a control, BACE1 5'UTR, to better interpretate this result. BACE1 5'UTR was

previously reported to strongly inhibit translation using the same experimental design, and this effect was associated also to its highly structured nature (De Pietri Tonelli et al., 2004; Lammich et al., 2004). As expected, since BACE1 5'UTR does not retain an IRES, the Fluc expression was found to be strongly inhibited, strengthening our idea of the involvement of an IRES.

Interestingly, we also included in this experiment a CDKL5 5'UTR (202up 5'UTR) carrying the unique SNP of uncertain significance from a patient associated with an atypical form of RTT, which could be diagnosed as CDD (Evans et al., 2005). SNPs of uncertain significance are defined as variants identified through genetic testing but for which there is a lack of information about their significance or role in disturbing the gene function (Richards et al., 2015). Therefore, we decided to use our assay to investigate the nature of Evans's SNP, since the absence of a previously described function of the CDKL5 5'UTR and a reliable assay, made it problematic to assess the role of this SNP. However, in a first attempt using a two-promoter vector, the SNP did not show an appreciable impact on the translation of the downstream CDS **(Fig. 5.8.1**). Starting from the evidence, obtained from the two-promoter approach of a good translatability of CDKL5 transcript in spite of the complex 5'UTR, we wanted to investigate the possible presence of an IRES. We repeated the same type of experiment but using a different vector suited for the study of IRES-mediated translation, excluding the impact of cap-dependent initiation: the dicistronic vector pBATmod2. In this vector the two cistrons are under the control of a unique T7 promoter, allowing the cap-dependent initiation of the upstream RN, whereas the Fluc - downstream to our 5'UTR of interest – is forced to start its translation in an IRES-mediated manner or it will not be translated at all (De Pietri Tonelli et al., 2003). This analysis strongly suggested the presence of an IRES within UTex2, but also extending in the first exon. In fact, the vectors carrying Ex1_202up and Ex1_205 also allow a good translation of Fluc, with expected progressive mild reductions in dependence of the length of the 5'UTRs. The involvement of first exons in the IRES could be also demonstrated by the effect of the Evans's SNP on the translation of Fluc. This was undoubtedly the most interesting result, showing how the presence of the SNP strongly reduces Fluc/RN ratio compared to the 202up 5'UTR control (**Fig. 5.8.2**). We propose that the SNP disturbs the IRES functionality, similarly to what was observed in c-Myc, connexin 32 and VEFGA (Chappell et al., 2000; Lambrechts et al., 2003; Shi et al., 2016; Péladeau et al., 2021). However, even if other controls are to be tested to confirm this first result, this is the

first experimental evidence of the significance of the SNP found by Evans, potentially linking its occurrence with CDKL5 shortage and thus with CDD etiopathology. Indeed, it was already reported in the literature that a single SNP in IRES can be causative of a complex human disease, ranging from melanoma to neurodegenerative disorders such as Charcot-Marie Tooth and ALS (Lambrechts et al., 2003; Shi et al., 2016; Péladeau et al., 2021). Therefore, we suggest that Evans's SNP could represent a pathological SNP, causing the reduction of the required physiological level of CDKL5 in cells. This assumption is in line with the fact that CDD abnormal neuronal morphology can be caused by low CDKL5 protein level *in vitro* (Ricciardi et al., 2012).

In the end, our results proved the importance of considering CDKL5 202 and 205 5'UTR in CDKL5 regulation, but they did not give an indication about the influence of these mechanism in the overall CDKL5 protein expression. In fact, the presence of the other 5'UTR variants, excluded from our analysis based on our bioinformatics screening, continues to complicate the framework of the information about CDKL5 5'UTR. In other words, we did not have an indication of the percentage of 5'UTRs carrying Ex1_205 and Ex1_ 202, and, as such, we could not evaluate the biological relevance of our results. For this reason, we performed CAGE peaks analysis in order to quantify the TSS usage of CDKL5 transcripts in different human tissues libraries obtained from the FANTOM5 server. Quantifying the percentage of CDKL5 TSSs usage corresponds to quantifying the percentage of the specific first exons contained in transcribed 5'UTRs (**par. 5.7**). The libraries obtained from CAGE sequencing are the most reliable source of information to approach this task, since this methodology does not involve amplification but is based on the count of single tag sequences, normalized to the size of the library (Haberle et al., 2015, **par. 9.3**). Moreover, the availability of several libraries for different adult human tissues also makes it possible to confront the CDKL5 TSS usage in different conditions, obtaining a complete perspective about the preferences of CDKL5 transcript composition. This could potentially underlie possible tissue-specific translational control mechanisms based on the composition of the CDKL5 5'UTRs and the ensemble of the selected cis-acting elements. Interestingly, our analysis returned different information (**Fig. 5.7.1**). Firstly, most of the reported alternative first exons, even if the ones experimentally quantified by Hector, are almost not present at all in CDKL5 transcripts. Indeed, the great majority of CDKL5 transcript adopt 205 5'UTR, since 205 TSS is confirmed as the dominant TSS of the gene. Ex1_205 is well expressed in all the analyzed body tissues, and it is especially enriched

in the brain, in accordance with the previous experimental findings in the literature, representing in about the 70% of the transcripts in the human adult brain. The second most represented first exon was named Ex1_205down, since its TSS is located at about 20 nucleotides downstream the dominant 205 TSS. This is a novel Ex1 previously unknown since amplification techniques mask downstream TSSs not allowing the identification of shorter 5'UTRs. It represents about one quarter of the total TSSs used in adult brain and about 20% in fetal. This variant differs from Ex1_205 for the absence of the predicted $G4_2$, thus, presumably, when it will be subjected to experimental analysis it will present a higher Fluc/RN ratio than 205 5'UTR. The Ex1_202 is the third exon in order of usage, and it is contained in 5% of the adult brain CDKL5 transcripts, while it is in about 20% of the fetal brain mRNAs. This might suggest how its differential regulatory role could be determinant in the brain development. This three TSSs constitute the CDKL5 main TSS cluster, that also includes an upstream TSS, located 10 nucleotides upstream the 202 TSS that we previously demonstrated to exist and named 202up - that does not contain further cis-regulatory elements (**Fig. 5.5.2** and **Fig. 5.7.2**). The evidence collected by the CAGEr analysis confirmed the result of our bioinformatics screening, conferring biological relevance to our results. Moreover, we demonstrated the utility of using CAGE libraries to study 5'UTR 5' termination, since this technique can overcome the limitation of the previous approaches used in this field of research. We demonstrated with reliability the existence of Ex1_202 and of Ex1_202up exons, excluding truncation hypothesis and other false interpretations. Further analyses are necessary to better evaluate Ex1_202 effect excluding the 10 nucleotides of 202up, but our protocol has proven to be useful to face the challenges of investigating 5'UTRs.

Taken together, our data suggests that the 5'UTR plays a crucial role in the CDKL5 protein expression, confirming our initial hypothesis. A fundamental transcriptional program supports the translational control provided by different 5'UTRs, chosen within a TSS cluster that we detailed here for the first time. The presence or the absence of inhibitory cis-acting elements that we individuated in G4s, can vary the translational efficiency of CDKL5, depending on the choice of Ex1 and the number of cis-acting elements contained in the sequence. Even though we did not distinguish if G4s role is to inhibit translation or to act as platform for RBPs, we experimentally found their presence to be of interest, since they could represent cis-acting elements for modulating CDKL5 expression. Certainly, the presence of an IRES is one of the most interesting

outcome of our work. We found that it partially lies in UTex2, but that it involves Ex1s expressed by the dominant TSS cluster (**Fig. 6.1**).

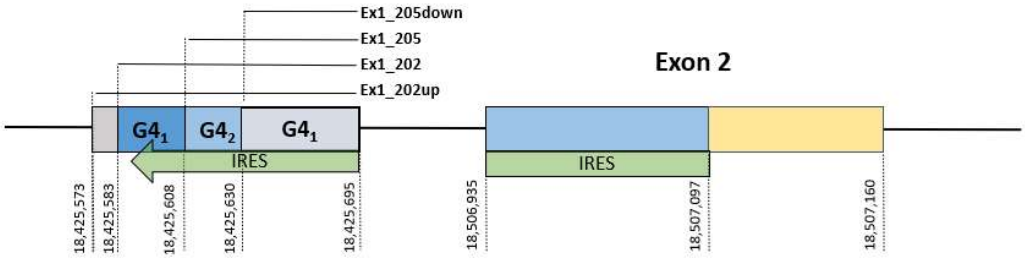


Fig. 6.1 CDKL5 5'UTR architecture mediating translational efficiency through cis-acting elements. *Schematic, not to scale, representation of CDKL5 5'UTR, reporting position of different portions of the sequence leader. 5'UTR is composed by a first region that can be represented by four alternative first exons of different length, based on the TSS used. The three TSSs at positions 18425583, 18425630 and 18425695 express the three mainly used first exons, differing each other for the presence of a predicted G4. The recently identified Ex1_205down probably excludes inhibitory structures and contains part of the first region of an IRES expanding in UTex2 sequence, here colored in light blue. Therefore, the selection of a TSS from the cluster, and, thus, of a longer first exon, influences the regulation of CDKL5 protein synthesis, whereas, in a still unclear condition, its translation can be supported by either cap-dependent or IRES-mediated initiation.*

Our results clearly prompt for further work and even if require confirmation and expansion. However, we believe that this field of research deserves particular attention for the following reason. In fact, these first, promising data about CDKL5 5'UTR could also potentially lead to new therapeutic strategies for treating CDD in the future. The use of ASOs targeting 5'UTRs is already considered an innovative and convenient system to cure complex human diseases, especially if they are of monogenetic nature, such as CDD. ASOs employment is already approved in the clinic, because of its safety of use on patients, as well as for its specificity, avoiding undesired interactions with other targets inside cells (Bennett et al., 2019). For example, ASOs has been already proposed for cystic fibrosis (CF), a monogenetic disease in which mutations disrupts the functionality and expression of the causative protein CFTR, leading to the onset of the pathology (Mall et al., 2014). Even if it has not been already demonstrated *in vivo* that the reduction of the level of an unmutated form of CDKL5 can be causative of the disease, this evidence has been already collected in *in vitro* experiments, as previously

mentioned (Ricciardi et al., 2012). Therefore, it cannot be excluded that mutations in CDKL5 gene reducing its expression at the translational level can be causative of the neurodevelopmental disease, not guaranteeing the proper phosphorylation of the kinase targets. Our data supports this hypothesis, since we collected the first experimental hints about the effect of the Evans's SNP on the translational efficiency of a transcript, suggesting the link between the SNP and the pathological condition of the patient in which it was found.

If our results will be confirmed, deepening the knowledge about translational control of CDKL5, 5'ASOs strategies could be considered to treat some forms of CDD, acting by modulating the translatability of CDKL5 through its cis-acting elements, as it is taking in consideration for treating CF (Sasaki et al., 2019), or boosting the translation of correct transcripts. For example, another class of ASOs acts by correcting splicing errors and thus fixing the expression of mutated pathological variants. This technology is considered to treat SMA, where 5'UTR ASOs are used to increase the expression of the correct transcripts (Winkelsan et al., 2021). Indeed, about 15% of the CDKL5 pathogenic mutations is reported to affect splicing (Balestra et al., 2019), and, therefore, the knowledge regarding the 5'UTR-mediated translational control of CDKL5 could potentially find indirect applications not only in treating mutations in 5'UTRs, but more importantly in boosting CDKL5 expression of correct isoforms.

Even though there is much ground still left to cover, we hope that our contribution can move the focus on the correct inclusion of CDKL5 5'UTR in the diagnostic screening for CDD and mark the beginning of a new point of view from which to consider CDD and its treatment. Moreover, we believe that our preliminary data about MeCP2 and eIF4B could aid in reaching a better understanding of the pathways involved in neurodevelopment in the future, helping to enlighten the complex mechanisms involved in neuronal biology.

7 MATERIALS AND METHODS

7.1 Bioinformatics methods

7.1.1 Data Banks

CDKL5 5'UTR sequences used for the bioinformatic analysis were taken from the nucleotide database of NCBI (Benson et al., 2018) and from the GRCh38.p13 assembly in Ensembl (Howe et al., 2021).

7.1.2 nBlast Analysis

Nucleotide Basic Local Alignment Search Tool (nBLAST) (McGinnis et al., 2004) was used to conduct different types of analysis, using two different databases:

The nucleotide collection, consisting of non-redundant GenBank, EMBL, PDB and RefSeq sequences.

The databases of Expressed Sequence Tags (ESTs), containing single-pass sequence reads from mRNA retrotranscribed in cDNA.

nBLAST search was performed using the BLASTN algorithm, setting a Expect threshold (E value) of 0.1 and word size to 7. The scoring parameters were chosen to promote the inclusion of more dissimilar sequence, because of the nature of untranslated regions:

- Match/Mismatch scores gave a reward of 1 and penalty -2 for marching and mismatching bases.
- Gap costs, representing the cost to create and extend a gap in an alignment, were set to Existence:1 and Extension 2.

Default parameters of the section "Filters and Masking" were maintained.

7.1.3 uAUGs prediction

uORFs prediction was performed using two different tools to detect upstream AUG start codons and to evaluate their Kozak context. Translation Initiation Start Site (TIS). Netstart (DTU Health Tech, Pedersen et al., 1997) uses a combination of local start codon context and information from the global sequence to return a score that can range from 0 to 1. The threshold of significance for predicting a functioning TIS is 0.5. To strengthen our prediction, we employed a second predictor that uses a different algorithm. In this manner, we could compare the results returned from Netstart with the ones returned from the second predictor, having a better evaluation of the Kozak contest

and the probability of TIS to being functioning. The second predictor chosen was TIS Miner, that returns a score ranging from 0 to 1. Threshold of significancy can be chosen by the operator depending on the accuracy and the precision of the predictions wanted. We set a threshold of 0.5, that allows an accuracy of 77.5%, a sensitivity of 71% and a precision of 81.6%. Moreover, it was reported the similarity of the context of the candidate TIS with the Kozak consensus sequence.

7.1.4 Watson and Crick Structural prediction

Analysis of Watson and Crick secondary structures was conducted using RNAfold Server by Vienna RNA Web Services (Gruber et al., 2008), after GC content evaluation. GC content evaluation was conducted calculating the GC percentage on the length of the sequences.

RNAfold returns thermodynamic secondary structures prediction of single stranded nucleic acid sequences. The tool allows to have two different structural predictions, each one with a distinct graphical output. Minimum free energy (MFE) prediction calculates the most stable folding of the input sequences, characterized by the lowest ΔG.The centroid (CE) prediction is the predicted structure with minimal base-pair distance to all structures in the thermodynamic ensemble. It is possible to evaluate the frequency of the MFE structure in the ensemble - that is a measurement of the reliability of the MFE prediction - and the ensemble diversity - that is the average base-pair distance between all the structures of the ensemble and it is a measurement of the heterogeneity of the ensemble.

Graphical outputs allow to recognize two dimensional structural motifs, as stem loops, bulge, and hairpin loops, that can be observed also through the forna package (Kerpedjiev et al., 2015). The structure was displayed with a color code that allows to visualize the positional entropy. Positional entropy shows that positions that have low entropy are more stable and well-defined in the structural prediction, while high entropic positions are ill-defined in the structure. Positional Entropy can be visualized also as a plot of entropy versus position, describing the entropic landscape of the input sequence in an alternative way. A Mountain plot is described as a representation of a predicted 2D secondary structure, plotted as height versus position. Height is obtained from the number of base pairs near the given position k. In a Mountain plot, the MFE structure, the thermodynamic ensemble of RNA structures, and the centroid structure are compared to show possible discrepancies or similarities.

All these parameters are used to have the best possible interpretation of the prediction.

The calculation of the MFE of partially random 5'UTRs and the consequent comparison with the MFE obtained from the original CDKL5 5'UTRs was performed using the package XNAString in RStudio (**par. 5.6.1**).

7.1.5 RNAsnp

RNAsnp web server (Sabarinathan et al., 2013) was used to predict the effect of the SNP of uncertain significance found by Evans (Evans et al., 2005) on local RNA secondary structure of Ex1_205 and of Ex1_202. The selected modes of operation were mode 1 and mode 2, based on global folding (RNAfold) and local folding (RNAplfold), respectively. We used default settings with a folding window of 200 nts.

7.1.6 G-Quadruplex Predictions

G-Quadruplex prediction was performed using the software QGRS Mapper, that allows to identify Quadruplex forming G-Rich Sequence (QGRS) in a nucleotide sequence (Kikin et al., 2006).

7.1.7 Alignments and Conservation Analysis

Multiple sequence alignments were performed using the software Clustal X (Thompson et al., 2002). The generated alignments were visualized through GeneDOC (Nicholas et al., 1997). To have a more detailed landscape of the sequence conservation, alignment files obtained from Clustal X were used as input file in ConSurf (Glaser at al., 2003), and open-source webserver for the identification of conserved local sequences. Consurf allows the operator to calculate, position per position, the degree of conservation of the sequence of interest when compared to other sequences in the alignment. We chose the Maximum Likelihood as the calculation method and GTR (generalized time-reversible) as the evolutionary substitution. The result was visualized through Prism as a plot in which the degree of conservation was shown as color score (1 = not conserved and very variable; 9 = very conserved and very stable) versus nucleotide position. Sliding window of the mean of five nucleotides was used to gain a better interpretation of the result.

7.1.8 CAGE peaks quantification

CAGE peaks quantification was performed on CAGE libraries of different tissues (adult brain, adult kidney, adult testis, adult lung, adult heart, adult liver, and fetal brain

taken by FANTOM 5 web service: (https://fantom.gsc.riken.jp/5/datafiles/latest/basic/human.tissue.hCAGE/). CAGE tags were analyzed using CAGEr package, freely distributed by R/Bioconductor(Haberle et al., 2015). Normalization of the analysis was performed in accordance withBalwierz et al., 2009. The workflow used for our analysis is reported in the appendix (**par. 9.3**).

7.2 Experimental methods

7.2.1 Cell Cultures

SHSY-5Y cells were maintained in RPMI 1640 Media (Gibco), supplied with 10% of Fetal Bovine Serum (FBS, ThermoFisher Scientific), 1% P/S (penicillin and streptomycin), and 1% of Non-Essential Amino Acid (NEAA, Cyagen). Cells were placed in an incubator at 37°C, 5% CO2. When cultures were at confluency, they were collected after trypsinization with a solution of 0.25% trypsin/EDTA (Sigma-Aldrich), after two washes in Phosphate Buffered Saline w/o Ca and Mg (PBS 1X, ThermoFisher Scientific). Cells were plated in wells of multiwell plates MW6 and MW24.

7.2.2 Experimental Animals

All the experiments were performed following the European Community Council Directive 2010/63/UE for the correct care and use of experimental animals, with protocols approved by the San Raffaele Scientific Institutional Animal Care. CD1 mouse strain animals were used in the experiments reported in the thesis.

7.2.3 Primary cultures of mouse cortical neurons

Primary cultures of cortical neurons were prepared from wild-type murine cortex at embryonic days 15 (E15). After dissection, cortices were collected in a 15 mL falcon tube containing ice-cold Dissection Media, composed of Hank's Balanced Salt Solution (HBSS 1X) supplemented with 10 mM HEPES/Na pH 7,4 and 100 u/ml P/S. Tissues were washed three times with fresh Dissection Media at 37° and subsequently incubated in a solution of Dissection Media with 2,5 mg/ul trypsin (Sigma-Aldrich) for 15 minutes. Every 5 minutes a mild agitation to facilitate the digestion was performed. Then, cortices were washed with Wash Media, composed of Neurobasal Media (Gibco cod. 21103049), 5 U/ml P/S and 10% FBS, in order to block the digestion. The second wash was performed with a second Wash Media, prepared as per the first Wash Media but without FBS. After discarding the wash, 2 ml of Neuron Culture Media were added,

and the cortices were mechanically dissociated with the aid of a Pasteur pipette. Neuron Culture Media was composed of Neurobasal medium, supplied with B27 1 X (Gibco 17504044) e 100 u/ml P/S. After the addition of Neuron Culture Media, the obtained solution was filtered using a 70um strainer. Cells were counted with an automatic Counter by Biorad twice and plated in MW6, 400'000 cells per well.

7.2.4 Cellular treatments

Neuronal cultures were treated with the following drugs:

- Ionomycin (Merck 407952), at the concentration of 5uM. Ionomycin powder was dissolved in DMSO and stored at -20°C before use.
- N-Methyl-D-aspartic acid (NMDA, Tocris Cat. No. 2224), at the concentration of 50uM. NMDA powder was dissolved in sterile milliQ water and stored at -20°C before use.
- (-)- Bicuculline methocloride salt (Tocris Cat. No. 0131), at the concentration of 10uM. Bicuculline powder was dissolved in water and stored at -20°C.

All treatments were administrated directly on cell cultures after dilution in 100 ul of media. The incubation period was of 10 minutes. After two washes in PBS 1X, cells were collected as described in 7.2.3.

7.2.5 In vivo Bicuculline Treatment

(+)- Bicuculline (Tocris Cat. No. 0130) was administrated to p30 wild type mice via intraperitoneal injection. Bicuculline powder was dissolved in DMSO (Sigma-Aldrich) and conveyed in a saline solution. Six animals were treated with 1 mg/ kg or 3 mg/kg of the drug, whereas a group of six animals was used as negative control for the experiment and was injected with the vehicle containing the same quantity of DMSO in which bicuculline was dissolved. After thirty minutes from the injection, mice were sacrificed, and cortices were collected in RNAlater (Thermo Fisher AM7020) to preserve the quality of proteins and mRNA of the tissues. Samples were stored at 20°C until experiments were performed.

7.2.6 Small Interfering RNA (siRNA) transfections

eIF4B small interfering RNA transfections were performed using Lipofectamine 3000 Reagent (Life Technologies, L3000001). Briefly, siRNA (100 pmol/ml) was incubated with P300 (4ul) in Opti-MEM Reduced Serum Media (Gibco, 31985062), whereas lipofectamine 3000 (4ul) was incubated in a different eppondorf, with the same quantity

of Opti-MEM. After 10 minutes of incubation at room temperature (RT), the two solutions were mixed and left for 15 minutes at RT. Cell cultures were transfected when arrive at 80% of confluency, for 72h at 37° in a sterile environment. After two washes with sterile PBS 1X, protein and RNA samples were prepared.

eIF4B siRNA was purchased by Dharmacon (ON-TARGETplus Human eIF4B siRNA, SMARTPool format, Catalog ID: L-020179-00-0005).

7.2.7 Protein Extraction and sample preparation

Protein lysis protocols were performed in ice, using RIPA Buffer (50mM Trizma pH 8, 150 mM NaCl, 1% NP40, 0.1% SDS, 0.5% sodium deoxicholate, 5mM EDTA). Cells cultures, plated in MW6, were washes twice and placed on ice. 75ul of RIPA Buffer, supplemented with protease inhibitor mix (CLAP) and phosphatase inhibitor mic (pI) 1:1000. After 5 minutes, cells were detached using a cell scrapper. The resulted solution was collected in an eppendorf and left to rest for 15 minutes on ice. At the end of the incubation period, samples were centrifuged for 10 minutes at 4°C, 12'500 rpm. Then, the supernatant was collected and quantified thought a Micro BCA Protein Assay Kit (Thermo Fisher Scientific, 23235).

The preparation of protein samples from mice tissue followed the same protocol but included a first step of homogenization. Tissues were washed twice in PBS 1X and placed in 500 ul of RIPA Buffer. Homogenization was performed through TissueLyser (Qiagen) with sterile metallic beads. Protein quantification was performed through Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, 23225).

7.2.8 SDS PAGE

After denaturing protein samples for 5 minutes at 95°C with Laemmli buffer 1:4, samples were centrifuged at max speed for 1 minute. Gels were prepared with 7.5% or 12.5% of polyacrylamide, starting from a solution of 30% 37.5:1 acrylamide/bisacrylamide solution (Bio-Rad, 1610158). Running buffer was composed of Tris-glicine 0.025 M, at 8.3 pH, with 0.1% SDS content. The electrophoresis was carried out for 35 minutes at 50 V, then for another 2 hours at 150 V. The protein ladder used in all the gels was SeeBlue™ Plus2 Pre-stained Protein Standard (Thermo Fisher, LC5925). For each sample 20 ug of proteins were loaded.

7.2.9 Western Blot analysis

The gel obtained from the electrophoresis run was transferred on a nitrocellulose membrane (porousness = 0.2μm), using a blotting apparatus by BioRad. The resulting

membranes were incubated with a 0.2% S Ponceau solution in 3% tricloracetic acid. Membranes were carefully decolored with twice-distilled water to correctly show the proteins transfer. After three washes in TBST-T 1X (Tris-based solution containing 0.2% Tween-20) and an hour of blocking with 5% milk or bovine serum albumin, membranes were incubated in a cold room (4°C), putting them on the rocker for 16 hours with primary antibodies. Secondary antibodies were diluted (1:5000) in a TBS-T solution containing 5% milk, and left to the rocker for 1 hour at RT. The detection of protein signals was performed through chemiluminescence, employing PICO ECL (Thermo Scientific, 334579) or FEMTO ECL (Thermo Scientific, 34094). The instrument employed was the ChemiDoc™ Imaging System. See Table 7.2.9 for a complete list of the antibodies used throughout this thesis.

Antibody	Host	Productor	Ref. Code	Dilution
αTubuline	Mouse	Cell Signaling	3873	1:500
βActine	Mouse	Sigma-Aldrich	A3853	1:500
CDKL5	Mouse	Santa Cruz Biotech.	sc-376314	1:1000
GAPDH	Rabbit	Abcam	Ab9485	1:500
eIF4B	Rabbit	Cell Signaling	3592	1:1000
ERK1/2	Rabbit	Cell Signaling	9102	1:1000
PhosphoERK1/2 (Thr202/Tyr204)	Rabbit	Cell Signaling	4377	1:1000
MeCP2	Rabbit	Cell Signaling	D4F3	1:1000
XIAP	Rabbit	Cell Signaling	14334	1:1000

Tab. 7.2.9 List of used Antibodies.

7.2.10 RNA Extraction and Retrotranscription

RNA Extraction was performed using PureZOL RNA isolation agent (Bio-Rad, 7326880), after the careful cleaning of the instruments with RNAseZAP (Invitrogen, AM9782). Cell cultures plated in MW6 were gently washed with PBS 1X before adding

300ul of PureZOL. The resulting solution was collected in a 1.5 ml Eppendorf tube. After 5 minutes of dissociation, 60 ul of chloroform was added. After vortexing the Eppendorf tube for 15 seconds and a brief incubation at RT, the solution was centrifuge (12'500 rcf, 15 minutes, 4°C). The resulted colorless aqueous phase was added to glycogen RNA grade (Thermo Scientific, R0551) and to 150ul of isopropyl alcohol. Samples were stored overnight at -20°C. The following day centrifugation (1200 g, 10 minutes, 4°C) was performed to precipitate the RNA. After washing with Ethanol 70% and millique water, the pellet was air dried and resuspended in RNAse free water. The same protocol was used to extract RNA from tissues, however an extra homogenization step with TissueLyser in PUREzol solution (500ul) was performed at the beginning. After extraction, RNA was quantified with Nanodrop (Thermo Fisher Scientific) and retrotranscribed with RT2 Easy First Strand Kit (Qiagen, Cat. No. / ID: 330421).

7.2.11 Real-Time qRT-PCR and RT-PCR

Real-Time qRT-PCR was performed using SYBR Green as detection chemistry. The quantitation approach used was the comparative Ct method, employing as reference genes GAPDH and RPL13. 9ng of cDNA was used for the amplification, for 30 cycles. CDKL5 murine primers was designed using NetPrimer (PREMIER Biosoft) and Benchling (https://benchling.com).

Here is a list of the primers used for Real-Time qRT-PCR.

- mCDKL5 FORWARD 5' – CCACTGGTGCCACAAGAACGACA - 3'
- mCDKL5 REVERSE 5' – TGCCCTCTGAGAGATTGCGAGC – 3'
- mGAPDH FORWARD 5' - AGGTCGGTGTGAACGGATTTG - 3'
- mGAPDH REVERSE 5' - TGTAGACCATGTAGTTGAGGTCA - 3'
- mRPL13 FORWARD 5' - TGGCTGGCATCCACAAGAAA - 3'
- mRPL13 REVERSE 5' - TTCTTCAGCAGAACTGTCTCCC - 3'

RT-PCR was used to analyze the presence of alternative Ex1_202 from samples obtained via different biological sources: wild-type p30 mice cortex and SHSY-5Y cells. PCR reaction was carried out in 50ul of reaction solution, starting from an initial cDNA amount of 100 ng. The amplification involved the transcripts CDKL5 and GAPDH. Below the list of primers used for the amplification of murine cDNA.

- mCDKL5 TOP1 5' – TACTTGTCGCTGCCGCTAGGGA – 3'
- mCDKL5 TOP2 5' – GCTCCGGCGAGAGGGCGGGG – 3'
- mCDKL5 TOP3 5' – GCAGACGGGGGCGGTGCGA – 3'
- mCDKL5 REVERSE 5' – TAATGTCCCAACGAAGAAATTCTC – 3'

- mGAPDH FORWARD 5’ - AGGTCGGTGTGAACGGATTTG - 3’
- mGAPDH REVERSE 5’ - TGTAGACCATGTAGTTGAGGTCA - 3’

Primers used for the amplification of human cDNA are the following:

- hCDKL5 TOP1 5’- TAGTTGTCTCTGCCGCTGGGGA-3’
- hCDKL5 TOP2 5’ -CTTCTGCTAGAGGGCGGGG -3’
- hCDKL5 TOP3 5’- GCTGGGGCGGGGCAGTTAG -3’
- hCDKL5 REVERSE 5’- CACTGGTTGGTGGGAACTTTCAC- 3’
- hGAPDH FORWARD 5’ – GTCTCCTCTGACTTCAACAGCG – 3’
- hGAPDH REVERSE 5’ – ACCACCCTGTTGCTGTAGCCAA – 3’

We collected samples (7ul) for GAPDH at 27th cycle, while CDKL5 7ul was picked up at cycle 35. The amplification products were loaded on a 2% agarose gel and run at 80V. The amplification bands were visualized through Bio-RED Gel Doc XR. After the run, amplification products were purified from the gel and sequenced using the LightRun Barcodes from Eurofins.

7.2.12 Two-promoter vectors

Synbio Technology synthesized the sequences of interest and cloned them in the pBRm2L vector. Here the list of plasmids used in the experiments presented in this thesis:

pBRm2L_Ex1_205, containing the CDKL5 5’UTR reported in RefSeq NM_001323289.2 (250nts) between the second T7 promoter and the firefly luciferase cistron:

5 - AGTTGTCTCTGCCGCTGGGGAAGGTAAAGCGGCGACGGCGTCCTC
AGGAGCTGTGGGGTCCCCTGCTAGAAGTGGGGGACTCGGCGGGGGAG
TCATTTAATACTTCATGATTAGAACAAATATGTGAAAGTTCCCACCAA
CCAGTGAGAATTTCTTCCTTCAGACGGTTTTGGATCTTACTGCACAGCT
TTCTGAGAAGTTCTTTTGGTGCCATGTTTTGTGGCTTGCATCAAAAGAG
GAGTTTGTCTTC – 3

pBRm2L_Ex1_202up, containing in the same position the first exon of CDKL5 TV 202 (285nts):

5 - GCTTCTGCTAGAGGGCGGGGCCGGAGGTTTCGATTAGTTGTCTCTG
CCGCTGGGGAAGGTAAAGCGGCGACGGCGTCCTCAGGAGCTGTGGGG
TCCCCTGCTAGAAGTGGGGGACTCGGCGGGGGAGTCATTTAATACTTC

ATGATTAGAACAAATATGTGAAAGTTCCCACCAACCAGTGAGAATTTC
TTCCTTCAGACGGTTTTGGATCTTACTGCACAGCTTTCTGAGAAGTTCT
TTTGGTGCCATGTTTTGTGGCTTGCATCAAAAGAGGAGTTTGTCTTC - 3

pBRm2L_UTex2, containing in the same position the untranslated region of exon 2 of CDKL5 (162nts):

5 – GGAGTCATTTAATACTTCATGATTAGAACAAATATGTGAAAGTTC
CCACCAACCAGTGAGAATTTCTTCCTTCAGACGGTTTTGGATCTTACTG
CACAGCTTTCTGAGAAGTTCTTTTGGTGCCATGTTTTGTGGCTTGCATC
AAAAGAGGAGTTTGTCTTC – 3

pBRm2L_EVA, containing in the same position the sequence indicated for the vector Ex1_202up, but reporting the SNP rs786204994 (Evans et al., 2005) – here highlighted in bold:

5 - GCTTCTGCTAGAGGGCGGGGCCGGAGGTTTCGATTAGTTGTCTCTG
CCGCTGGGGAAGGTAAAGCGGCGACGGCGTCCTCAGGAGCTGTGGGG
TC**T**CCTGCTAGAAGTGGGGGACTCGGCGGGGGAGTCATTTAATACTTC
ATGATTAGAACAAATATGTGAAAGTTCCCACCAACCAGTGAGAATTTC
TTCCTTCAGACGGTTTTGGATCTTACTGCACAGCTTTCTGAGAAGTTCT
TTTGGTGCCATGTTTTGTGGCTTGCATCAAAAGAGGAGTTTGTCTTC - 3

The map of the original empty pBRm2L vector was reported in par. 9.4 and it was used as control for the assay. pBRm2L_BACE - a plasmid carrying the 5'UTR of the gene BACE1 and demonstrated to be functional (Lammich et al., 2004), was used as an extra control of the goodness of the analysis.

7.2.13 Dicistronic Vectors

Inserts of pBRm2L vectors were amplified with primers, adding to the ends of the amplicons SalI and NcoI restriction sites.

- Ex1205fSalI: 5' – TTCCGTCGACAGTTGTCTCTGCCG – 3'
- Ex1202fSalI: 5' – TTCCGTCGACGCTTCTGCTAGAGG – 3'
- Utex2SalI: 5' – TTCCGTCGACGGAGTCATTTAATACTTC - 3'
- UTex2NcoI: 5' – TTCCCCATGGAAGACAAACTCCTCTTTTGATG - 3'

After purification of the amplicons, they were digested with both the enzymes (purchased by NEB), for 2h at 37°C. At the same time, the original pBAT was cut with the same enzymes, subjected to SAP reaction (rSAP, NEB) and purified through gel extraction. Then, ligation was performed at RT for 1h with Quick Ligation Kit (NEB).

The new vectors were verified through sequencing with LightRun Barcodes (Eurofins Genomics).

Primers designed for sequencing of the vectors: Primers designed for sequencing of the vectors:

- nRLSEQf: 5' - AATATATCAAATCGTTCGTTGAG - 3'
- nFLSEQr: 5' - CATCTTCCAGCGGATAGAATG - 3'

7.2.14 MVA-T7 virus

The modified vaccinia virus Ankara strain recombinant for T7 RNA polymerase (MVA-T7) was originally provided by Gerd Sutter (Institute of Molecular Virology of the GSF-Forschungszentrum fuer Umwelt und Gesundheit GmbH, Oberschleissheim, Germany) to Daniele Zacchetti and the stock was prepared by replication in chicken embryo fibroblast (Sutter et al, 1995; De Pietri Tonelli et al, 2003 and 2004).

Cells between 70-80% of confluency were washed twice with sterile PBS 1X and infected by adding 10ul of prepared virus. After this step, cells were incubated in the incubator for 30 minutes, before proceeding with further washes and plasmid transfection.

7.2.15 Vector Transfection

Vector transfections were performed using Lipofectamine 3000 Reagent (Life Technologies, L3000001), following the official protocol. Vectors (500ng) and P3000 reagent (1ul) were incubated together with 50 ul of Opti-MEM for 10 minutes, while Lipofectamine 3000 (1ul) was incubated separately in 50 ul of Opti-MEM for the same amount of time. At the end of the first incubation, a master mix was made up mixing the two preparations. It was incubated for 10 minutes. After the second incubation, the mix was used to transfect SHSY-5Y cells plated in a MW24 (100ul per well). Cells were incubated for six hours before being utilized in further experimental steps.

7.2.16 Dual Luciferase Reporter Assay

After six hours of plasmid transfection, cells were washed twice and were lysed by using 100 ul Passive Lysis Buffer 1X (PLB 1X), supplied by the Dual-Luciferase® Reporter Assay (Promega, E1910). PLB promotes a rapid lysis of cultured mammalian cells without the need to scrape. After 15 minutes of incubation at RT, lysates were

harvested in a 1.5ml tube. For each measurement, 20ul of total lysate were used and prepared in a new tube. The tube was inserted inside the proper chamber in the GloMax 20/20 luminometer (Promega), where injector 1 injected 100 ul of LARII Solution at room temperature; injector 2 then proceeded by injecting 100 ul of Stop&Glo Reagent (prepared as indicated in the official protocol). The program of the luminometer was set considering 2 seconds of pre-measurement and 10 seconds of integration. Five samples of 20 ul PLB 1X were measured to assess the background of the measurements. Each experiment was designed to have at least two biological replicates per MW24. Each biological replicate was measured twice (to have two technical replicates) and the remaining sample was stored at -20°C in case of further measurements.

7.2.17 Circular Dichroism Spectroscopy

The following RNA oligonucleotide, called CDKL51, was purchased from Sigma-Aldrich: 5' – GCUUCUGCUAGAGGGCGGGGCCGGAGGUUUCGAUU – 3'.
The oligonucleotide sample (5uM) was prepared in RNase-free water with 10mM Tris/HCl, 0.1 EDTA, at pH 7.4 in a final volume of 200ul. The sample was annealed by heating at 90°C for 10 minutes, following by slow cooling to 20°C at a constant rate of 2°C/min. CD measurements were performed using Jasco-J100 spectropolarimeter equipped with a Peltier temperature cooler. Samples were inserted in a 0.1-cm cell and scanned at a speed of 50nm/min. The spectra were obtained through an average of 9 scans from 200 to 320 nm. A buffer-specific baseline was considered for each different sample. The used salts were KCl, NaCl and LiCl at 1mM. Spectra were analyzed using the software SpectraGryph 1.2 (Menges 2016) and Prism.

7.3 Data Analysis

All the data analysis and the relative plots presented in this thesis - unless otherwise specified – were performed with the software Prism from GraphPad. Student's t-test for unpaired samples was performed to evaluate the statistical significance of the results.In plots, error bars represent the Standard Error of the Mean (SEM). In addition, one-way ANOVA, followed by Tukey's test, was performed to compare CDKL5 protein expression level.

8 REFERENCES

Abramson RD, Dever TE, Lawson TG, Ray BK, Thach RE, Merrick WC. The ATP-dependent interaction of eukaryotic initiation factors with mRNA. J Biol Chem. 1987 Mar 15;262(8):3826-32. PMID: 2950099.

Agarwala P, Pandey S, Ekka MK, Chakraborty D, Maiti S. Combinatorial role of two G-quadruplexes in 5' UTR of transforming growth factor β2 (TGFβ2). Biochim Biophys Acta Gen Subj. 2019 Nov;1863(11):129416. doi: 10.1016/j.bbagen.2019.129416. Epub 2019 Aug 16. PMID: 31425729.

Agarwala P, Pandey S, Maiti S. The tale of RNA G-quadruplex. Org Biomol Chem. 2015 May 28;13(20):5570-85. doi: 10.1039/c4ob02681k. PMID: 25879384.

Amendola E, Zhan Y, Mattucci C, Castroflorio E, Calcagno E, Fuchs C, Lonetti G, Silingardi D, Vyssotski AL, Farley D, Ciani E, Pizzorusso T, Giustetto M, Gross CT. Mapping pathological phenotypes in a mouse model of CDKL5 disorder. PLoS One. 2014 May 16;9(5):e91613. doi: 10.1371/journal.pone.0091613. PMID: 24838000; PMCID: PMC4023934.

Andreev DE, O'Connor PB, Fahey C, Kenny EM, Terenin IM, Dmitriev SE, Cormican P, Morris DW, Shatsky IN, Baranov PV. Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. Elife. 2015 Jan 26;4:e03971. doi: 10.7554/eLife.03971. PMID: 25621764; PMCID: PMC4383229.

Araujo PR, Yoon K, Ko D, Smith AD, Qiao M, Suresh U, Burns SC, Penalva LO. Before It Gets Started: Regulating Translation at the 5' UTR. Comp Funct Genomics. 2012;2012:475731. doi: 10.1155/2012/475731. Epub 2012 May 28. PMID: 22693426; PMCID: PMC3368165.

Archer HL, Evans JC, Millar DS, Thompson PW, Kerr AM, Leonard H, Christodoulou J, Ravine D, Lazarou L, Grove L, Verity C, Whatley SD, Pilz DT, Sampson JR, Clarke AJ. NTNG1 mutations are a rare cause of Rett syndrome. Am J Med Genet A. 2006 Apr 1;140(7):691-4. doi: 10.1002/ajmg.a.31133. PMID: 16502428; PMCID: PMC2577736.

Arora A, Dutkiewicz M, Scaria V, Hariharan M, Maiti S, Kurreck J. Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. RNA. 2008 Jul;14(7):1290-6. doi: 10.1261/rna.1001708. Epub 2008 May 30. PMID: 18515550; PMCID: PMC2441988.

Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 2010 Jul;38(Web Server issue):W529-33. doi: 10.1093/nar/gkq399. Epub 2010 May 16. PMID: 20478830; PMCID: PMC2896094.

Audigier S, Guiramand J, Prado-Lourenco L, Conte C, Gonzalez-Herrera IG, Cohen-Solal C, Récasens M, Prats AC. Potent activation of FGF-2 IRES-dependent

mechanism of translation during brain development. RNA. 2008 Sep;14(9):1852-64. doi: 10.1261/rna.790608. Epub 2008 Aug 1. PMID: 18676616; PMCID: PMC2525950.

Babendure JR, Babendure JL, Ding JH, Tsien RY. Control of mammalian translation by mRNA structure near caps. RNA. 2006 May;12(5):851-61. doi: 10.1261/rna.2309906. Epub 2006 Mar 15. PMID: 16540693; PMCID: PMC1440912.

Bagga JS, D'Antonio LA. Role of conserved cis-regulatory elements in the post-transcriptional regulation of the human MECP2 gene involved in autism. Hum Genomics. 2013 Sep 16;7(1):19. doi: 10.1186/1479-7364-7-19. PMID: 24040966; PMCID: PMC3844687.

Bahi-Buisson N, Bienvenu T. CDKL5-Related Disorders: From Clinical Description to Molecular Genetics. Mol Syndromol. 2012 Apr;2(3-5):137-152. doi: 10.1159/000331333. Epub 2011 Sep 13. PMID: 22670135; PMCID: PMC3366705.

Bahi-Buisson N, Girard B, Gautier A, Nectoux J, Fichou Y, Saillour Y, Poirier K, Chelly J, Bienvenu T. Epileptic encephalopathy in a girl with an interstitial deletion of Xp22 comprising promoter and exon 1 of the CDKL5 gene. Am J Med Genet B Neuropsychiatr Genet. 2010 Jan 5;153B(1):202-7. doi: 10.1002/ajmg.b.30974. PMID: 19455595.

Balestra D, Giorgio D, Bizzotto M, Fazzari M, Ben Zeev B, Pinotti M, Landsberger N, Frasca A. Splicing Mutations Impairing CDKL5 Expression and Activity Can be Efficiently Rescued by U1snRNA-Based Therapy. Int J Mol Sci. 2019 Aug 24;20(17):4130. doi: 10.3390/ijms20174130. PMID: 31450582; PMCID: PMC6747535.

Baltussen LL, Negraes PD, Silvestre M, Claxton S, Moeskops M, Christodoulou E, Flynn HR, Snijders AP, Muotri AR, Ultanir SK. Chemical genetic identification of CDKL5 substrates reveals its role in neuronal microtubule dynamics. EMBO J. 2018 Dec 14;37(24):e99763. doi: 10.15252/embj.201899763. Epub 2018 Sep 28. PMID: 30266824; PMCID: PMC6293278.

Balwierz PJ, Carninci P, Daub CO, Kawai J, Hayashizaki Y, Van Belle W, Beisel C, van Nimwegen E. Methods for analyzing deep sequencing expression data:constructing the human and mouse promoterome with deepCAGE data. Genome Biol. 2009;10(7):R79. doi: 10.1186/gb-2009-10-7-r79. Epub 2009 Jul 22. PMID: 19624849; PMCID: PMC2728533.

Barbiero I, De Rosa R, Kilstrup-Nielsen C. Microtubules: A Key to Understand and Correct Neuronal Defects in CDKL5 Deficiency Disorder? Int J Mol Sci. 2019 Aug 21;20(17):4075. doi: 10.3390/ijms20174075. PMID: 31438497; PMCID: PMC6747382.

Barbiero I, Peroni D, Tramarin M, Chandola C, Rusconi L, Landsberger N, Kilstrup-Nielsen C. The neurosteroid pregnenolone reverts microtubule derangement induced by the loss of a functional CDKL5-IQGAP1 complex. Hum Mol Genet. 2017 Sep 15;26(18):3520-3530. doi: 10.1093/hmg/ddx237. PMID: 28641386.

Barbosa C, Peixeiro I, Romão L. Gene expression regulation by upstream open reading

frames and human disease. PLoS Genet. 2013;9(8):e1003529. doi: 10.1371/journal.pgen.1003529. Epub 2013 Aug 8. PMID: 23950723; PMCID: PMC3738444.

Bartnik M, Derwińska K, Gos M, Obersztyn E, Kołodziejska KE, Erez A, Szpecht-Potocka A, Fang P, Terczyńska I, Mierzewska H, Lohr NJ, Bellus GA, Reimschisel T, Bocian E, Mazurczak T, Cheung SW, Stankiewicz P. Early-onset seizures due to mosaic exonic deletions of CDKL5 in a male and two females. Genet Med. 2011 May;13(5):447-52. doi: 10.1097/GIM.0b013e31820605f5. PMID: 21293276.

Bassell GJ, Warren ST. Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function. Neuron. 2008 Oct 23;60(2):201-14. doi: 10.1016/j.neuron.2008.10.004. PMID: 18957214; PMCID: PMC3691995.

Belsham GJ, Jackson RJ. Translation initiation on picornavirus RNA. In: Sonenberg N, Hershey JWB, Mathews MB, editors. Translational Control of Gene Expression. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (2000). p. 869–900.

Ben-Asouli Y, Banai Y, Pel-Or Y, Shir A, Kaempfer R. Human interferon-gamma mRNA autoregulates its translation through a pseudoknot that activates the interferon-inducible protein kinase PKR. Cell. 2002 Jan 25;108(2):221-32. doi: 10.1016/s0092-8674(02)00616-5. PMID: 11832212.

Bennett CF. Therapeutic Antisense Oligonucleotides Are Coming of Age. Annu Rev Med. 2019 Jan 27;70:307-321. doi: 10.1146/annurev-med-041217-010829. PMID: 30691367.

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW. GenBank. Nucleic Acids Res. 2018 Jan 4;46(D1):D41-D47. doi: 10.1093/nar/gkx1094. PMID: 29140468; PMCID: PMC5753231.

Benson MD, Pandey S, Witchell D, Jazayeri A, Siwkowski A, Monia B, Kluve-Beckerman B. Antisense oligonucleotide therapy for TTR amyloidosis. Amyloid. 2011 Jun;18 Suppl 1:60. doi: 10.3109/13506129.2011.574354021. Erratum in: Amyloid. 2011 Jun;18 Suppl 1:60. PMID: 21838433.

Bertani I, Rusconi L, Bolognese F, Forlani G, Conca B, De Monte L, Badaracco G, Landsberger N, Kilstrup-Nielsen C. Functional consequences of mutations in CDKL5, an X-linked gene involved in infantile spasms and mental retardation. J Biol Chem. 2006 Oct 20;281(42):32048-56. doi: 10.1074/jbc.M606325200. Epub 2006 Aug 24. PMID: 16935860.

Bettegazzi B, Bellani S, Roncon P, Guarnieri FC, Bertero A, Codazzi F, Valtorta F, Simonato M, Grohovaz F, Zacchetti D. eIF4B phosphorylation at Ser504 links synaptic activity with protein translation in physiology and pathology. Sci Rep. 2017 Sep 5;7(1):10563. doi: 10.1038/s41598-017-11096-1. PMID: 28874824; PMCID: PMC5585320.

Bettegazzi B, Sebastian Monasor L, Bellani S, Codazzi F, Restelli LM, Colombo AV, Deigendesch N, Frank S, Saito T, Saido TC, Lammich S, Tahirovic S, Grohovaz F, Zacchetti D. Casein Kinase 2 dependent phosphorylation of eIF4B regulates BACE1

expression in Alzheimer's disease. Cell Death Dis. 2021 Aug 4;12(8):769. doi: 10.1038/s41419-021-04062-3. PMID: 34349120; PMCID: PMC8339060.

Biswas J, Liu Y, Singer RH, Wu B. Fluorescence Imaging Methods to Investigate Translation in Single Cells. Cold Spring Harb Perspect Biol. 2019 Apr 1;11(4):a032722. doi: 10.1101/cshperspect.a032722. PMID: 30082468; PMCID: PMC6442198.

Blais JD, Filipenko V, Bi M, Harding HP, Ron D, Koumenis C, Wouters BG, Bell JC. Activating transcription factor 4 is translationally regulated by hypoxic stress. Mol Cell Biol. 2004 Sep;24(17):7469-82. doi: 10.1128/MCB.24.17.7469-7482.2004. PMID: 15314157; PMCID: PMC506979.

Bolger TA, Yao TP. Intracellular trafficking of histone deacetylase 4 regulates neuronal cell death. J Neurosci. 2005 Oct 12;25(41):9544-53. doi: 10.1523/JNEUROSCI.1826-05.2005. PMID: 16221865; PMCID: PMC6725694.

Bonnal S, Pileur F, Orsini C, Parker F, Pujol F, Prats AC, Vagner S. Heterogeneous nuclear ribonucleoprotein A1 is a novel internal ribosome entry site trans-acting factor that modulates alternative initiation of translation of the fibroblast growth factor 2 mRNA. J Biol Chem. 2005 Feb 11;280(6):4144-53. doi: 10.1074/jbc.M411492200. Epub 2004 Nov 3. PMID: 15525641.

Borden KLB, Volpon L. The diversity, plasticity, and adaptability of cap-dependent translation initiation and the associated machinery. RNA Biol. 2020 Sep;17(9):1239-1251. doi: 10.1080/15476286.2020.1766179. Epub 2020 Jun 4. PMID: 32496897; PMCID: PMC7549709.

Borg I, Freude K, Kübart S, Hoffmann K, Menzel C, Laccone F, Firth H, Ferguson-Smith MA, Tommerup N, Ropers HH, Sargan D, Kalscheuer VM. Disruption of Netrin G1 by a balanced chromosome translocation in a girl with Rett syndrome. Eur J Hum Genet. 2005 Aug;13(8):921-7. doi: 10.1038/sj.ejhg.5201429. PMID: 15870826.

Bramham CR, Worley PF, Moore MJ, Guzowski JF. The immediate early gene arc/arg3.1: regulation, mechanisms, and function. J Neurosci. 2008 Nov 12;28(46):11760-7. doi: 10.1523/JNEUROSCI.3864-08.2008. PMID: 19005037; PMCID: PMC2615463.

Buffington SA, Huang W, Costa-Mattioli M. Translational control in synaptic plasticity and cognitive dysfunction. Annu Rev Neurosci. 2014;37:17-38. doi: 10.1146/annurev-neuro-071013-014100. PMID: 25032491; PMCID: PMC4721605.

Bugaut A, Balasubramanian S. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. Nucleic Acids Res. 2012 Jun;40(11):4727-41. doi: 10.1093/nar/gks068. Epub 2012 Feb 20. PMID: 22351747; PMCID: PMC3367173.

Burgess DJ. The TOPMed genomic resource for human health. Nat Rev Genet. 2021 Apr;22(4):200. doi: 10.1038/s41576-021-00343-x. PMID: 33654294.

Buttgereit F, Brand MD. A hierarchy of ATP-consuming processes in mammalian cells. Biochem J. 1995 Nov 15;312 (Pt 1)(Pt 1):163-7. doi: 10.1042/bj3120163. PMID: 7492307; PMCID: PMC1136240.

Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci U S A. 2009 May 5;106(18):7507-12. doi: 10.1073/pnas.0810916106. Epub 2009 Apr 16. PMID: 19372376; PMCID: PMC2669787.

Cao H, Chén OY, Chung Y, Forsyth JK, McEwen SC, Gee DG, Bearden CE, Addington J, Goodyear B, Cadenhead KS, Mirzakhanian H, Cornblatt BA, Carrión RE, Mathalon DH, McGlashan TH, Perkins DO, Belger A, Seidman LJ, Thermenos H, Tsuang MT, van Erp TGM, Walker EF, Hamann S, Anticevic A, Woods SW, Cannon TD. Cerebello-thalamo-cortical hyperconnectivity as a state-independent functional neural signature for psychosis prediction and characterization. Nat Commun. 2018 Sep 21;9(1):3836. doi: 10.1038/s41467-018-06350-7. PMID: 30242220; PMCID: PMC6155100.

Carouge D, Host L, Aunis D, Zwiller J, Anglard P. CDKL5 is a brain MeCP2 target gene regulated by DNA methylation. Neurobiol Dis. 2010 Jun;38(3):414-24. doi: 10.1016/j.nbd.2010.02.014. Epub 2010 Mar 6. PMID: 20211261.

Chahil G, Bollu PC. Rett Syndrome. 2021 Aug 11. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan–. PMID: 29489169.

Chan SW. Establishment of chronic hepatitis C virus infection: translational evasion of oxidative defence. World J Gastroenterol. 2014 Mar 21;20(11):2785-800. doi: 10.3748/wjg.v20.i11.2785. PMID: 24659872; PMCID: PMC3961964.

Chang LS, Oblinger JL, Burns SS, Huang J, Anderson LW, Hollingshead MG, Shen R, Pan L, Agarwal G, Ren Y, Roberts RD, O'Keefe BR, Kinghorn AD, Collins JM. Targeting Protein Translation by Rocaglamide and Didesmethylrocaglamide to Treat MPNST and Other Sarcomas. Mol Cancer Ther. 2020 Mar;19(3):731-741. doi: 10.1158/1535-7163.MCT-19-0809. Epub 2019 Dec 17. PMID: 31848295; PMCID: PMC7056570.

Chappell SA, LeQuesne JP, Paulin FE, deSchoolmeester ML, Stoneley M, Soutar RL, Ralston SH, Helfrich MH, Willis AE. A mutation in the c-myc-IRES leads to enhanced internal ribosome entry in multiple myeloma: a novel mechanism of oncogene de-regulation. Oncogene. 2000 Sep 7;19(38):4437-40. doi: 10.1038/sj.onc.1203791. PMID: 10980620.

Chemudupati M, Kenney AD, Bonifati S, Zani A, McMichael TM, Wu L, Yount JS. From APOBEC to ZAP: Diverse mechanisms used by cellular restriction factors to inhibit virus infections. Biochim Biophys Acta Mol Cell Res. 2019 Mar;1866(3):382-394. doi: 10.1016/j.bbamcr.2018.09.012. Epub 2018 Oct 2. PMID: 30290238; PMCID: PMC6334645.

Chen Q, Zhu YC, Yu J, Miao S, Zheng J, Xu L, Zhou Y, Li D, Zhang C, Tao J, Xiong ZQ. CDKL5, a protein associated with rett syndrome, regulates neuronal morphogenesis via Rac1 signaling. J Neurosci. 2010 Sep 22;30(38):12777-86. doi: 10.1523/JNEUROSCI.1102-10.2010. PMID: 20861382; PMCID: PMC6633570.

Chen Q, Zhu YC, Yu J, Miao S, Zheng J, Xu L, Zhou Y, Li D, Zhang C, Tao J, Xiong

ZQ. CDKL5, a protein associated with rett syndrome, regulates neuronal morphogenesis via Rac1 signaling. J Neurosci. 2010 Sep 22;30(38):12777-86. doi: 10.1523/JNEUROSCI.1102-10.2010. PMID: 20861382; PMCID: PMC6633570.

Cheung YN, Maag D, Mitchell SF, Fekete CA, Algire MA, Takacs JE, Shirokikh N, Pestova T, Lorsch JR, Hinnebusch AG. Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection in vivo. Genes Dev. 2007 May 15;21(10):1217-30. doi: 10.1101/gad.1528307. PMID: 17504939; PMCID: PMC1865493.

Choi JH, Kim SH, Jeong YH, Kim SW, Min KT, Kim KT. hnRNP Q Regulates Internal Ribosome Entry Site-Mediated fmr1 Translation in Neurons. Mol Cell Biol. 2019 Feb 4;39(4):e00371-18. doi: 10.1128/MCB.00371-18. PMID: 30478144; PMCID: PMC6362317.

Cogoi S, Shchekotikhin AE, Xodo LE. HRAS is silenced by two neighboring G-quadruplexes and activated by MAZ, a zinc-finger transcription factor with DNA unfolding property. Nucleic Acids Res. 2014 Jul;42(13):8379-88. doi: 10.1093/nar/gku574. Epub 2014 Jul 10. PMID: 25013182; PMCID: PMC4117790.

Cohen-Chalamish S, Hasson A, Weinberg D, Namer LS, Banai Y, Osman F, Kaempfer R. Dynamic refolding of IFN-gamma mRNA enables it to function as PKR activator and translation template. Nat Chem Biol. 2009 Dec;5(12):896-903. doi: 10.1038/nchembio.234. Epub 2009 Oct 4. PMID: 19801993.

Coldwell MJ, Mitchell SA, Stoneley M, MacFarlane M, Willis AE. Initiation of Apaf-1 translation by internal ribosome entry. Oncogene. 2000 Feb 17;19(7):899-905. doi: 10.1038/sj.onc.1203407. PMID: 10702798.

Comery TA, Harris JB, Willems PJ, Oostra BA, Irwin SA, Weiler IJ, Greenough WT. Abnormal dendritic spines in fragile X knockout mice: maturation and pruning deficits. Proc Natl Acad Sci U S A. 1997 May 13;94(10):5401-4. doi: 10.1073/pnas.94.10.5401. PMID: 9144249; PMCID: PMC24690.

Conte C, Ainaoui N, Delluc-Clavières A, Khoury MP, Azar R, Pujol F, Martineau Y, Pyronnet S, Prats AC. Fibroblast growth factor 1 induced during myogenesis by a transcription-translation coupling mechanism. Nucleic Acids Res. 2009 Sep;37(16):5267-78. doi: 10.1093/nar/gkp550. Epub 2009 Jun 26. PMID: 19561198; PMCID: PMC2760804.

Crooke ST. Molecular Mechanisms of Antisense Oligonucleotides. Nucleic Acid Ther. 2017 Apr;27(2):70-77. doi: 10.1089/nat.2016.0656. Epub 2017 Jan 12. PMID: 28080221; PMCID: PMC5372764.

Cunningham TA, Chapman E, Schatz JH. eIF4A inhibition: ready for primetime? Oncotarget. 2018 Oct 30;9(85):35515-35516. doi: 10.18632/oncotarget.26268. PMID: 30473746; PMCID: PMC6238976.

Córdova-Fletes C, Rademacher N, Müller I, Mundo-Ayala JN, Morales-Jeanhs EA, García-Ortiz JE, León-Gil A, Rivera H, Domínguez MG, Kalscheuer VM. CDKL5 truncation due to a t(X;2)(p22.1;p25.3) in a girl with X-linked infantile spasm syndrome.

Clin Genet. 2010 Jan;77(1):92-6.

Davuluri RV, Suzuki Y, Sugano S, Zhang MQ. CART classification of human 5' UTR sequences. Genome Res. 2000 Nov;10(11):1807-16. doi: 10.1101/gr.gr-1460r. PMID: 11076865; PMCID: PMC310970.

De Pietri Tonelli D, Mihailovich M, Di Cesare A, Codazzi F, Grohovaz F, Zacchetti D. Translational regulation of BACE-1 expression in neuronal and non-neuronal cells. Nucleic Acids Res. 2004 Mar 19;32(5):1808-17. doi: 10.1093/nar/gkh348. PMID: 15034149; PMCID: PMC390341.

De Pietri Tonelli D, Mihailovich M, Schnurbus R, Pesole G, Grohovaz F, Zacchetti D. Translational control of Scamper expression via a cell-specific internal ribosome entry site. Nucleic Acids Res. 2003 May 15;31(10):2508-13. doi: 10.1093/nar/gkg357. PMID: 12736299; PMCID: PMC156039.

Dever TE, Dinman JD, Green R. Translation Elongation and Recoding in Eukaryotes. Cold Spring Harb Perspect Biol. 2018 Aug 1;10(8):a032649. doi: 10.1101/cshperspect.a032649. PMID: 29610120; PMCID: PMC6071482.

Ding Q, Markesbery WR, Chen Q, Li F, Keller JN. Ribosome dysfunction is an early event in Alzheimer's disease. J Neurosci. 2005 Oct 5;25(40):9171-5. doi: 10.1523/JNEUROSCI.3040-05.2005. Erratum in: J Neurosci. 2006 Mar 15;26(11):3077. PMID: 16207876; PMCID: PMC6725754.

Dobbyn HC, Hill K, Hamilton TL, Spriggs KA, Pickering BM, Coldwell MJ, de Moor CH, Bushell M, Willis AE. Regulation of BAG-1 IRES-mediated translation following chemotoxic stress. Oncogene. 2008 Feb 14;27(8):1167-74. doi: 10.1038/sj.onc.1210723. Epub 2007 Aug 13. PMID: 17700523; PMCID: PMC2570717.

Doyle M, Kiebler MA. Mechanisms of dendritic mRNA transport and its role in synaptic tagging. EMBO J. 2011 Aug 31;30(17):3540-52. doi: 10.1038/emboj.2011.278. PMID: 21878995; PMCID: PMC3181491.

Dresios J, Chappell SA, Zhou W, Mauro VP. An mRNA-rRNA base-pairing mechanism for translation initiation in eukaryotes. Nat Struct Mol Biol. 2006 Jan;13(1):30-4. doi: 10.1038/nsmb1031. Epub 2005 Dec 11. PMID: 16341227.

Ducani C, Bernardinelli G, Högberg B, Keppler BK, Terenzi A. Interplay of Three G-Quadruplex Units in the KIT Promoter. J Am Chem Soc. 2019 Jul 3;141(26):10205-10213. doi: 10.1021/jacs.8b12753. Epub 2019 Jun 20. PMID: 31244182.

EBOUE BONIS D, CHAMBAUT AM, VOLFIN P, CLAUSER H. ACTION OF INSULIN ON THE ISOLATED RAT DIAPHRAGM IN THE PRESENCE OF ACTINOMYCIN D AND PUROMYCIN. Nature. 1963 Sep 21;199:1183-4. doi: 10.1038/1991183a0. PMID: 14072040.

Eliscovich C, Buxbaum AR, Katz ZB, Singer RH. mRNA on the move: the road to its biological destiny. J Biol Chem. 2013 Jul 12;288(28):20361-8. doi: 10.1074/jbc.R113.452094. Epub 2013 May 28. PMID: 23720759; PMCID: PMC3711302.

Evans JC, Archer HL, Colley JP, Ravn K, Nielsen JB, Kerr A, Williams E, Christodoulou J, Gécz J, Jardine PE, Wright MJ, Pilz DT, Lazarou L, Cooper DN, Sampson JR, Butler R, Whatley SD, Clarke AJ. Early onset seizures and Rett-like features associated with mutations in CDKL5. Eur J Hum Genet. 2005 Oct;13(10):1113-20. doi: 10.1038/sj.ejhg.5201451. PMID: 16015284.

Fan H, Penman S. Regulation of protein synthesis in mammalian cells. II. Inhibition of protein synthesis at the level of initiation during mitosis. J Mol Biol. 1970 Jun 28;50(3):655-70. doi: 10.1016/0022-2836(70)90091-4. PMID: 5529301.

Fazzari M, Frasca A, Bifari F, Landsberger N. Aminoglycoside drugs induce efficient read-through of CDKL5 nonsense mutations, slightly restoring its kinase activity. RNA Biol. 2019 Oct;16(10):1414-1423. doi: 10.1080/15476286.2019.1632633. Epub 2019 Jun 23. PMID: 31232219; PMCID: PMC6779400.

Fehr S, Leonard H, Ho G, Williams S, de Klerk N, Forbes D, Christodoulou J, Downs J. There is variability in the attainment of developmental milestones in the CDKL5 disorder. J Neurodev Disord. 2015;7(1):2. doi: 10.1186/1866-1955-7-2. Epub 2015 Jan 5. PMID: 25657822; PMCID: PMC4318547.

Fekete CA, Mitchell SF, Cherkasova VA, Applefield D, Algire MA, Maag D, Saini AK, Lorsch JR, Hinnebusch AG. N- and C-terminal residues of eIF1A have opposing effects on the fidelity of start codon selection. EMBO J. 2007 Mar 21;26(6):1602-14. doi: 10.1038/sj.emboj.7601613. Epub 2007 Mar 1. PMID: 17332751; PMCID: PMC1829380.

Fichou Y, Nectoux J, Bahi-Buisson N, Chelly J, Bienvenu T. An isoform of the severe encephalopathy-related CDKL5 gene, including a novel exon with extremely high sequence conservation, is specifically expressed in brain. J Hum Genet. 2011 Jan;56(1):52-7. doi: 10.1038/jhg.2010.143. Epub 2010 Dec 2. PMID: 21124335.

Fisette JF, Montagna DR, Mihailescu MR, Wolfe MS. A G-rich element forms a G-quadruplex and regulates BACE1 mRNA alternative splicing. J Neurochem. 2012 Jun;121(5):763-73. doi: 10.1111/j.1471-4159.2012.07680.x. Epub 2012 Mar 13. PMID: 22303960; PMCID: PMC3342435.

Fuchs C, Medici G, Trazzi S, Gennaccaro L, Galvani G, Berteotti C, Ren E, Loi M, Ciani E. CDKL5 deficiency predisposes neurons to cell death through the deregulation of SMAD3 signaling. Brain Pathol. 2019 Sep;29(5):658-674. doi: 10.1111/bpa.12716. Epub 2019 Mar 22. Erratum in: Brain Pathol. 2020 May;30(3):721. PMID: 30793413; PMCID: PMC8028508.

Fuchs C, Trazzi S, Torricella R, Viggiano R, De Franceschi M, Amendola E, Gross C, Calzà L, Bartesaghi R, Ciani E. Loss of CDKL5 impairs survival and dendritic growth of newborn neurons by altering AKT/GSK-3β signaling. Neurobiol Dis. 2014 Oct;70(100):53-68. doi: 10.1016/j.nbd.2014.06.006. Epub 2014 Jun 18. PMID: 24952363; PMCID: PMC4146476.

Gardner PP, Giegerich R. A comprehensive comparison of comparative RNA structure prediction approaches. BMC Bioinformatics. 2004 Sep 30;5:140. doi: 10.1186/1471-

2105-5-140. PMID: 15458580; PMCID: PMC526219.

GARREN LD, HOWELL RR, TOMKINS GM, CROCCO RM. A PARADOXICAL EFFECT OF ACTINOMYCIN D: THE MECHANISM OF REGULATION OF ENZYME SYNTHESIS BY HYDROCORTISONE. Proc Natl Acad Sci U S A. 1964 Oct;52(4):1121-9. doi: 10.1073/pnas.52.4.1121. PMID: 14224391; PMCID: PMC300404.

Gatto CL, Broadie K. Genetic controls balancing excitatory and inhibitory synaptogenesis in neurodevelopmental disorder models. Front Synaptic Neurosci. 2010 Jun 7;2:4. doi: 10.3389/fnsyn.2010.00004. PMID: 21423490; PMCID: PMC3059704.

Ghilardi N, Wiestner A, Skoda RC. Thrombopoietin production is inhibited by a translational mechanism. Blood. 1998 Dec 1;92(11):4023-30. PMID: 9834204.

Ghosh A, Mizuno K, Tiwari SS, Proitsi P, Gomez Perez-Nievas B, Glennon E, Martinez-Nunez RT, Giese KP. Alzheimer's disease-related dysregulation of mRNA translation causes key pathological features with ageing. Transl Psychiatry. 2020 Jun 16;10(1):192. doi: 10.1038/s41398-020-00882-7. PMID: 32546772; PMCID: PMC7297996.

Gingras AC, Raught B, Sonenberg N. mTOR signaling to translation. Curr Top Microbiol Immunol. 2004;279:169-97. doi: 10.1007/978-3-642-18930-2_11. PMID: 14560958.

Gkogkas CG, Khoutorsky A, Ran I, Rampakakis E, Nevarko T, Weatherill DB, Vasuta C, Yee S, Truitt M, Dallaire P, Major F, Lasko P, Ruggero D, Nader K, Lacaille JC, Sonenberg N. Autism-related deficits via dysregulated eIF4E-dependent translational control. Nature. 2013 Jan 17;493(7432):371-7. doi: 10.1038/nature11628. Epub 2012 Nov 21. PMID: 23172145; PMCID: PMC4133997.

Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics. 2003 Jan;19(1):163-4. doi: 10.1093/bioinformatics/19.1.163. PMID: 12499312.

Glass JD, Fromm GH, Chattha AS. Bicuculline and neuronal activity in motor cortex. Electroencephalogr Clin Neurophysiol. 1980 Jan;48(1):16-24. doi: 10.1016/0013-4694(80)90039-5. PMID: 6153317.

Godet AC, David F, Hantelys F, Tatin F, Lacazette E, Garmy-Susini B, Prats AC. IRES Trans-Acting Factors, Key Actors of the Stress Response. Int J Mol Sci. 2019 Feb 20;20(4):924. doi: 10.3390/ijms20040924. PMID: 30791615; PMCID: PMC6412753.

Gonzalez-Herrera IG, Prado-Lourenco L, Pileur F, Conte C, Morin A, Cabon F, Prats H, Vagner S, Bayard F, Audigier S, Prats AC. Testosterone regulates FGF-2 expression during testis maturation by an IRES-dependent translational mechanism. FASEB J. 2006 Mar;20(3):476-8. doi: 10.1096/fj.04-3314fje. Epub 2006 Jan 19. Erratum in: FASEB J. 2006 Jul;20(9):1573-4. PMID: 16423876.

Górska A, Plucinska M, Pedersen L, Kielpinski L, Tehler D, Hagedorn P (2022).

XNAString: Efficient Manipulation of Modified Oligonucleotide Sequences. R package version 1.4.0.

Gray NK, Hentze MW. Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs. EMBO J. 1994 Aug 15;13(16):3882-91. PMID: 8070415; PMCID: PMC395301.

Gray RD, Trent JO, Arumugam S, Chaires JB. Folding Landscape of a Parallel G-Quadruplex. J Phys Chem Lett. 2019 Mar 7;10(5):1146-1151. doi: 10.1021/acs.jpclett.9b00227. Epub 2019 Feb 27. PMID: 30802054; PMCID: PMC6455816.

Griffin E, Re A, Hamel N, Fu C, Bush H, McCaffrey T, Asch AS. A link between diabetes and atherosclerosis: Glucose regulates expression of CD36 at the level of translation. Nat Med. 2001 Jul;7(7):840-6. doi: 10.1038/89969. PMID: 11433350.

Gross JD, Moerke NJ, von der Haar T, Lugovskoy AA, Sachs AB, McCarthy JE, Wagner G. Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. Cell. 2003 Dec 12;115(6):739-50. doi: 10.1016/s0092-8674(03)00975-9. PMID: 14675538.

GROSS PR, MALKIN LI, MOYER WA. TEMPLATES FOR THE FIRST PROTEINS OF EMBRYONIC DEVELOPMENT. Proc Natl Acad Sci U S A. 1964 Mar;51(3):407-14. doi: 10.1073/pnas.51.3.407. PMID: 14171452; PMCID: PMC300086.

Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA websuite. Nucleic Acids Res. 2008 Jul 1;36(Web Server issue):W70-4. doi: 10.1093/nar/gkn188. Epub 2008 Apr 19. PMID: 18424795; PMCID: PMC2447809.

Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA websuite. Nucleic Acids Res. 2008 Jul 1;36(Web Server issue):W70-4. doi: 10.1093/nar/gkn188. Epub 2008 Apr 19. PMID: 18424795; PMCID: PMC2447809.

Grüner S, Peter D, Weber R, Wohlbold L, Chung MY, Weichenrieder O, Valkov E, Igreja C, Izaurralde E. The Structures of eIF4E-eIF4G Complexes Reveal an Extended Interface to Regulate Translation Initiation. Mol Cell. 2016 Nov 3;64(3):467-479. doi: 10.1016/j.molcel.2016.09.020. Epub 2016 Oct 20. PMID: 27773676.

Guerrini R, Parrini E. Epilepsy in Rett syndrome, and CDKL5- and FOXG1-gene-related encephalopathies. Epilepsia. 2012 Dec;53(12):2067-78. doi: 10.1111/j.1528-1167.2012.03656.x. Epub 2012 Sep 21. PMID: 22998673.

Guo S, Lu H. Conjunction of G-quadruplex and stem-loop in the 5' untranslated region of mouse hepatocyte nuclear factor 4-alpha1 mediates strong inhibition of protein expression. Mol Cell Biochem. 2018 Sep;446(1-2):73-81. doi: 10.1007/s11010-018-3274-3. Epub 2018 Jan 13. PMID: 29332143; PMCID: PMC8827217.

Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol. 1999 Mar;19(3):1720-30. doi: 10.1128/MCB.19.3.1720. PMID: 10022859; PMCID: PMC83965.

Haberle V, Forrest AR, Hayashizaki Y, Carninci P, Lenhard B. CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. Nucleic Acids Res. 2015 Apr 30;43(8):e51. doi: 10.1093/nar/gkv054. Epub 2015 Feb 4. PMID: 25653163; PMCID: PMC4417143.

Hagebeuk EE, Marcelis CL, Alders M, Kaspers A, de Weerd AW. Two Siblings With a CDKL5 Mutation: Genotype and Phenotype Evaluation. J Child Neurol. 2015 Oct;30(11):1515-9. doi: 10.1177/0883073815573317. Epub 2015 Mar 11. PMID: 25762588.

Hagebeuk EE, van den Bossche RA, de Weerd AW. Respiratory and sleep disorders in female children with atypical Rett syndrome caused by mutations in the CDKL5 gene. Dev Med Child Neurol. 2013 May;55(5):480-4. doi: 10.1111/j.1469-8749.2012.04432.x. Epub 2012 Nov 14. PMID: 23151060.

Hagerman RJ, Berry-Kravis E, Hazlett HC, Bailey DB Jr, Moine H, Kooy RF, Tassone F, Gantois I, Sonenberg N, Mandel JL, Hagerman PJ. Fragile X syndrome. Nat Rev Dis Primers. 2017 Sep 29;3:17065. doi: 10.1038/nrdp.2017.65. PMID: 28960184.

Haimov O, Sinvani H, Martin F, Ulitsky I, Emmanuel R, Tamarkin-Ben-Harush A, Vardy A, Dikstein R. Efficient and Accurate Translation Initiation Directed by TISU Involves RPS3 and RPS10e Binding and Differential Eukaryotic Initiation Factor 1A Regulation. Mol Cell Biol. 2017 Jul 14;37(15):e00150-17. doi: 10.1128/MCB.00150-17. PMID: 28584194; PMCID: PMC5514451.

Halmai JANM, Deng P, Gonzalez CE, Coggins NB, Cameron D, Carter JL, Buchanan FKB, Waldo JJ, Lock SR, Anderson JD, O'Geen H, Segal DJ, Nolta J, Fink KD. Artificial escape from XCI by DNA methylation editing of the CDKL5 gene. Nucleic Acids Res. 2020 Mar 18;48(5):2372-2387. doi: 10.1093/nar/gkz1214. PMID: 31925439; PMCID: PMC7049732.

Harbers M, Carninci P. Tag-based approaches for transcriptome research and genome annotation. Nat Methods. 2005 Jul;2(7):495-502. doi: 10.1038/nmeth768. PMID: 15973418.

HARDESTY B, MILLER R, SCHWEET R. POLYRIBOSOME BREAKDOWN AND HEMOGLOBIN SYNTHESIS. Proc Natl Acad Sci U S A. 1963 Nov;50(5):924-31. doi: 10.1073/pnas.50.5.924. PMID: 14082359; PMCID: PMC221949.

Harms U, Andreou AZ, Gubaev A, Klostermeier D. eIF4B, eIF4G and RNA regulate eIF4A activity in translation initiation by modulating the eIF4A conformational cycle. Nucleic Acids Res. 2014 Jul;42(12):7911-22. doi: 10.1093/nar/gku440. Epub 2014 May 21. PMID: 24848014; PMCID: PMC4081068.

Hector RD, Dando O, Landsberger N, Kilstrup-Nielsen C, Kind PC, Bailey ME, Cobb SR. Characterisation of CDKL5 Transcript Isoforms in Human and Mouse. PLoS One. 2016 Jun 17;11(6):e0157758. doi: 10.1371/journal.pone.0157758. PMID: 27315173; PMCID: PMC4912119.

Hector RD, Dando O, Ritakari TE, Kind PC, Bailey ME, Cobb SR. Characterisation of Cdkl5 transcript isoforms in rat. Gene. 2017 Mar 1;603:21-26. doi:

10.1016/j.gene.2016.12.001. Epub 2016 Dec 7. PMID: 27940108.

Hentze MW, Caughman SW, Rouault TA, Barriocanal JG, Dancis A, Harford JB, Klausner RD. Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. Science. 1987 Dec 11;238(4833):1570-3. doi: 10.1126/science.3685996. PMID: 3685996.

Herbert S, Waxman, Marco Rabinovitz. Control of reticulocyte polyribosome content and hemoglobin synthesis by heme. Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis,1966. https://doi.org/10.1016/0005-2787(66)90379-0.

Hernández G, Altmann M, Lasko P. Origins and evolution of the mechanisms regulating translation initiation in eukaryotes. Trends Biochem Sci. 2010 Feb;35(2):63-73. doi: 10.1016/j.tibs.2009.10.009. Epub 2009 Nov 18. PMID: 19926289.

Hernández-Ortega K, Garcia-Esparcia P, Gil L, Lucas JJ, Ferrer I. Altered Machinery of Protein Synthesis in Alzheimer's: From the Nucleolus to the Ribosome. Brain Pathol. 2016 Sep;26(5):593-605. doi: 10.1111/bpa.12335. Epub 2015 Dec 14. PMID: 26512942; PMCID: PMC8029302.

Hershey JWB, Sonenberg N, Mathews MB. Principles of Translational Control. Cold Spring Harb Perspect Biol. 2019 Sep 3;11(9):a032607. doi: 10.1101/cshperspect.a032607. PMID: 29959195; PMCID: PMC6719596.

Heuser JE, Reese TS. Evidence for recycling of synaptic vesicle membrane during transmitter release at the frog neuromuscular junction. J Cell Biol. 1973 May;57(2):315-44. doi: 10.1083/jcb.57.2.315. PMID: 4348786; PMCID: PMC2108984.

Heywood SM. Specificity of mRNA binding factor in eukaryotes. Proc Natl Acad Sci U S A. 1970 Dec;67(4):1782-8. doi: 10.1073/pnas.67.4.1782. PMID: 5275376; PMCID: PMC283427.

Hinnebusch AG, Ivanov IP, Sonenberg N. Translational control by 5'-untranslated regions of eukaryotic mRNAs. Science. 2016 Jun 17;352(6292):1413-6. doi: 10.1126/science.aad9868. PMID: 27313038; PMCID: PMC7422601.

Hinnebusch AG. The scanning mechanism of eukaryotic translation initiation. Annu Rev Biochem. 2014;83:779-812. doi: 10.1146/annurev-biochem-060713-035802.Epub 2014 Jan 29. PMID: 24499181.

Hinnebusch AG. Translational regulation of GCN4 and the general amino acid control of yeast. Annu Rev Microbiol. 2005;59:407-50. doi: 10.1146/annurev.micro.59.031805.133833. PMID: 16153175.

Holt CE, Martin KC, Schuman EM. Local translation in neurons: visualization and function. Nat Struct Mol Biol. 2019 Jul;26(7):557-566. doi: 10.1038/s41594-019-0263-5. Epub 2019 Jul 3. PMID: 31270476.

Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, Bhai J, Billis K, Boddu S, Charkhchi M, Cummins C, Da Rin

Fioretto L, Davidson C, Dodiya K, El Houdaigui B, Fatima R, Gall A, Garcia Giron C, Grego T, Guijarro-Clarke C, Haggerty L, Hemrom A, Hourlier T, Izuogu OG, Juettemann T, Kaikala V, Kay M, Lavidas I, Le T, Lemos D, Gonzalez Martinez J, Marugán JC, Maurel T, McMahon AC, Mohanan S, Moore B, Muffato M, Oheh DN, Paraschas D, Parker A, Parton A, Prosovetskaia I, Sakthivel MP, Salam AIA, Schmitt BM, Schuilenburg H, Sheppard D, Steed E, Szpak M, Szuba M, Taylor K, Thormann A, Threadgold G, Walts B, Winterbottom A, Chakiachvili M, Chaubal A, De Silva N, Flint B, Frankish A, Hunt SE, IIsley GR, Langridge N, Loveland JE, Martin FJ, Mudge JM, Morales J, Perry E, Ruffier M, Tate J, Thybert D, Trevanion SJ, Cunningham F, Yates AD, Zerbino DR, Flicek P. Ensembl 2021. Nucleic Acids Res. 2021 Jan 8;49(D1):D884-D891. doi: 10.1093/nar/gkaa942. PMID: 33137190; PMCID: PMC7778975.

Huez I, Bornes S, Bresson D, Créancier L, Prats H. New vascular endothelial growth factor isoform generated by internal ribosome entry site-driven CUG translation initiation. Mol Endocrinol. 2001 Dec;15(12):2197-210. doi: 10.1210/mend.15.12.0738. PMID: 11731620.

Huez I, Créancier L, Audigier S, Gensac MC, Prats AC, Prats H. Two independent internal ribosome entry sites are involved in translation initiation of vascular endothelial growth factor mRNA. Mol Cell Biol. 1998 Nov;18(11):6178-90. doi: 10.1128/MCB.18.11.6178. PMID: 9774635; PMCID: PMC109205.

HULTIN T. Activation of ribosomes in sea urchin eggs in response to fertilization. Exp Cell Res. 1961 Nov;25:405-17. doi: 10.1016/0014-4827(61)90290-7. PMID: 14449935.

Iacono M, Mignone F, Pesole G. uAUG and uORFs in human and rodent 5'untranslated mRNAs. Gene. 2005 Apr 11;349:97-105. doi: 10.1016/j.gene.2004.11.041. PMID: 15777708.

Imai Y, Gehrke S, Wang HQ, Takahashi R, Hasegawa K, Oota E, Lu B. Phosphorylation of 4E-BP by LRRK2 affects the maintenance of dopaminergic neurons in Drosophila. EMBO J. 2008 Sep 17;27(18):2432-43. doi: 10.1038/emboj.2008.163. Epub 2008 Aug 14. PMID: 18701920; PMCID: PMC2543051.

Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. The ribosomeprofiling strategy for monitoring translation in vivo by deep sequencing of ribosome- protected mRNA fragments. Nat Protoc. 2012 Jul 26;7(8):1534-50. doi: 10.1038/nprot.2012.086. PMID: 22836135; PMCID: PMC3535016.

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009 Apr 10;324(5924):218-23. doi: 10.1126/science.1168978. Epub 2009 Feb 12. PMID: 19213877; PMCID: PMC2746483.

Ingolia NT, Hussmann JA, Weissman JS. Ribosome Profiling: Global Views of Translation. Cold Spring Harb Perspect Biol. 2019 May 1;11(5):a032698. doi: 10.1101/cshperspect.a032698. PMID: 30037969; PMCID: PMC6496350.

Jackson RJ, Hellen CU, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol. 2010 Feb;11(2):113-27. doi:

10.1038/nrm2838. PMID: 20094052; PMCID: PMC4461372.

Jackson RJ. The current status of vertebrate cellular mRNA IRESs. Cold Spring Harb Perspect Biol. 2013 Feb 1;5(2):a011569. doi: 10.1101/cshperspect.a011569. PMID: 23378589; PMCID: PMC3552511.

Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol. 1961 Jun;3:318-56. doi: 10.1016/s0022-2836(61)80072-7. PMID: 13718526.

Jacquemont S, Pacini L, Jønch AE, Cencelli G, Rozenberg I, He Y, D'Andrea L, Pedini G, Eldeeb M, Willemsen R, Gasparini F, Tassone F, Hagerman R, Gomez-Mancilla B, Bagni C. Protein synthesis levels are increased in a subset of individuals with fragile X syndrome. Hum Mol Genet. 2018 Jun 15;27(12):2039-2051. doi: 10.1093/hmg/ddy099. Erratum in: Hum Mol Genet. 2018 Nov 1;27(21):3825. PMID: 29590342; PMCID: PMC5985734.

Jakimiec M, Paprocka J, Śmigiel R. CDKL5 Deficiency Disorder - A ComplexEpileptic Encephalopathy. Brain Sci. 2020 Feb 17;10(2):107. doi: 10.3390/brainsci10020107. PMID: 32079229; PMCID: PMC7071516.

Jakimiec M, Paprocka J, Śmigiel R. CDKL5 Deficiency Disorder-A Complex Epileptic Encephalopathy. Brain Sci. 2020 Feb 17;10(2):107. doi: 10.3390/brainsci10020107. PMID: 32079229; PMCID: PMC7071516.

Jambhekar A, Derisi JL. Cis-acting determinants of asymmetric, cytoplasmic RNA transport. RNA. 2007 May;13(5):625-42. doi: 10.1261/rna.262607. PMID: 17449729; PMCID: PMC1852811.

Jang SK, Kräusslich HG, Nicklin MJ, Duke GM, Palmenberg AC, Wimmer E. A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. J Virol. 1988 Aug;62(8):2636-43. doi: 10.1128/JVI.62.8.2636-2643.1988. PMID: 2839690; PMCID: PMC253694.

Jdila MB, Triki C, Rhouma BB, Jomaa RB, Issa AB, Ammar-Keskes L, Kamoun F, Fakhfakh F. A novel C-terminal truncated mutation in hCDKL5 protein causing a severe West syndrome: Comparison with previous truncated mutations and genotype/phenotype correlation. Int J Dev Neurosci. 2019 Feb;72:22-30. doi: 10.1016/j.ijdevneu.2018.09.006. Epub 2018 Sep 17. PMID: 30236769.

Jishi A, Qi X, Miranda HC. Implications of mRNA translation dysregulation for neurological disorders. Semin Cell Dev Biol. 2021 Jun;114:11-19. doi: 10.1016/j.semcdb.2020.09.005. Epub 2020 Oct 21. PMID: 34024497; PMCID: PMC8144541.

Johnston GA. Advantages of an antagonist: bicuculline and other GABA antagonists. Br J Pharmacol. 2013 May;169(2):328-36. doi: 10.1111/bph.12127. PMID: 23425285; PMCID: PMC3651659.

Kalscheuer VM, Tao J, Donnelly A, et al. Disruption of the serine/threonine kinase 9 gene causes severe X-linked infantile spasms and mental retardation. Am J Hum Genet. 2003 Jun;72(6):1401-11. doi: 10.1086/375538. Epub 2003 May 7. PMID: 12736870;

PMCID: PMC1180301.

Kameshita I, Sekiguchi M, Hamasaki D, Sugiyama Y, Hatano N, Suetake I, Tajima S, Sueyoshi N. Cyclin-dependent kinase-like 5 binds and phosphorylates DNA methyltransferase 1. Biochem Biophys Res Commun. 2008 Dec 26;377(4):1162-7. doi: 10.1016/j.bbrc.2008.10.113. Epub 2008 Oct 31. PMID: 18977197.

Karginov TA, Pastor DPH, Semler BL, Gomez CM. Mammalian Polycistronic mRNAs and Disease. Trends Genet. 2017 Feb;33(2):129-142. doi: 10.1016/j.tig.2016.11.007. Epub 2016 Dec 21. PMID: 28012572; PMCID: PMC5285393.

Katayama S, Sueyoshi N, Inazu T, Kameshita I. Cyclin-Dependent Kinase-Like 5 (CDKL5): Possible Cellular Signalling Targets and Involvement in CDKL5 Deficiency Disorder. Neural Plast. 2020;2020:6970190. Published 2020 Jun 5. doi:10.1155/2020/6970190

Kelleher RJ 3rd, Bear MF. The autistic neuron: troubled translation? Cell. 2008 Oct 31;135(3):401-6. doi: 10.1016/j.cell.2008.10.017. PMID: 18984149.

Kelleher RJ 3rd, Govindarajan A, Jung HY, Kang H, Tonegawa S. Translational control by MAPK signaling in long-term synaptic plasticity and memory. Cell. 2004 Feb 6;116(3):467-79. doi: 10.1016/s0092-8674(04)00115-1. PMID: 15016380.

Kerpedjiev P, Hammer S, Hofacker IL. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. Bioinformatics. 2015 Oct 15;31(20):3377-9. doi: 10.1093/bioinformatics/btv372. Epub 2015 Jun 22. PMID: 26099263; PMCID: PMC4595900.

Khanam T, Muñoz I, Weiland F, Carroll T, Morgan M, Borsos BN, Pantazi V, Slean M, Novak M, Toth R, Appleton P, Pankotai T, Zhou H, Rouse J. CDKL5 kinase controls transcription-coupled responses to DNA damage. EMBO J. 2021 Dec 1;40(23):e108271. doi: 10.15252/embj.2021108271. Epub 2021 Oct 4. PMID: 34605059; PMCID: PMC8634139.

Khanam T, Muñoz I, Weiland F, Carroll T, Morgan M, Borsos BN, Pantazi V, Slean M, Novak M, Toth R, Appleton P, Pankotai T, Zhou H, Rouse J. CDKL5 kinase controls transcription-coupled responses to DNA damage. EMBO J. 2021 Dec 1;40(23):e108271. doi: 10.15252/embj.2021108271. Epub 2021 Oct 4. PMID: 34605059; PMCID: PMC8634139.

Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. 2006 Jul 1;34(Web Server issue):W676-82. doi: 10.1093/nar/gkl253. PMID: 16845096; PMCID: PMC1538864.

Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. 2006 Jul 1;34(Web Server issue):W676-82. doi: 10.1093/nar/gkl253. PMID: 16845096; PMCID: PMC1538864.

Kim JY, Bai Y, Jayne LA, Hector RD, Persaud AK, Ong SS, Rojesh S, Raj R, Feng MJHH, Chung S, Cianciolo RE, Christman JW, Campbell MJ, Gardner DS, Baker SD, Sparreboom A, Govindarajan R, Singh H, Chen T, Poi M, Susztak K, Cobb SR, Pabla

NS. A kinome-wide screen identifies a CDKL5-SOX9 regulatory axis in epithelial cell death and kidney injury. Nat Commun. 2020 Apr 21;11(1):1924. doi: 10.1038/s41467-020-15638-6. PMID: 32317630; PMCID: PMC7174303.

Kimura H, Shiota K. Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. J Biol Chem. 2003 Feb 14;278(7):4806-12. doi: 10.1074/jbc.M209923200. Epub 2002 Dec 6. PMID: 12473678.

King HA, Cobbold LC, Willis AE. The role of IRES trans-acting factors in regulating translation initiation. Biochem Soc Trans. 2010 Dec;38(6):1581-6. doi: 10.1042/BST0381581. PMID: 21118130.

Koch L. Exploring human genomic diversity with gnomAD. Nat Rev Genet. 2020 Aug;21(8):448. doi: 10.1038/s41576-020-0255-7. PMID: 32488197.

Koh DC, Edelman GM, Mauro VP. Physical evidence supporting a ribosomal shunting mechanism of translation initiation for BACE1 mRNA. Translation (Austin). 2013 Apr 1;1(1):e24400. doi: 10.4161/trla.24400. PMID: 26824018; PMCID: PMC4718059.

Kolupaeva VG, Unbehaun A, Lomakin IB, Hellen CU, Pestova TV. Binding of eukaryotic initiation factor 3 to ribosomal 40S subunits and its role in ribosomal dissociation and anti-association. RNA. 2005 Apr;11(4):470-86. doi: 10.1261/rna.7215305. Epub 2005 Feb 9. PMID: 15703437; PMCID: PMC1370736.

Komar AA, Hatzoglou M. Cellular IRES-mediated translation: the war of ITAFs in pathophysiological states. Cell Cycle. 2011 Jan 15;10(2):229-40. doi: 10.4161/cc.10.2.14472. Epub 2011 Jan 15. PMID: 21220943; PMCID: PMC3048795.

Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, de Lima RL, Daack-Hirsch S, Sander A, McDonald-McGinn DM, Zackai EH, Lammer EJ, Aylsworth AS, Ardinger HH, Lidral AC, Pober BR, Moreno L, Arcos-Burgos M, Valencia C, Houdayer C, Bahuau M, Moretti-Ferreira D, Richieri-Costa A, Dixon MJ, Murray JC. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat Genet. 2002 Oct;32(2):285-9. doi: 10.1038/ng985. Epub 2002 Sep 3. PMID: 12219090; PMCID: PMC3169431.

Kondo T, Okabe M, Sanada M, Kurosawa M, Suzuki S, Kobayashi M, Hosokawa M, Asaka M. Familial essential thrombocythemia associated with one-base deletion in the 5'-untranslated region of the thrombopoietin gene. Blood. 1998 Aug 15;92(4):1091-6. PMID: 9694695.

Kothur K, Holman K, Farnsworth E, Ho G, Lorentzos M, Troedson C, Gupta S, Webster R, Procopis PG, Menezes MP, Antony J, Ardern-Holmes S, Dale RC, Christodoulou J, Gill D, Bennetts B. Diagnostic yield of targeted massively parallel sequencing in children with epileptic encephalopathy. Seizure. 2018 Jul;59:132-140. doi: 10.1016/j.seizure.2018.05.005. Epub 2018 May 28. PMID: 29852413.

Kozak M. A second look at cellular mRNA sequences said to function as internal ribosome entry sites. Nucleic Acids Res. 2005 Nov 28;33(20):6593-602. doi: 10.1093/nar/gki958. PMID: 16314320; PMCID: PMC1298923.

Kozak M. Constraints on reinitiation of translation in mammals. Nucleic Acids Res. 2001 Dec 15;29(24):5226-32. doi: 10.1093/nar/29.24.5226. PMID: 11812856; PMCID: PMC97554.

Kozak M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc Natl Acad Sci U S A. 1986 May;83(9):2850-4. doi: 10.1073/pnas.83.9.2850. PMID: 3458245; PMCID: PMC323404.

Kozak M. Lessons (not) learned from mistakes about translation. Gene. 2007 Nov 15;403(1-2):194-203. doi: 10.1016/j.gene.2007.08.017. Epub 2007 Sep 5. PMID: 17888589.

Kozera B, Rapacz M. Reference genes in real-time PCR. J Appl Genet. 2013 Nov;54(4):391-406. doi: 10.1007/s13353-013-0173-x. PMID: 24078518; PMCID: PMC3825189.

Krishnaraj R, Ho G, Christodoulou J. RettBASE: Rett syndrome database update. Hum Mutat. 2017 Aug;38(8):922-931. doi: 10.1002/humu.23263. Epub 2017 Jun 9. PMID: 28544139.

Kumari S, Bugaut A, Huppert JL, Balasubramanian S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. Nat Chem Biol. 2007

Apr;3(4):218-21. doi: 10.1038/nchembio864. Epub 2007 Feb 25. PMID: 17322877; PMCID: PMC2206252.

Kwan T, Thompson SR. Noncanonical Translation Initiation in Eukaryotes. Cold Spring Harb Perspect Biol. 2019 Apr 1;11(4):a032672. doi: 10.1101/cshperspect.a032672. PMID: 29959190; PMCID: PMC6442200.

Lambrechts D, Storkebaum E, Morimoto M, Del-Favero J, Desmet F, Marklund SL, Wyns S, Thijs V, Andersson J, van Marion I, Al-Chalabi A, Bornes S, Musson R, Hansen V, Beckman L, Adolfsson R, Pall HS, Prats H, Vermeire S, Rutgeerts P, Katayama S, Awata T, Leigh N, Lang-Lazdunski L, Dewerchin M, Shaw C, Moons L, Vlietinck R, Morrison KE, Robberecht W, Van Broeckhoven C, Collen D, Andersen PM, Carmeliet P. VEGF is a modifier of amyotrophic lateral sclerosis in mice and humans and protects motoneurons against ischemic death. Nat Genet. 2003 Aug;34(4):383-94. doi: 10.1038/ng1211. PMID: 12847526.

Lammich S, Schöbel S, Zimmer AK, Lichtenthaler SF, Haass C. Expression of the Alzheimer protease BACE1 is suppressed via its 5'-untranslated region. EMBO Rep. 2004 Jun;5(6):620-5. doi: 10.1038/sj.embor.7400166. Epub 2004 May 28. PMID: 15167888; PMCID: PMC1299076.

Lammich S, Kamp F, Wagner J, Nuscher B, Zilow S, Ludwig AK, Willem M, Haass C. Translational repression of the disintegrin and metalloprotease ADAM10 by a stable G-quadruplex secondary structure in its 5'-untranslated region. J Biol Chem. 2011 Dec 30;286(52):45063-72. doi: 10.1074/jbc.M111.296921. Epub 2011 Nov 7. PMID: 22065584; PMCID: PMC3248004.

Lavoie G, Estève PO, Laulan NB, Pradhan S, St-Pierre Y. PKC isoforms interact with

and phosphorylate DNMT1. BMC Biol. 2011 May 27;9:31. doi: 10.1186/1741-7007-9-31. PMID: 21619587; PMCID: PMC3118390.

Lee KZ, Liao W. Loss of CDKL5 disrupts respiratory function in mice. Respir Physiol Neurobiol. 2018 Jan;248:48-54. doi: 10.1016/j.resp.2017.11.010. Epub 2017 Dec 5. PMID: 29208534.

Lee S, Liu B, Lee S, Huang SX, Shen B, Qian SB. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc Natl Acad Sci U S A. 2012 Sep 11;109(37):E2424-32. doi: 10.1073/pnas.1207846109. Epub 2012 Aug 27. PMID: 22927429; PMCID: PMC3443142.

Lefebvre S, Bürglen L, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, Cruaud C, Millasseau P, Zeviani M, et al. Identification and characterization of a spinal muscular atrophy-determining gene. Cell. 1995 Jan 13;80(1):155-65. doi: 10.1016/0092-8674(95)90460-3. PMID: 7813012.

Lefebvre S, Bürglen L, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, Cruaud C, Millasseau P, Zeviani M, et al. Identification and characterization of a spinal muscular atrophy-determining gene. Cell. 1995 Jan 13;80(1):155-65. doi: 10.1016/0092-8674(95)90460-3. PMID: 7813012.

Leppek K, Das R, Barna M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. Nat Rev Mol Cell Biol. 2018 Mar;19(3):158-174. doi: 10.1038/nrm.2017.103. Epub 2017 Nov 22. Erratum in: Nat Rev Mol Cell Biol. 2018 Oct;19(10):673. PMID: 29165424; PMCID: PMC5820134.

Lewerenz J, Sato H, Albrecht P, Henke N, Noack R, Methner A, Maher P. Mutation of ATF4 mediates resistance of neuronal cell lines against oxidative stress by inducing xCT expression. Cell Death Differ. 2012 May;19(5):847-58. doi: 10.1038/cdd.2011.165. Epub 2011 Nov 18. PMID: 22095285; PMCID: PMC3321624.

Lewis SM, Veyrier A, Hosszu Ungureanu N, Bonnal S, Vagner S, Holcik M. Subcellular relocalization of a trans-acting factor regulates XIAP IRES-dependent translation. Mol Biol Cell. 2007 Apr;18(4):1302-11. doi: 10.1091/mbc.e06-06-0515. Epub 2007 Feb 7. PMID: 17287399; PMCID: PMC1838995.

Li J, Chen J, Ricupero CL, Hart RP, Schwartz MS, Kusnecov A, Herrup K. Nuclear accumulation of HDAC4 in ATM deficiency promotes neurodegeneration in ataxia telangiectasia. Nat Med. 2012 May;18(5):783-90. doi: 10.1038/nm.2709. PMID: 22466704; PMCID: PMC3378917.

Liang JS, Huang H, Wang JS, Lu JF. Phenotypic manifestations between male and female children with CDKL5 mutations. Brain Dev. 2019 Oct;41(9):783-789. doi: 10.1016/j.braindev.2019.05.003. Epub 2019 May 20. PMID: 31122804.

Liang XH, Shen W, Sun H, Migawa MT, Vickers TA, Crooke ST. Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. Nat Biotechnol. 2016 Aug;34(8):875-80. doi: 10.1038/nbt.3589. Epub 2016 Jul 11. PMID: 27398791.

Liang XH, Sun H, Shen W, Wang S, Yao J, Migawa MT, Bui HH, Damle SS, Riney S, Graham MJ, Crooke RM, Crooke ST. Antisense oligonucleotides targeting translation inhibitory elements in 5' UTRs can selectively increase protein levels. Nucleic Acids Res. 2017 Sep 19;45(16):9528-9546. doi: 10.1093/nar/gkx632. PMID: 28934489; PMCID: PMC5766168.

Lin C, Franco B, Rosner MR. CDKL5/Stk9 kinase inactivation is associated with neuronal developmental disorders. Hum Mol Genet. 2005 Dec 15;14(24):3775-86. doi: 10.1093/hmg/ddi391. Epub 2005 Dec 5. PMID: 16330482.

Lin C, Franco B, Rosner MR. CDKL5/Stk9 kinase inactivation is associated with neuronal developmental disorders. Hum Mol Genet. 2005 Dec 15;14(24):3775-86. doi: 10.1093/hmg/ddi391. Epub 2005 Dec 5. PMID: 16330482.

Lindqvist L, Pelletier J. Inhibitors of translation initiation as cancer therapeutics. Future Med Chem. 2009 Dec;1(9):1709-22. doi: 10.4155/fmc.09.122. PMID: 21425987.

Lindy AS, Stosser MB, Butler E, Downtain-Pickersgill C, Shanmugham A, Retterer K, Brandt T, Richard G, McKnight DA. Diagnostic outcomes for genetic testing of 70 genes in 8565 patients with epilepsy and neurodevelopmental disorders. Epilepsia. 2018 May;59(5):1062-1071. doi: 10.1111/epi.14074. Epub 2018 Apr 14. PMID: 29655203.

Liu XS, Wu H, Krzisch M, Wu X, Graef J, Muffat J, Hnisz D, Li CH, Yuan B, Xu C, Li Y, Vershkov D, Cacace A, Young RA, Jaenisch R. Rescue of Fragile X Syndrome Neurons by DNA Methylation Editing of the FMR1 Gene. Cell. 2018 Feb 22;172(5):979-992.e6. doi: 10.1016/j.cell.2018.01.012. Epub 2018 Feb 15. PMID: 29456084; PMCID: PMC6375087.

Loi M, Trazzi S, Fuchs C, Galvani G, Medici G, Gennaccaro L, Tassinari M, Ciani E. Increased DNA Damage and Apoptosis in CDKL5-Deficient Neurons. Mol Neurobiol. 2020 May;57(5):2244-2262. doi: 10.1007/s12035-020-01884-8. Epub 2020 Jan 30. Erratum in: Mol Neurobiol. 2020 Feb 17;: PMID: 32002787.

Luukkonen BG, Tan W, Schwartz S. Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance. J Virol. 1995 Jul;69(7):4086-94. doi: 10.1128/JVI.69.7.4086-4094.1995. PMID: 7769666; PMCID: PMC189143.

Lynch M, Scofield DG, Hong X. The evolution of transcription-initiation sites. Mol Biol Evol. 2005 Apr;22(4):1137-46. doi: 10.1093/molbev/msi100. Epub 2005 Feb 2. PMID: 15689525.

Ma T, Trinh MA, Wexler AJ, Bourbon C, Gatti E, Pierre P, Cavener DR, Klann E. Suppression of eIF2α kinases alleviates Alzheimer's disease-related plasticity and memory deficits. Nat Neurosci. 2013 Sep;16(9):1299-305. doi: 10.1038/nn.3486. Epub 2013 Aug 11. PMID: 23933749; PMCID: PMC3756900.

Majumdar R, Bandyopadhyay A, Maitra U. Mammalian translation initiation factor eIF1 functions with eIF1A and eIF3 in the formation of a stable 40 S preinitiation complex. J Biol Chem. 2003 Feb 21;278(8):6580-7. doi: 10.1074/jbc.M210357200.

Epub 2002 Dec 18. PMID: 12493757.

Mall MA, Hartl D. CFTR: cystic fibrosis and beyond. Eur Respir J. 2014 Oct;44(4):1042-54. doi: 10.1183/09031936.00228013. Epub 2014 Jun 12. PMID: 24925916.

Mangatt M, Wong K, Anderson B, Epstein A, Hodgetts S, Leonard H, Downs J. Prevalence and onset of comorbidities in the CDKL5 disorder differ from Rett syndrome. Orphanet J Rare Dis. 2016 Apr 14;11:39. doi: 10.1186/s13023-016-0418-y. PMID: 27080038; PMCID: PMC4832563.

Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science. 2002 Dec 6;298(5600):1912-34. doi: 10.1126/science.1075762. PMID: 12471243.

Manzella JM, Blackshear PJ. Regulation of rat ornithine decarboxylase mRNA translation by its 5'-untranslated region. J Biol Chem. 1990 Jul 15;265(20):11817-22. PMID: 2365701.

Mari F, Azimonti S, Bertani I, Bolognese F, Colombo E, Caselli R, Scala E, Longo I, Grosso S, Pescucci C, Ariani F, Hayek G, Balestri P, Bergo A, Badaracco G, Zappella M, Broccoli V, Renieri A, Kilstrup-Nielsen C, Landsberger N. CDKL5 belongs to the same molecular pathway of MeCP2 and it is responsible for the early-onset seizure variant of Rett syndrome. Hum Mol Genet. 2005 Jul 15;14(14):1935-46. doi: 10.1093/hmg/ddi198. Epub 2005 May 25. PMID: 15917271.

Marintchev A, Ito T. eIF2B and the Integrated Stress Response: A Structural and Mechanistic View. Biochemistry. 2020 Apr 7;59(13):1299-1308. doi: 10.1021/acs.biochem.0c00132. Epub 2020 Mar 26. PMID: 32200625; PMCID: PMC7189779.

Martineau Y, Le Bec C, Monbrun L, Allo V, Chiu IM, Danos O, Moine H, Prats H, Prats AC. Internal ribosome entry site structural motifs conserved among mammalian fibroblast growth factor 1 alternatively spliced mRNAs. Mol Cell Biol. 2004 Sep;24(17):7622-35. doi: 10.1128/MCB.24.17.7622-7635.2004. PMID: 15314170; PMCID: PMC507008.

McCaskill JS. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers. 1990 May-Jun;29(6-7):1105-19. doi: 10.1002/bip.360290621. PMID: 1695107.

McCormick W, Penman S. Regulation of protein synthesis in HeLa cells: translation at elevated temperatures. J Mol Biol. 1969 Jan;39(2):315-33. doi: 10.1016/0022-2836(69)90320-9. PMID: 5362672.

McGillivray P, Ault R, Pawashe M, Kitchen R, Balasubramanian S, Gerstein M. A comprehensive catalog of predicted functional upstream open reading frames in humans. Nucleic Acids Res. 2018 Apr 20;46(7):3326-3338. doi: 10.1093/nar/gky188. PMID: 29562350; PMCID: PMC6283423.

McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence

analysis tools. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W20-5. doi: 10.1093/nar/gkh435. PMID: 15215342; PMCID: PMC441573.

Meier S, Bell M, Lyons DN, Rodriguez-Rivera J, Ingram A, Fontaine SN, Mechas E, Chen J, Wolozin B, LeVine H 3rd, Zhu H, Abisambra JF. Pathological Tau Promotes Neuronal Damage by Impairing Ribosomal Function and Decreasing Protein Synthesis. J Neurosci. 2016 Jan 20;36(3):1001-7. doi: 10.1523/JNEUROSCI.3029-15.2016. PMID: 26791227; PMCID: PMC4719006.

Merrick WC, Pavitt GD. Protein Synthesis Initiation in Eukaryotic Cells. Cold Spring Harb Perspect Biol. 2018 Dec 3;10(12):a033092. doi: 10.1101/cshperspect.a033092. PMID: 29735639; PMCID: PMC6280705.

Menges F. 2016. Spectragryph 1.2. Optical spectroscopy software http://www.effemm2.de/spectragryph/

Merrick WC. eIF4F: a retrospective. J Biol Chem. 2015 Oct 2;290(40):24091-9. doi: 10.1074/jbc.R115.675280. Epub 2015 Aug 31. PMID: 26324716; PMCID: PMC4591800.

Meyer KD, Patil DP, Zhou J, Zinoviev A, Skabkin MA, Elemento O, Pestova TV, Qian SB, Jaffrey SR. 5' UTR m(6)A Promotes Cap-Independent Translation. Cell. 2015 Nov 5;163(4):999-1010. doi: 10.1016/j.cell.2015.10.012. Epub 2015 Oct 22. PMID: 26593424; PMCID: PMC4695625.

Migeon BR. X-linked diseases: susceptible females. Genet Med. 2020 Jul;22(7):1156-1174. doi: 10.1038/s41436-020-0779-4. Epub 2020 Apr 14. PMID: 32284538; PMCID: PMC7332419.

Mignone F, Gissi C, Liuni S, Pesole G. Untranslated regions of mRNAs. Genome Biol. 2002;3(3):REVIEWS0004. doi: 10.1186/gb-2002-3-3-reviews0004. Epub 2002 Feb 28. PMID: 11897027; PMCID: PMC139023.

Mihailovich M, Thermann R, Grohovaz F, Hentze MW, Zacchetti D. Complex translational regulation of BACE1 involves upstream AUGs and stimulatory elements within the 5' untranslated region. Nucleic Acids Res. 2007;35(9):2975-85. doi: 10.1093/nar/gkm191. Epub 2007 Apr 16. PMID: 17439957; PMCID: PMC1888809.

Minatohara K, Akiyoshi M, Okuno H. Role of Immediate-Early Genes in Synaptic Plasticity and Neuronal Ensembles Underlying the Memory Trace. Front Mol Neurosci. 2016 Jan 5;8:78. doi: 10.3389/fnmol.2015.00078. PMID: 26778955; PMCID: PMC4700275.

Miranda JG, Schleicher WE, Wells KL, Ramirez DG, Landgrave SP, Benninger RKP. Dynamic changes in β-cell [Ca^{2+}] regulate NFAT activation, gene transcription, and islet gap junction communication. Mol Metab. 2022 Mar;57:101430. doi: 10.1016/j.molmet.2021.101430. Epub 2021 Dec 31. PMID: 34979329; PMCID: PMC8804269.

Mitchell SA, Spriggs KA, Coldwell MJ, Jackson RJ, Willis AE. The Apaf-1 internal ribosome entry segment attains the correct structural conformation for function via

interactions with PTB and unr. Mol Cell. 2003 Mar;11(3):757-71. doi: 10.1016/s1097-2765(03)00093-5. PMID: 12667457.

Moine H, Vitale N. Of local translation control and lipid signaling in neurons. Adv Biol Regul. 2019 Jan;71:194-205. doi: 10.1016/j.jbior.2018.09.005. Epub 2018 Sep 15. PMID: 30262213.

Monani UR, Lorson CL, Parsons DW, Prior TW, Androphy EJ, Burghes AH, McPherson JD. A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. Hum Mol Genet. 1999 Jul;8(7):1177-83. doi: 10.1093/hmg/8.7.1177. PMID: 10369862.

Montini E, Andolfi G, et al. Identification and characterization of a novel serine-threonine kinase gene from the Xp22 region. Genomics. 1998 Aug 1;51(3):427-33. doi: 10.1006/geno.1998.5391. PMID: 9721213.

Moon SL, Sonenberg N, Parker R. Neuronal Regulation of eIF2α Function in Health and Neurological Disorders. Trends Mol Med. 2018 Jun;24(6):575-589. doi: 10.1016/j.molmed.2018.04.001. Epub 2018 Apr 30. PMID: 29716790.

Morfoisse F, Kuchnio A, Frainay C, Gomez-Brouchet A, Delisle MB, Marzi S, Helfer AC, Hantelys F, Pujol F, Guillermet-Guibert J, Bousquet C, Dewerchin M, Pyronnet S, Prats AC, Carmeliet P, Garmy-Susini B. Hypoxia induces VEGF-C expression in metastatic tumor cells via a HIF-1α-independent translation-mediated mechanism. Cell Rep. 2014 Jan 16;6(1):155-67. doi: 10.1016/j.celrep.2013.12.011. Epub 2014 Jan 2. PMID: 24388748.

Morfoisse F, Tatin F, Hantelys F, Adoue A, Helfer AC, Cassant-Sourdy S, Pujol F, Gomez-Brouchet A, Ligat L, Lopez F, Pyronnet S, Courty J, Guillermet-Guibert J, Marzi S, Schneider RJ, Prats AC, Garmy-Susini BH. Nucleolin Promotes Heat Shock-Associated Translation of VEGF-D to Promote Tumor Lymphangiogenesis. Cancer Res. 2016 Aug 1;76(15):4394-405. doi: 10.1158/0008-5472.CAN-15-3140. Epub 2016 Jun 8. PMID: 27280395.

Morgan AJ, Jacob R. Ionomycin enhances Ca2+ influx by stimulating store-regulated cation entry and not by a direct action at the plasma membrane. Biochem J. 1994 Jun 15;300 (Pt 3)(Pt 3):665-72. doi: 10.1042/bj3000665. PMID: 8010948; PMCID: PMC1138219.

Morris DR, Geballe AP. Upstream open reading frames as regulators of mRNA translation. Mol Cell Biol. 2000 Dec;20(23):8635-42. doi: 10.1128/MCB.20.23.8635-8642.2000. PMID: 11073965; PMCID: PMC86464.

Mouton-Liger F, Paquet C, Dumurgier J, Bouras C, Pradier L, Gray F, Hugon J. Oxidative stress increases BACE1 protein levels through activation of the PKR-eIF2α pathway. Biochim Biophys Acta. 2012 Jun;1822(6):885-96. doi: 10.1016/j.bbadis.2012.01.009. Epub 2012 Jan 28. PMID: 22306812.

Muckenthaler M, Gray NK, Hentze MW. IRP-1 binding to ferritin mRNA prevents the recruitment of the small ribosomal subunit by the cap-binding complex eIF4F. Mol Cell. 1998 Sep;2(3):383-8. doi: 10.1016/s1097-2765(00)80282-8. PMID: 9774976.

Mueller PP, Hinnebusch AG. Multiple upstream AUG codons mediate translational control of GCN4. Cell. 1986 Apr 25;45(2):201-7. doi: 10.1016/0092-8674(86)90384-3. PMID: 3516411.

Munzarová V, Pánek J, Gunišová S, Dányi I, Szamecz B, Valášek LS. Translation reinitiation relies on the interaction between eIF3a/TIF32 and progressively folded cis-acting mRNA elements preceding short uORFs. PLoS Genet. 2011 Jul;7(7):e1002137. doi: 10.1371/journal.pgen.1002137. Epub 2011 Jul 7. PMID: 21750682; PMCID: PMC3131280.

Murphy DM, Buckley PG, Das S, Watters KM, Bryan K, Stallings RL. Co-localization of the oncogenic transcription factor MycN and the DNA methyl binding protein MeCP2 at genomic sites in neuroblastoma. PLoS One. 2011;6(6):e21436. doi: 10.1371/journal.pone.0021436. Epub 2011 Jun 22. PMID: 21731748; PMCID: PMC3120883.

Muñoz IM, Morgan ME, Peltier J, Weiland F, Gregorczyk M, Brown FC, Macartney T, Toth R, Trost M, Rouse J. Phosphoproteomic screening identifies physiological substrates of the CDKL5 kinase. EMBO J. 2018 Dec 14;37(24):e99559. doi: 10.15252/embj.201899559. Epub 2018 Sep 28. PMID: 30266825; PMCID: PMC6293279. F

Müller C, Schulte FW, Lange-Grünweller K, Obermann W, Madhugiri R, Pleschka S, Ziebuhr J, Hartmann RK, Grünweller A. Broad-spectrum antiviral activity of the eIF4A inhibitor silvestrol against corona- and picornaviruses. Antiviral Res. 2018 Feb;150:123-129. doi: 10.1016/j.antiviral.2017.12.010. Epub 2017 Dec 16. PMID: 29258862; PMCID: PMC7113723.

Nanbru C, Lafon I, Audigier S, Gensac MC, Vagner S, Huez G, Prats AC. Alternative translation of the proto-oncogene c-myc by an internal ribosome entry site. J Biol Chem. 1997 Dec 19;272(51):32061-6. doi: 10.1074/jbc.272.51.32061. PMID: 9405401.

Nectoux J, Girard B, Bahi-Buisson N, Prieur F, Afenjar A, Rosas-Vargas H, Chelly J, Bienvenu T. Netrin G1 mutations are an uncommon cause of atypical Rett syndrome with or without epilepsy. Pediatr Neurol. 2007 Oct;37(4):270-4. doi: 10.1016/j.pediatrneurol.2007.06.002. PMID: 17903671.

Nectoux J, Heron D, Tallot M, Chelly J, Bienvenu T. Maternal origin of a novel C-terminal truncation mutation in CDKL5 causing a severe atypical form of Rett syndrome. Clin Genet. 2006 Jul;70(1):29-33. doi: 10.1111/j.1399-0004.2006.00629.x. PMID: 16813600.

NEMER M. Interrelation of messenger polyribonucleotides and ribosomes in the sea urchin egg during embryonic development. Biochem Biophys Res Commun. 1962 Aug 31;8:511-5. doi: 10.1016/0006-291x(62)90307-8. PMID: 14478972.

Nemos C, Lambert L, Giuliano F, Doray B, Roubertie A, Goldenberg A, Delobel B, Layet V, N'guyen MA, Saunier A, Verneau F, Jonveaux P, Philippe C. Mutational spectrum of CDKL5 in early-onset encephalopathies: a study of a large collection of French patients and review of the literature. Clin Genet. 2009 Oct;76(4):357-71. doi:

10.1111/j.1399-0004.2009.01194.x. PMID: 19793311.

Nemos C, Lambert L, Giuliano F, Doray B, Roubertie A, Goldenberg A, Delobel B, Layet V, N'guyen MA, Saunier A, Verneau F, Jonveaux P, Philippe C. Mutational spectrum of CDKL5 in early-onset encephalopathies: a study of a large collection of French patients and review of the literature. Clin Genet. 2009 Oct;76(4):357-71. doi: 10.1111/j.1399-0004.2009.01194.x. PMID: 19793311.

Nishimura-Akiyoshi S, Niimi K, Nakashiba T, Itohara S. Axonal netrin-Gs transneuronally determine lamina-specific subdendritic segments. Proc Natl Acad Sci U S A. 2007 Sep 11;104(37):14801-6. doi: 10.1073/pnas.0706919104. Epub 2007 Sep 4. PMID: 17785411; PMCID: PMC1964543.

Noguchi S, Arakawa T, Fukuda S, Furuno M, Hasegawa A, Hori F, Ishikawa-Kato S, Kaida K, Kaiho A, Kanamori-Katayama M, Kawashima T, Kojima M, Kubosaki A, Manabe RI, Murata M, Nagao-Sato S, Nakazato K, Ninomiya N, Nishiyori-Sueki H, Noma S, Saijyo E, Saka A, Sakai M, Simon C, Suzuki N, Tagami M, Watanabe S, Yoshida S, Arner P, Axton RA, Babina M, Baillie JK, Barnett TC, Beckhouse AG, Blumenthal A, Bodega B, Bonetti A, Briggs J, Brombacher F, Carlisle AJ, Clevers HC, Davis CA, Detmar M, Dohi T, Edge ASB, Edinger M, Ehrlund A, Ekwall K, Endoh M, Enomoto H, Eslami A, Fagiolini M, Fairbairn L, Farach-Carson MC, Faulkner GJ, Ferrai C, Fisher ME, Forrester LM, Fujita R, Furusawa JI, Geijtenbeek TB, Gingeras T, Goldowitz D, Guhl S, Guler R, Gustincich S, Ha TJ, Hamaguchi M, Hara M, Hasegawa Y, Herlyn M, Heutink P, Hitchens KJ, Hume DA, Ikawa T, Ishizu Y, Kai C, Kawamoto H, Kawamura YI, Kempfle JS, Kenna TJ, Kere J, Khachigian LM, Kitamura T, Klein S, Klinken SP, Knox AJ, Kojima S, Koseki H, Koyasu S, Lee W, Lennartsson A, Mackay-Sim A, Mejhert N, Mizuno Y, Morikawa H, Morimoto M, Moro K, Morris KJ, Motohashi H, Mummery CL, Nakachi Y, Nakahara F, Nakamura T, Nakamura Y, Nozaki T, Ogishima S, Ohkura N, Ohno H, Ohshima M, Okada-Hatakeyama M, Okazaki Y, Orlando V, Ovchinnikov DA, Passier R, Patrikakis M, Pombo A, Pradhan-Bhatt S, Qin XY, Rehli M, Rizzu P, Roy S, Sajantila A, Sakaguchi S, Sato H, Satoh H, Savvi S, Saxena A, Schmidl C, Schneider C, Schulze-Tanzil GG, Schwegmann A, Sheng G, Shin JW, Sugiyama D, Sugiyama T, Summers KM, Takahashi N, Takai J, Tanaka H, Tatsukawa H, Tomoiu A, Toyoda H, van de Wetering M, van den Berg LM, Verardo R, Vijayan D, Wells CA, Winteringham LN, Wolvetang E, Yamaguchi Y, Yamamoto M, Yanagi-Mizuochi C, Yoneda M, Yonekura Y, Zhang PG, Zucchelli S, Abugessaisa I, Arner E, Harshbarger J, Kondo A, Lassmann T, Lizio M, Sahin S, Sengstag T, Severin J, Shimoji H, Suzuki M, Suzuki H, Kawai J, Kondo N, Itoh M, Daub CO, Kasukawa T, Kawaji H, Carninci P, Forrest ARR, Hayashizaki Y. FANTOM5 CAGE profiles of human and mouse samples. Sci Data. 2017 Aug 29;4:170112. doi: 10.1038/sdata.2017.112. PMID: 28850106; PMCID: PMC5574368.

Nogueras-Ortiz CJ, De Jesús-Cortes HJ, Vaquer-Alicea J, Vega IE. Novel autoimmune response in a tauopathy mouse model. Front Neurosci. 2014 Jan 10;7:277. doi: 10.3389/fnins.2013.00277. PMID: 24454278; PMCID: PMC3887318.

O'Connor T, Sadleir KR, Maus E, Velliquette RA, Zhao J, Cole SL, Eimer WA, Hitt B, Bembinster LA, Lammich S, Lichtenthaler SF, Hébert SS, De Strooper B, Haass C, Bennett DA, Vassar R. Phosphorylation of the translation initiation factor eIF2alpha increases BACE1 levels and promotes amyloidogenesis. Neuron. 2008 Dec 26;60(6):988-1009. doi: 10.1016/j.neuron.2008.10.047. PMID: 19109907; PMCID:

PMC2667382.

Oh SK, Scott MP, Sarnow P. Homeotic gene Antennapedia mRNA contains 5'-noncoding sequences that confer translational initiation by internal ribosome binding. Genes Dev. 1992 Sep;6(9):1643-53. doi: 10.1101/gad.6.9.1643. PMID: 1355457.

Ohno M. Roles of eIF2α kinases in the pathogenesis of Alzheimer's disease. Front Mol Neurosci. 2014 Apr 16;7:22. doi: 10.3389/fnmol.2014.00022. PMID: 24795560; PMCID: PMC3997008.

OHTAKA Y, SPIEGELMAN S. TRANSLATIONAL CONTROL OF PROTEIN SYNTHESIS IN A CELL-FREE SYSTEM DIRECTED BY A POLYCISTRONIC VIRAL RNA. Science. 1963 Oct 25;142(3591):493-7. doi: 10.1126/science.142.3591.493. PMID: 14064448.

Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H, Yamanaka I, Kiyosawa H, Yagi K, Tomaru Y, Hasegawa Y, Nogami A, Schönbach C, Gojobori T, Baldarelli R, Hill DP, Bult C, Hume DA, Quackenbush J, Schriml LM, Kanapin A, Matsuda H, Batalov S, Beisel KW, Blake JA, Bradt D, Brusic V, Chothia C, Corbani LE, Cousins S, Dalla E, Dragani TA, Fletcher CF, Forrest A, Frazer KS, Gaasterland T, Gariboldi M, Gissi C, Godzik A, Gough J, Grimmond S, Gustincich S, Hirokawa N, Jackson IJ, Jarvis ED, Kanai A, Kawaji H, Kawasawa Y, Kedzierski RM, King BL, Konagaya A, Kurochkin IV, Lee Y, Lenhard B, Lyons PA, Maglott DR, Maltais L, Marchionni L, McKenzie L, Miki H, Nagashima T, Numata K, Okido T, Pavan WJ, Pertea G, Pesole G, Petrovsky N, Pillai R, Pontius JU, Qi D, Ramachandran S, Ravasi T, Reed JC, Reed DJ, Reid J, Ring BZ, Ringwald M, Sandelin A, Schneider C, Semple CA, Setou M, Shimada K, Sultana R, Takenaka Y, Taylor MS, Teasdale RD, Tomita M, Verardo R, Wagner L, Wahlestedt C, Wang Y, Watanabe Y, Wells C, Wilming LG, Wynshaw-Boris A, Yanagisawa M, Yang I, Yang L, Yuan Z, Zavolan M, Zhu Y, Zimmer A, Carninci P, Hayatsu N, Hirozane-Kishikawa T, Konno H, Nakamura M, Sakazume N, Sato K, Shiraki T, Waki K, Kawai J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashizume W, Imotani K, Ishii Y, Itoh M, Kagawa I, Miyazaki A, Sakai K, Sasaki D, Shibata K, Shinagawa A, Yasunishi A, Yoshino M, Waterston R, Lander ES, Rogers J, Birney E, Hayashizaki Y; FANTOM Consortium; RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature. 2002 Dec 5;420(6915):563-73. doi: 10.1038/nature01266. PMID: 12466851.

Okuda K, Kobayashi S, Fukaya M, Watanabe A, Murakami T, Hagiwara M, Sato T, Ueno H, Ogonuki N, Komano-Inoue S, Manabe H, Yamaguchi M, Ogura A, Asahara H, Sakagami H, Mizuguchi M, Manabe T, Tanaka T. CDKL5 controls postsynaptic localization of GluN2B-containing NMDA receptors in the hippocampus and regulates seizure susceptibility. Neurobiol Dis. 2017 Oct;106:158-170. doi: 10.1016/j.nbd.2017.07.002. Epub 2017 Jul 6. PMID: 28688852.

Olsen DS, Savner EM, Mathew A, Zhang F, Krishnamoorthy T, Phan L, Hinnebusch AG. Domains of eIF1A that mediate binding to eIF2, eIF3 and eIF5B and promote ternary complex recruitment in vivo. EMBO J. 2003 Jan 15;22(2):193-204. doi: 10.1093/emboj/cdg030. PMID: 12514125; PMCID: PMC140105.

Olson HE, Demarest ST, Pestana-Knight EM, Swanson LC, Iqbal S, Lal D, Leonard H,

Cross JH, Devinsky O, Benke TA. Cyclin-Dependent Kinase-Like 5 Deficiency Disorder: Clinical Review. Pediatr Neurol. 2019 Aug;97:18-25. doi: 10.1016/j.pediatrneurol.2019.02.015. Epub 2019 Feb 23. PMID: 30928302; PMCID: PMC7120929.

Ozretić P, Bisio A, Musani V, Trnski D, Sabol M, Levanat S, Inga A. Regulation of human PTCH1b expression by different 5' untranslated region cis-regulatory elements. RNA Biol. 2015;12(3):290-304. doi: 10.1080/15476286.2015.1008929. PMID: 25826662; PMCID: PMC4615190.

Palii SS, Kays CE, Deval C, Bruhat A, Fafournoux P, Kilberg MS. Specificity of amino acid regulated gene expression: analysis of genes subjected to either complete or single amino acid deprivation. Amino Acids. 2009 May;37(1):79-88. doi: 10.1007/s00726-008-0199-2. Epub 2008 Nov 14. PMID: 19009228; PMCID: PMC3595573.

Pany SPP, Sapra M, Sharma J, Dhamodharan V, Patankar S, Pradeepkumar PI. Presence of Potential G-Quadruplex RNA-Forming Motifs at the 5'-UTR of PP2Acα mRNA Repress Translation. Chembiochem. 2019 Dec 2;20(23):2955-2960. doi: 10.1002/cbic.201900336. Epub 2019 Sep 26. PMID: 31206965.

Parsyan A, Svitkin Y, Shahbazian D, Gkogkas C, Lasko P, Merrick WC, Sonenberg N. mRNA helicases: the tacticians of translational control. Nat Rev Mol Cell Biol. 2011 Apr;12(4):235-45. doi: 10.1038/nrm3083. PMID: 21427765.

Pedersen AG, Nielsen H. Neural network prediction of translation initiation sites in eukaryotes: perspectives for EST and genome analysis. Proc Int Conf Intell Syst Mol Biol. 1997;5:226-33. PMID: 9322041.

Peer E, Moshitch-Moshkovitz S, Rechavi G, Dominissini D. The Epitranscriptome in Translation Regulation. Cold Spring Harb Perspect Biol. 2019 Aug 1;11(8):a032623. doi: 10.1101/cshperspect.a032623. PMID: 30037968; PMCID: PMC6671940.

Pelham HR, Jackson RJ. An efficient mRNA-dependent translation system from reticulocyte lysates. Eur J Biochem. 1976 Aug 1;67(1):247-56. doi: 10.1111/j.1432-1033.1976.tb10656.x. PMID: 823012.

Pelletier J, Sonenberg N. Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translational efficiency. Cell. 1985 Mar;40(3):515-26. doi: 10.1016/0092-8674(85)90200-4. PMID: 2982496.

Pelletier J, Sonenberg N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Nature. 1988 Jul 28;334(6180):320-5. doi: 10.1038/334320a0. PMID: 2839775.

Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S. Structural and functional features of eukaryotic mRNA untranslated regions. Gene. 2001 Oct 3;276(1-2):73-81. doi: 10.1016/s0378-1119(01)00674-6. PMID: 11591473.

Pestova TV, Hellen CU, Shatsky IN. Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. Mol Cell Biol. 1996

Dec;16(12):6859-69. doi: 10.1128/MCB.16.12.6859. PMID: 8943341; PMCID: PMC231689.

Pfeiffer BE, Huber KM. The state of synapses in fragile X syndrome. Neuroscientist. 2009 Oct;15(5):549-67. doi: 10.1177/1073858409333075. Epub 2009 Mar 26. PMID: 19325170; PMCID: PMC2762019.

Pickering BM, Mitchell SA, Spriggs KA, Stoneley M, Willis AE. Bag-1 internal ribosome entry segment activity is promoted by structural changes mediated by poly(rC) binding protein 1 and recruitment of polypyrimidine tract binding protein 1. Mol Cell Biol. 2004 Jun;24(12):5595-605. doi: 10.1128/MCB.24.12.5595-5605.2004. PMID: 15169918; PMCID: PMC419896.

Pini G, Bigoni S, et al. Variant of Rett syndrome and CDKL5 gene: clinical and autonomic description of 10 cases. Neuropediatrics. 2012 Feb;43(1):37-43. doi:10.1055/s-0032-1308856. Epub 2012 Mar 19. Erratum in: Neuropediatrics. 2013 Aug;44(4):237. PMID: 22430159.

Pradhan SK, Su T, Yen L, Jacquet K, Huang C, Côté J, Kurdistani SK, Carey MF. EP400 Deposits H3.3 into Promoters and Enhancers during Gene Activation. Mol Cell. 2016 Jan 7;61(1):27-38. doi: 10.1016/j.molcel.2015.10.039. Epub 2015 Dec 6. PMID: 26669263; PMCID: PMC4707986.

Prats AC, Prats H. Translational control of gene expression: role of IRESs and consequences for cell transformation and angiogenesis. Prog Nucleic Acid Res Mol Biol. 2002;72:367-413. doi: 10.1016/s0079-6603(02)72075-8. PMID: 12206457.

Proud CG. Phosphorylation and Signal Transduction Pathways in Translational Control. Cold Spring Harb Perspect Biol. 2019 Jul 1;11(7):a033050. doi: 10.1101/cshperspect.a033050. PMID: 29959191; PMCID: PMC6601458.

Péladeau C, Jasmin BJ. Targeting IRES-dependent translation as a novel approach for treating Duchenne muscular dystrophy. RNA Biol. 2021 Sep;18(9):1238-1251. doi: 10.1080/15476286.2020.1847894. Epub 2020 Nov 19. PMID: 33164678; PMCID: PMC8354616.

Randazzo A, Spada GP, da Silva MW. Circular dichroism of quadruplex structures. Top Curr Chem. 2013;330:67-86. doi: 10.1007/128_2012_331. PMID: 22752576.

Ricciardi S, Ungaro F, Hambrock M, Rademacher N, Stefanelli G, Brambilla D, Sessa A, Magagnotti C, Bachi A, Giarda E, Verpelli C, Kilstrup-Nielsen C, Sala C, Kalscheuer VM, Broccoli V. CDKL5 ensures excitatory synapse stability by reinforcing NGL-1-PSD95 interaction in the postsynaptic compartment and is impaired in patient iPSC-derived neurons. Nat Cell Biol. 2012 Sep;14(9):911-23. doi: 10.1038/ncb2566. Epub 2012 Aug 26. PMID: 22922712; PMCID: PMC6485419.

Ricciardi S, Ungaro F, Hambrock M, Rademacher N, Stefanelli G, Brambilla D, Sessa A, Magagnotti C, Bachi A, Giarda E, Verpelli C, Kilstrup-Nielsen C, Sala C, Kalscheuer VM, Broccoli V. CDKL5 ensures excitatory synapse stability by reinforcing NGL-1-PSD95 interaction in the postsynaptic compartment and is impaired in patient iPSC-derived neurons. Nat Cell Biol. 2012 Sep;14(9):911-23. doi:

10.1038/ncb2566. Epub 2012 Aug 26. PMID: 22922712; PMCID: PMC6485419.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May;17(5):405-24. doi: 10.1038/gim.2015.30. Epub 2015 Mar 5. PMID: 25741868; PMCID: PMC4544753.

Richards SEV, Moore AR, Nam AY, Saxena S, Paradis S, Van Hooser SD. Experience-Dependent Development of Dendritic Arbors in Mouse Visual Cortex. J Neurosci. 2020 Aug 19;40(34):6536-6556. doi: 10.1523/JNEUROSCI.2910-19.2020. Epub 2020 Jul 15. PMID: 32669356; PMCID: PMC7486652.

Richter JD, Bassell GJ, Klann E. Dysregulation and restoration of translational homeostasis in fragile X syndrome. Nat Rev Neurosci. 2015 Oct;16(10):595-605. doi: 10.1038/nrn4001. Epub 2015 Sep 9. PMID: 26350240; PMCID: PMC4688896.

Rigo R, Sissi C. Characterization of G4-G4 Crosstalk in the c-KIT Promoter Region. Biochemistry. 2017 Aug 22;56(33):4309-4312. doi: 10.1021/acs.biochem.7b00660. Epub 2017 Aug 8. PMID: 28763217.

Robichaud N, Sonenberg N, Ruggero D, Schneider RJ. Translational Control in Cancer. Cold Spring Harb Perspect Biol. 2019 Jul 1;11(7):a032896. doi: 10.1101/cshperspect.a032896. PMID: 29959193; PMCID: PMC6601465.

Rolfe DF, Brown GC. Cellular energy utilization and molecular origin of standard metabolic rate in mammals. Physiol Rev. 1997 Jul;77(3):731-58. doi: 10.1152/physrev.1997.77.3.731. PMID: 9234964.

Rouleau SG, Beaudoin JD, Bisaillon M, Perreault JP. Small antisense oligonucleotides against G-quadruplexes: specific mRNA translational switches. Nucleic Acids Res. 2015 Jan;43(1):595-606. doi: 10.1093/nar/gku1311. Epub 2014 Dec 15. PMID: 25510493; PMCID: PMC4288198.

Rowlands AG, Panniers R, Henshaw EC. The catalytic mechanism of guanine nucleotide exchange factor action and competitive inhibition by phosphorylated eukaryotic initiation factor 2. J Biol Chem. 1988 Apr 25;263(12):5526-33. PMID: 3356695.

Rozen F, Edery I, Meerovitch K, Dever TE, Merrick WC, Sonenberg N. Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. Mol Cell Biol. 1990 Mar;10(3):1134-44. doi: 10.1128/mcb.10.3.1134-1144.1990. PMID: 2304461; PMCID: PMC360981.

Rubio CA, Weisburd B, Holderfield M, Arias C, Fang E, DeRisi JL, Fanidi A. Transcriptome-wide characterization of the eIF4A signature highlights plasticity in translation regulation. Genome Biol. 2014;15(10):476. doi: 10.1186/s13059-014-0476-1. PMID: 25273840; PMCID: PMC4203936.

Rusconi L, Salvatoni L, Giudici L, et al. CDKL5 expression is modulated during neuronal development and its subcellular distribution is tightly regulated by the C-terminal tail. J Biol Chem. 2008;283(44):30101-30111. doi:10.1074/jbc.M804613200

Sabarinathan R, Tafer H, Seemann SE, Hofacker IL, Stadler PF, Gorodkin J. The RNAsnp web server: predicting SNP effects on local RNA secondary structure. Nucleic Acids Res. 2013 Jul;41(Web Server issue):W475-9. doi: 10.1093/nar/gkt291. Epub 2013 Apr 29. PMID: 23630321; PMCID: PMC3977658.

Santini E, Klann E. Reciprocal signaling between translational control pathways and synaptic proteins in autism spectrum disorders. Sci Signal. 2014 Oct 28;7(349):re10. doi: 10.1126/scisignal.2005832. PMID: 25351249; PMCID: PMC6002803.

Sasaki S, Sun R, Bui HH, Crosby JR, Monia BP, Guo S. Steric Inhibition of 5' UTR Regulatory Elements Results in Upregulation of Human CFTR. Mol Ther. 2019 Oct 2;27(10):1749-1757. doi: 10.1016/j.ymthe.2019.06.016. Epub 2019 Jul 12. PMID: 31351782; PMCID: PMC6822282.

Sato K, Hamada M, Asai K, Mituyama T. CENTROIDFOLD: a web server for RNA secondary structure prediction. Nucleic Acids Res. 2009 Jul;37(Web Server issue):W277-80. doi: 10.1093/nar/gkp367. Epub 2009 May 12. PMID: 19435882; PMCID: PMC2703931.

Schedel J, Distler O, Woenckhaus M, Gay RE, Simmen B, Michel BA, Müller-Ladner U, Gay S. Discrepancy between mRNA and protein expression of tumour suppressor maspin in synovial tissue may contribute to synovial hyperplasia in rheumatoid arthritis. Ann Rheum Dis. 2004 Oct;63(10):1205-11. doi: 10.1136/ard.2003.006312. PMID: 15361372; PMCID: PMC1754744

Sekiguchi M., Katayama S., Hatano N., Shigeri Y., Sueyoshi N., Kameshita I. Identification of amphiphysin 1 as an endogenous substrate for CDKL5, a protein kinase associated with X-linked neurodevelopmental disorder. Archives of Biochemistry and Biophysics. 2013;535(2):257–267. doi: 10.1016/j.abb.2013.04.012.

Senga Y, Nagamine T, Sekiguchi M, Kaneko K, Sueyoshi N, Kameshita I. Detection of protein kinase substrates in tissue extracts after separation by isoelectric focusing. Anal Biochem. 2011 Jan 15;408(2):345-7. doi: 10.1016/j.ab.2010.08.036. Epub 2010 Aug 31. PMID: 20807496.

Sethna F, Moon C, Wang H. From FMRP function to potential therapies for fragile X syndrome. Neurochem Res. 2014 Jun;39(6):1016-31. doi: 10.1007/s11064-013-1229-3. Epub 2013 Dec 18. PMID: 24346713; PMCID: PMC4024105.

Shahbazian D, Parsyan A, Petroulakis E, Topisirovic I, Martineau Y, Gibbs BF, Svitkin Y, Sonenberg N. Control of cell survival and proliferation by mammalian eukaryotic initiation factor 4B. Mol Cell Biol. 2010 Mar;30(6):1478-85. doi: 10.1128/MCB.01218-09. Epub 2010 Jan 19. PMID: 20086100; PMCID: PMC2832492.

Sherrill KW, Lloyd RE. Translation of cIAP2 mRNA is mediated exclusively by a stress-modulated ribosome shunt. Mol Cell Biol. 2008 Mar;28(6):2011-22. doi: 10.1128/MCB.01446-07. Epub 2008 Jan 14. PMID: 18195037; PMCID: PMC2268400.

Shi Y, Yang Y, Hoang B, Bardeleben C, Holmes B, Gera J, Lichtenstein A. Therapeutic potential of targeting IRES-dependent c-myc translation in multiple myeloma cells during ER stress. Oncogene. 2016 Feb 25;35(8):1015-24. doi: 10.1038/onc.2015.156. Epub 2015 May 11. PMID: 25961916; PMCID: PMC5104155.

Silva J, Fernandes R, Romão L. Translational Regulation by Upstream Open Reading Frames and Human Diseases. Adv Exp Med Biol. 2019;1157:99-116. doi: 10.1007/978-3-030-19966-1_5. PMID: 31342439.

Simone R, Fratta P, Neidle S, Parkinson GN, Isaacs AM. G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome. FEBS Lett. 2015 Jun 22;589(14):1653-68. doi: 10.1016/j.febslet.2015.05.003. Epub 2015 May 13. PMID: 25979

Sonenberg N, Hershey JWB, Mathews MB. Translational control of gene expression. Cold Spring Harbor Laboratory Press, 2000

Spriggs KA, Cobbold LC, Jopling CL, Cooper RE, Wilson LA, Stoneley M, Coldwell MJ, Poncet D, Shen YC, Morley SJ, Bushell M, Willis AE. Canonical initiation factor requirements of the Myc family of internal ribosome entry segments. Mol Cell Biol. 2009 Mar;29(6):1565-74. doi: 10.1128/MCB.01283-08. Epub 2009 Jan 5. PMID: 19124605; PMCID: PMC2648239.

Spriggs KA, Bushell M, Willis AE. Translational regulation of gene expression during conditions of cell stress. Mol Cell. 2010 Oct 22;40(2):228-37. doi: 10.1016/j.molcel.2010.09.028. PMID: 20965418.

Steward DL, Shaeffer JR, Humphrey RM. Breakdown and assembly of polyribosomes in synchronized Chinese hamster cells. Science. 1968 Aug 23;161(3843):791-3. doi: 10.1126/science.161.3843.791. PMID: 5663806.

Steward O, Levy WB. Preferential localization of polyribosomes under the base of dendritic spines in granule cells of the dentate gyrus. J Neurosci. 1982 Mar;2(3):284-91. doi: 10.1523/JNEUROSCI.02-03-00284.1982. PMID: 7062109; PMCID: PMC6564334.

Stoneley M, Chappell SA, Jopling CL, Dickens M, MacFarlane M, Willis AE. c-Myc protein synthesis is initiated from the internal ribosome entry segment during apoptosis. Mol Cell Biol. 2000 Feb;20(4):1162-9. doi: 10.1128/MCB.20.4.1162-1169.2000. PMID: 10648601; PMCID: PMC85234.

Stoneley M, Paulin FE, Le Quesne JP, Chappell SA, Willis AE. C-Myc 5' untranslated region contains an internal ribosome entry segment. Oncogene. 1998 Jan 22;16(3):423-8. doi: 10.1038/sj.onc.1201763. PMID: 9467968.

Stoneley M, Willis AE. Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. Oncogene. 2004 Apr 19;23(18):3200-7. doi: 10.1038/sj.onc.1207551. PMID: 15094769.

Subramanian M, Rage F, Tabet R, Flatter E, Mandel JL, Moine H. G-quadruplex RNA

structure as a signal for neurite mRNA targeting. EMBO Rep. 2011 Jul 1;12(7):697-704. doi: 10.1038/embor.2011.76. PMID: 21566646; PMCID: PMC3128965.

Sugiyama Y, Hatano N, Sueyoshi N, Suetake I, Tajima S, Kinoshita E, Kinoshita-Kikuta E, Koike T, Kameshita I. The DNA-binding activity of mouse DNA methyltransferase 1 is regulated by phosphorylation with casein kinase 1delta/epsilon. Biochem J. 2010 Apr 14;427(3):489-97. doi: 10.1042/BJ20091856. PMID: 20192920.

Supekar K, Uddin LQ, Khouzam A, Phillips J, Gaillard WD, Kenworthy LE, Yerys BE, Vaidya CJ, Menon V. Brain hyperconnectivity in children with autism and its links to social deficits. Cell Rep. 2013 Nov 14;5(3):738-47. doi: 10.1016/j.celrep.2013.10.001. Epub 2013 Nov 7. PMID: 24210821; PMCID: PMC3894787.

Tahmasebi S, Sonenberg N, Hershey JWB, Mathews MB. Protein Synthesis and Translational Control: A Historical Perspective. Cold Spring Harb Perspect Biol. 2019 Sep 3;11(9):a035584. doi: 10.1101/cshperspect.a035584. PMID: 30082466; PMCID: PMC6719597.

Takahashi H, Kato S, Murata M, Carninci P. CAGE (cap analysis of gene expression): a protocol for the detection of promoter and transcriptional networks. Methods Mol Biol. 2012;786:181-200. doi: 10.1007/978-1-61779-292-2_11. PMID: 21938627; PMCID: PMC4094367.

Takahashi H, Lassmann T, Murata M, Carninci P. 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. Nat Protoc. 2012 Feb 23;7(3):542-61. doi: 10.1038/nprot.2012.005. PMID: 22362160; PMCID: PMC4094379.

Takahashi S, Sugimoto N. Watson-Crick versus Hoogsteen Base Pairs: Chemical Strategy to Encode and Express Genetic Information in Life. Acc Chem Res. 2021 May 4;54(9):2110-2120. doi: 10.1021/acs.accounts.0c00734. Epub 2021 Feb 16. PMID: 33591181.

Tatsumoto S, Adati N, Tohtoki Y, Sakaki Y, Boroviak T, Habu S, Okano H, Suemizu H, Sasaki E, Satake M. Development and characterization of cDNA resources for the common marmoset: one of the experimental primate models. DNA Res. 2013 Jun;20(3):255-62. doi: 10.1093/dnares/dst007. Epub 2013 Mar 29. PMID: 23543116; PMCID: PMC3686431.

Tegha-Dunghu J, Bausch E, Neumann B, Wuensche A, Walter T, Ellenberg J, Gruss OJ. MAP1S controls microtubule stability throughout the cell cycle in human cells. J Cell Sci. 2014 Dec 1;127(Pt 23):5007-13. doi: 10.1242/jcs.136457. Epub 2014 Oct 9. PMID: 25300793.

Teleman AA, Chen YW, Cohen SM. 4E-BP functions as a metabolic brake used under stress conditions but not during normal growth. Genes Dev. 2005 Aug 15;19(16):1844-8. doi: 10.1101/gad.341505. PMID: 16103212; PMCID: PMC1186183.

Tennyson VM. The fine structure of the axon and growth cone of the dorsal root neuroblast of the rabbit embryo. J Cell Biol. 1970 Jan;44(1):62-79. doi: 10.1083/jcb.44.1.62. PMID: 5409464; PMCID: PMC2107779.

Thomas GP, Mathews MB. Alterations of transcription and translation in HeLa cells exposed to amino acid analogs. Mol Cell Biol. 1984 Jun;4(6):1063-72. doi: 10.1128/mcb.4.6.1063-1072.1984. PMID: 6610822; PMCID: PMC368874.

Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics. 2002 Aug;Chapter 2:Unit 2.3. doi: 10.1002/0471250953.bi0203s00. PMID: 18792934.

Tomkins GM, Garren LD, Howell RR, Peterkofsky B. The regulation of enzyme synthesis by steroid hormones: the role of translation. J Cell Physiol. 1965 Oct;66(2):Suppl 1:137-51. doi: 10.1002/jcp.1030660413. PMID: 4379390.

Torrance V, Lydall D. Overlapping open reading frames strongly reduce human and yeast STN1 gene expression and affect telomere function. PLoS Genet. 2018 Aug 1;14(8):e1007523. doi: 10.1371/journal.pgen.1007523. PMID: 30067734; PMCID: PMC6089452.

Trazzi S, Fuchs C, Viggiano R, De Franceschi M, Valli E, Jedynak P, Hansen FK, Perini G, Rimondini R, Kurz T, Bartesaghi R, Ciani E. HDAC4: a key factor underlyingbrain developmental alterations in CDKL5 disorder. Hum Mol Genet. 2016 Sep 15;25(18):3887-3907. doi: 10.1093/hmg/ddw231. Epub 2016 Jul 27. PMID: 27466189.

Trotta E. On the normalization of the minimum free energy of RNAs by sequence length. PLoS One. 2014 Nov 18;9(11):e113380. doi: 10.1371/journal.pone.0113380. PMID: 25405875; PMCID: PMC4236180.

Vagner S, Gensac MC, Maret A, Bayard F, Amalric F, Prats H, Prats AC. Alternative translation of human fibroblast growth factor 2 mRNA occurs by internal entry of ribosomes. Mol Cell Biol. 1995 Jan;15(1):35-44. doi: 10.1128/MCB.15.1.35. PMID: 7799942; PMCID: PMC231905.

Valli E, Trazzi S, Fuchs C, Erriquez D, Bartesaghi R, Perini G, Ciani E. CDKL5, a novel MycN-repressed gene, blocks cell cycle and promotes differentiation of neuronal cells. Biochim Biophys Acta. 2012 Nov-Dec;1819(11-12):1173-85. doi: 10.1016/j.bbagrm.2012.08.001. Epub 2012 Aug 19. PMID: 22921766; PMCID: PMC3787793.

Van de Sompele S, Pécheux L, Couso J, Meunier A, Sanchez M, De Baere E.Functional characterization of a novel non-coding mutation "Ghent +49A > G" in the iron-responsive element of L-ferritin causing hereditary hyperferritinaemia-cataract syndrome. Sci Rep. 2017 Dec 21;7(1):18025. doi: 10.1038/s41598-017-18326-6. PMID: 29269865; PMCID: PMC5740175.

Van Esch H, Jansen A, Bauters M, Froyen G, Fryns JP. Encephalopathy and bilateral cataract in a boy with an interstitial deletion of Xp22 comprising the CDKL5 and NHS genes. Am J Med Genet A. 2007 Feb 15;143(4):364-9. doi: 10.1002/ajmg.a.31572. PMID: 17256798.

von Bohlen AE, Böhm J, Pop R, Johnson DS, Tolmie J, Stücker R, Morris-Rosendahl D, Scherer G. A mutation creating an upstream initiation codon in the SOX9 5' UTR

causes acampomelic campomelic dysplasia. Mol Genet Genomic Med. 2017 Mar 21;5(3):261-268. doi: 10.1002/mgg3.282. PMID: 28546996; PMCID: PMC5441400.

Wang H, Wang Y, Yang J, Zhao Q, Tang N, Chen C, Li H, Cheng C, Xie M, Yang Y, Xie Z. Tissue- and stage-specific landscape of the mouse translatome. Nucleic Acids Res. 2021 Jun 21;49(11):6165-6180. doi: 10.1093/nar/gkab482. PMID: 34107020; PMCID: PMC8216458

Wang IT, Allen M, Goffin D, Zhu X, Fairless AH, Brodkin ES, Siegel SJ, Marsh ED, Blendy JA, Zhou Z. Loss of CDKL5 disrupts kinome profile and event-related potentials leading to autistic-like phenotypes in mice. Proc Natl Acad Sci U S A. 2012 Dec 26;109(52):21516-21. doi: 10.1073/pnas.1216988110. Epub 2012 Dec 10. PMID: 23236174; PMCID: PMC3535652.

Wang T, Cui Y, Jin J, Guo J, Wang G, Yin X, He QY, Zhang G. Translating mRNAs strongly correlate to proteins in a multivariate manner and their translation ratios are phenotype specific. Nucleic Acids Res. 2013 May;41(9):4743-54. doi: 10.1093/nar/gkt178. Epub 2013 Mar 21. PMID: 23519614; PMCID: PMC3643591

Wang XQ, Rothnagel JA. 5'-untranslated regions with multiple upstream AUG codons can support low-level translation via leaky scanning and reinitiation. Nucleic Acids Res. 2004 Feb 27;32(4):1382-91. doi: 10.1093/nar/gkh305. PMID: 14990743; PMCID: PMC390293.

Wang Y, Hou L, Ardehali MB, Kingston RE, Dynlacht BD. Elongin A regulates transcription in vivo through enhanced RNA polymerase processivity. J Biol Chem. 2021 Jan-Jun;296:100170. doi: 10.1074/jbc.RA120.015876. Epub 2020 Dec 13. PMID: 33298525; PMCID: PMC7948402.

Warnakulasuriyarachchi D, Cerquozzi S, Cheung HH, Holcík M. Translational induction of the inhibitor of apoptosis protein HIAP2 during endoplasmic reticulum stress attenuates cell death and is mediated via an inducible internal ribosome entry site element. J Biol Chem. 2004 Apr 23;279(17):17148-57. doi: 10.1074/jbc.M308737200. Epub 2004 Feb 11. PMID: 14960583.

Washbourne P. Synapse assembly and neurodevelopmental disorders. Neuropsychopharmacology. 2015 Jan;40(1):4-15. doi: 10.1038/npp.2014.163. Epub 2014 Jul 3. PMID: 24990427; PMCID: PMC4262893.

Wek RC. Role of eIF2α Kinases in Translational Control and Adaptation to Cellular Stress. Cold Spring Harb Perspect Biol. 2018 Jul 2;10(7):a032870. doi: 10.1101/cshperspect.a032870. PMID: 29440070; PMCID: PMC6028073.

Wethmar K, Bégay V, Smink JJ, Zaragoza K, Wiesenthal V, Dörken B, Calkhoven CF, Leutz A. C/EBPbetaDeltauORF mice--a genetic model for uORF-mediated translational control in mammals. Genes Dev. 2010 Jan 1;24(1):15-20. doi: 10.1101/gad.557910. PMID: 20047998; PMCID: PMC2802187.

Wethmar K, Smink JJ, Leutz A. Upstream open reading frames: molecular switches in (patho)physiology. Bioessays. 2010 Oct;32(10):885-93. doi: 10.1002/bies.201000037. Epub 2010 Aug 19. PMID: 20726009; PMCID: PMC3045505.

Williamson SL, Giudici L, Kilstrup-Nielsen C, Gold W, Pelka GJ, Tam PP, Grimm A, Prodi D, Landsberger N, Christodoulou J. A novel transcript of cyclin-dependent kinase-like 5 (CDKL5) has an alternative C-terminus and is the predominant transcript in brain. Hum Genet. 2012 Feb;131(2):187-200. doi: 10.1007/s00439-011-1058-x. Epub 2011 Jul 12. PMID: 21748340.

Winkelsas AM, Grunseich C, Harmison GG, Chwalenia K, Rinaldi C, Hammond SM, Johnson K, Bowerman M, Arya S, Talbot K, Wood MJ, Fischbeck KH. Targeting the 5' untranslated region of SMN2 as a therapeutic strategy for spinal muscular atrophy. Mol Ther Nucleic Acids. 2021 Jan 5;23:731-742. doi: 10.1016/j.omtn.2020.12.027. PMID: 33575118; PMCID: PMC7851419.

Winkelsas AM, Grunseich C, Harmison GG, Chwalenia K, Rinaldi C, Hammond SM, Johnson K, Bowerman M, Arya S, Talbot K, Wood MJ, Fischbeck KH. Targeting the 5' untranslated region of SMN2 as a therapeutic strategy for spinal muscular atrophy. Mol Ther Nucleic Acids. 2021 Jan 5;23:731-742. doi: 10.1016/j.omtn.2020.12.027. PMID: 33575118; PMCID: PMC7851419.

Wolfe AL, Singh K, Zhong Y, Drewe P, Rajasekhar VK, Sanghvi VR, Mavrakis KJ, Jiang M, Roderick JE, Van der Meulen J, Schatz JH, Rodrigo CM, Zhao C, Rondou P, de Stanchina E, Teruya-Feldstein J, Kelliher MA, Speleman F, Porco JA Jr, Pelletier J, Rätsch G, Wendel HG. RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. Nature. 2014 Sep 4;513(7516):65-70. doi: 10.1038/nature13485. Epub 2014 Jul 27. PMID: 25079319; PMCID: PMC4492470.

Xu J, Patassini S, Rustogi N, Riba-Garcia I, Hale BD, Phillips AM, Waldvogel H, Haines R, Bradbury P, Stevens A, Faull RLM, Dowsey AW, Cooper GJS, Unwin RD. Regional protein expression in human Alzheimer's brain correlates with disease severity. Commun Biol. 2019 Feb 4;2:43. doi: 10.1038/s42003-018-0254-9. PMID: 30729181; PMCID: PMC6361956.

Xue S, Tian S, Fujii K, Kladwang W, Das R, Barna M. RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation. Nature. 2015 Jan 1;517(7532):33-8. doi: 10.1038/nature14010. Epub 2014 Nov 19. PMID: 25409156; PMCID: PMC4353651.

Yang DQ, Halaby MJ, Zhang Y. The identification of an internal ribosomal entry site in the 5'-untranslated region of p53 mRNA provides a novel mechanism for the regulation of its translation following DNA damage. Oncogene. 2006 Aug 3;25(33):4613-9. doi: 10.1038/sj.onc.1209483. Epub 2006 Apr 10. PMID: 16607284.

Yang Y, Wang C, Zhao K, Zhang G, Wang D, Mei Y. TRMP, a p53-inducible long noncoding RNA, regulates G1/S cell cycle progression by modulating IRES-dependent p27 translation. Cell Death Dis. 2018 Aug 30;9(9):886. doi: 10.1038/s41419-018-0884-3. PMID: 30166522; PMCID: PMC6117267.

Yang Y, Wang Z. IRES-mediated cap-independent translation, a path leading to hidden proteome. J Mol Cell Biol. 2019 Oct 25;11(10):911-919. doi: 10.1093/jmcb/mjz091. PMID: 31504667; PMCID: PMC6884710.

Ye Y, Liang Y, Yu Q, Hu L, Li H, Zhang Z, Xu X. Analysis of human upstream open

reading frames and impact on gene expression. Hum Genet. 2015 Jun;134(6):605-12. doi: 10.1007/s00439-015-1544-7. Epub 2015 Mar 24. PMID: 25800702.

Yennawar M, White RS, Jensen FE. AMPA Receptor Dysregulation and Therapeutic Interventions in a Mouse Model of CDKL5 Deficiency Disorder. J Neurosci. 2019 Jun 12;39(24):4814-4828. doi: 10.1523/JNEUROSCI.2041-18.2019. Epub 2019 Apr 5. PMID: 30952813; PMCID: PMC6561688.

Zeraati M, Moye AL, Wong JWH, Perera D, Cowley MJ, Christ DU, Bryan TM, Dinger ME. Cancer-associated noncoding mutations affect RNA G-quadruplex-mediated regulation of gene expression. Sci Rep. 2017 Apr 6;7(1):708. doi: 10.1038/s41598-017-00739-y. PMID: 28386116; PMCID: PMC5429658.

Zhou Y, Koelling N, Fenwick AL, McGowan SJ, Calpena E, Wall SA, Smithson SF, Wilkie AOM, Twigg SRF. Disruption of TWIST1 translation by 5' UTR variants in Saethre-Chotzen syndrome. Hum Mutat. 2018 Oct;39(10):1360-1365. doi: 10.1002/humu.23598. Epub 2018 Aug 7. PMID: 30040876; PMCID: PMC6175480.

Zhu YC, Xiong ZQ. Molecular and Synaptic Bases of CDKL5 Disorder. Dev Neurobiol. 2019 Jan;79(1):8-19. doi: 10.1002/dneu.22639. Epub 2018 Oct 19. PMID: 30246934.

9 APPENDIX

9.1 Consurf Analysis Raw Data

In this paragraph we reported the nucleic acid conservation scores (color score) of the sequence Ex1_202up obtained from Consurf (**Tab. 9.1.1**). This data was employed to perform the conservation analysis provided in the **par. 5.6**. The sequences of the alignment used for feeding the tool are listed in **Tab. 9.1.2**, where are reported the species of provenience and the phylogenetic group of belonging for each sequence. These sequences were obtained through nBLAST search described in **par. 5.4.4**. Incomplete sequences (with no complete query coverage) were completed trough manual research in NCBI Genome Data Viewer.

POS	SEQ	SCORE	COLOR	MSA DATA	RESIDUE VARIETY
1	G	0.055	5	66/66	G,T,C
2	C	-0.809	9	66/66	C
3	T	2.138	1	66/66	G,T,C
4	T	0.585	2	66/66	T,C,A
5	C	0.746	1	66/66	G,C,T
6	T	2.334	1	66/66	T,A,G
7	G	0.614	2	66/66	G,T
8	C	0.231	4	66/66	A,T,C
9	T	1.64	1	66/66	C,T,G,A
10	A	-0.454	8	66/66	G,A
11	G	-0.395	7	66/66	G,T
12	A	-0.809	9	66/66	A
13	G	-0.394	7	66/66	A,G
14	G	-0.809	9	66/66	G
15	G	-0.809	9	66/66	G
16	C	-0.809	9	66/66	C
17	G	-0.809	9	66/66	G
18	G	-0.809	9	66/66	G
19	G	-0.809	9	66/66	G
20	G	-0.809	9	66/66	G
21	C	0.637	1	66/66	C,T
22	C	-0.128	6	66/66	T,C,A
23	G	0.027	5	66/66	G,A
24	G	-0.399	7	66/66	A,G
25	A	-0.461	8	66/66	A,G
26	G	-0.809	9	66/66	G
27	G	0.058	5	66/66	G,A,C
28	T	0.624	2	66/66	A,G,C,T
29	T	-0.809	9	66/66	T

30	T	-0.358	7	66/66	T,C
31	C	-0.472	8	66/66	T,C
32	G	0.023	5	66/66	G,C,T
33	A	0.678	1	66/66	T,A,G
34	T	-0.809	9	66/66	T
35	T	-0.809	9	66/66	T
36	A	-0.809	9	66/66	A
37	G	-0.392	7	66/66	G,C
38	T	-0.809	9	66/66	T
39	T	-0.334	7	66/66	T,G
40	G	-0.384	7	66/66	G,A
41	T	-0.363	7	66/66	C,T
42	C	0.743	1	66/66	C,T
43	T	3.28	1	66/66	C,T,A,G
44	C	0.725	1	66/66	T,C,G
45	T	0.102	4	66/66	T,C
46	G	-0.396	7	66/66	A,G
47	C	-0.809	9	66/66	C
48	C	0.647	1	66/66	T,C,G,A
49	G	1.016	1	66/66	A,G,T
50	C	-0.122	6	66/66	C,T
51	T	-0.809	9	66/66	T
52	G	-0.366	7	66/66	G,A
53	G	-0.392	7	66/66	G,T
54	G	-0.809	9	66/66	G
55	G	0.033	5	66/66	T,G,A
56	A	-0.435	7	66/66	T,A
57	A	-0.452	8	66/66	A,C
58	G	-0.395	7	66/66	A,G
59	G	-0.404	7	66/66	G,A
60	T	-0.339	7	66/66	G,T
61	A	-0.46	8	66/66	A,G
62	A	-0.075	5	66/66	A,C,T
63	A	-0.809	9	66/66	A
64	G	0.473	2	66/66	G,A,T
65	C	-0.465	8	66/66	C,A
66	G	0.544	2	66/66	A,G
67	G	-0.389	7	66/66	G,T
68	C	1.59	1	66/66	C,T,A
69	G	-0.389	7	66/66	A,G
70	A	-0.096	6	66/66	T,C,A
71	C	-0.47	8	66/66	C,G
72	G	-0.391	7	66/66	G,C
73	G	-0.809	9	66/66	G
74	C	0.233	4	66/66	A,C,T
75	G	-0.37	7	66/66	T,G
76	T	-0.389	7	66/66	T,G

77	C	0.612	2	66/66	G,T,C
78	C	1.537	1	66/66	G,C,T
79	T	0.149	4	66/66	G,T
80	C	-0.809	9	66/66	C
81	A	4.668	1	66/66	C,T,G,A
82	G	-0.383	7	66/66	A,G
83	G	-0.37	7	66/66	G,T
84	A	-0.1	6	66/66	A,G,C
85	G	-0.809	9	66/66	G
86	C	-0.443	7	66/66	C,T
87	T	2.269	1	66/66	T,C,G,A
88	G	1	1	66/66	A,G,T,C
89	T	0.596	2	66/66	A,T
90	G	0.052	5	66/66	G,A,T
91	G	0.057	5	66/66	G,C
92	G	0.441	3	66/66	C,A,G
93	G	2.026	1	66/66	C,T,G,A
94	T	0.557	2	66/66	G,T,C
95	C	-0.809	9	66/66	C
96	C	0.232	4	66/66	T,C
97	C	-0.101	6	66/66	C,A,G
98	C	2.504	1	66/66	G,A,C,T
99	T	2.479	1	66/66	A,G,C,T
100	G	0.84	1	66/66	G,C
101	C	0.019	5	66/66	T,C
102	T	0.638	1	66/66	T,C,A
103	A	0.524	2	66/66	G,A
104	G	-0.373	7	66/66	A,G
105	A	-0.058	5	66/66	C,A,G
106	A	0.035	5	66/66	A,G
107	G	-0.37	7	66/66	C,G
108	T	2.584	1	66/66	C,T,A
109	G	-0.388	7	66/66	G,T
110	G	-0.389	7	66/66	T,G
111	G	-0.394	7	66/66	C,G
112	G	0.934	1	66/66	G,A
113	G	1.075	1	66/66	G,A,C,T
114	A	0.295	3	66/66	A,G,T
115	C	-0.064	5	66/66	T,C,A
116	T	0.199	4	66/66	G,T
117	C	0.258	4	66/66	T,C
118	G	1.08	1	66/66	G,A
119	G	0.608	2	66/66	T,C,G,A
120	C	0.892	1	66/66	C,T
121	G	1.738	1	66/66	T,G,A
122	G	-0.396	7	66/66	G,A
123	G	-0.387	7	66/66	G,A
124	G	0.032	5	66/66	G,A

125	G	1.483	1	66/66	G,A
126	A	-0.071	5	66/66	G,A
127	G	1.736	1	66/66	C,T,A,G
128	T	-0.809	9	66/66	T
129	C	0.293	3	66/66	C,T,A
130	A	-0.069	5	66/66	A,G,C
131	T	0.659	1	66/66	G,T,C
132	T	-0.809	9	66/66	T
133	T	-0.357	7	66/66	T,C
134	A	0.372	3	66/66	A,G,C,T
135	A	0.818	1	66/66	A,T,C
136	T	-0.809	9	66/66	T
137	A	1.366	1	66/66	T,A,G
138	C	0.281	3	66/66	A,T,C
139	T	0.681	1	66/66	T,A,G
140	T	0.118	4	66/66	T,C,G
141	C	0.275	3	66/66	T,C,G
142	A	-0.091	6	66/66	A,G
143	T	1.59	1	66/66	A,G,T,C
144	G	2.257	1	66/66	C,T,A,G
145	A	-0.085	5	66/66	A,G
146	T	0.653	1	66/66	T,C,G
147	T	-0.361	7	66/66	C,T
148	A	0.073	5	66/66	A,G,T
149	G	2.1	1	66/66	T,C,G,A
150	A	0.312	3	66/66	G,A,C
151	A	-0.46	8	66/66	A,T
152	C	0.265	4	66/66	G,T,C
153	A	-0.452	8	66/66	A,G
154	A	-0.1	6	66/66	G,A
155	A	0.741	1	66/66	A,G
156	T	-0.36	7	66/66	A,T
157	A	-0.442	7	66/66	A,G
158	T	-0.362	7	66/66	C,T
159	G	0.11	4	66/66	G,A
160	T	-0.338	7	66/66	A,T
161	G	-0.809	9	66/66	G
162	A	-0.809	9	66/66	A
163	A	0.309	3	66/66	C,A,G
164	A	-0.086	5	66/66	C,A,G
165	G	1.499	1	66/66	C,G,A
166	T	-0.362	7	66/66	T,C
167	T	0.096	4	66/66	T,C
168	C	-0.116	6	66/66	C,T
169	C	-0.809	9	66/66	C
170	C	0.074	5	66/66	C,T
171	A	0.433	3	66/66	G,A
172	C	0.643	1	66/66	T,C

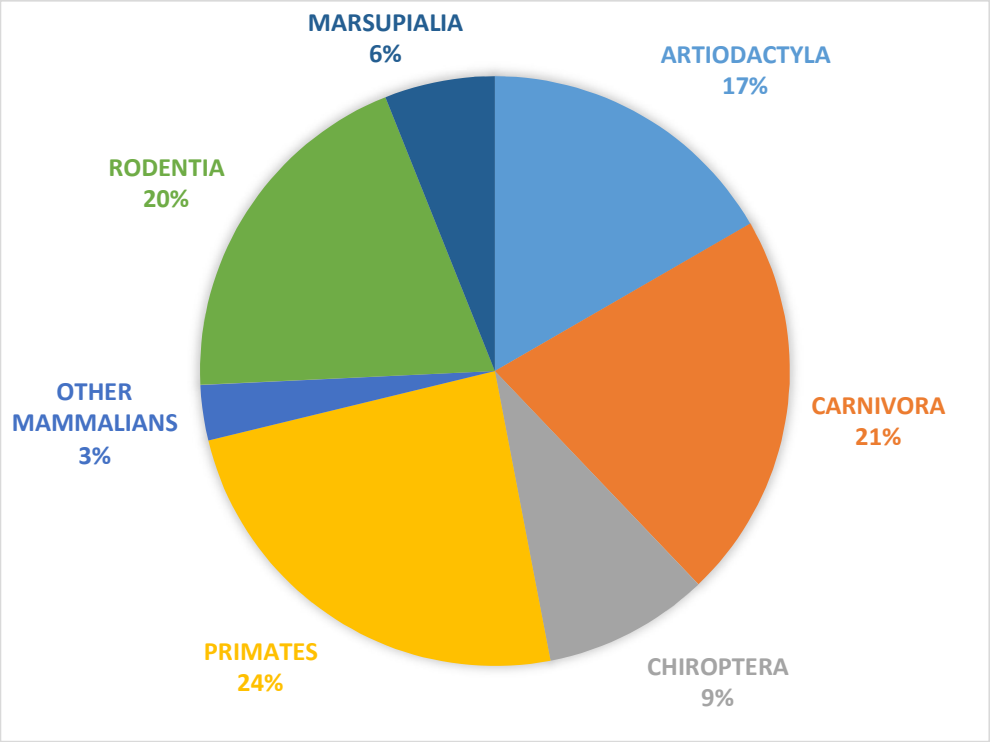
173	C	-0.809	9	66/66	C
174	A	-0.101	6	66/66	C,A,G
175	A	1.609	1	66/66	A,G,C
176	C	1.633	1	66/66	T,C,G,A
177	C	0.264	4	66/66	G,C,T
178	A	0.833	1	66/66	T,A,G
179	G	-0.372	7	66/66	G,T
180	T	0.613	2	66/66	C,T
181	G	-0.809	9	66/66	G
182	A	-0.41	7	66/66	A,T
183	G	0.7	1	66/66	A,G,C
184	A	0.458	2	66/66	T,A,G
185	A	-0.413	7	66/66	G,A
186	T	-0.809	9	66/66	T
187	T	-0.809	9	66/66	T
188	T	-0.809	9	66/66	T
189	C	0.309	3	66/66	C,T,A
190	T	-0.319	7	66/66	C,T
191	T	-0.809	9	66/66	T
192	C	-0.031	5	66/66	T,C
193	C	1.339	1	66/66	T,C,G,A
194	T	0.196	4	66/66	G,T
195	T	-0.326	7	66/66	T,C
196	C	3.388	1	66/66	T,C,A,G
197	A	1.021	1	66/66	G,A
198	G	0.262	4	66/66	G,A,C
199	A	0.36	3	66/66	A,G
200	C	0.308	3	66/66	A,C,T
201	G	0.939	1	66/66	C,T,G,A
202	G	0.645	1	66/66	G,A,C,T
203	T	0.745	1	66/66	C,T,G
204	T	-0.303	7	66/66	T,A
205	T	0.717	1	66/66	C,T,G
206	T	1.73	1	66/66	G,A,C,T
207	G	5.273	1	66/66	T,C,A,G
208	G	3.306	1	66/66	G,A,C,T
209	A	2.217	1	66/66	T,C,G,A
210	T	0.896	1	66/66	A,T,C
211	C	-0.809	9	66/66	C
212	T	0.32	3	66/66	A,T,C
213	T	2.596	1	66/66	T,G,A
214	A	1.441	1	66/66	T,C,A
215	C	-0.448	7	66/66	C,A
216	T	0.702	1	66/66	G,T,C
217	G	2.1	1	66/66	A,G,T
218	C	2.362	1	66/66	T,C,A,G
219	A	-0.46	8	66/66	G,A
220	C	-0.396	7	66/66	T,C

221	A	-0.809	9	66/66	A
222	G	-0.089	5	66/66	G,A
223	C	0.264	4	66/66	A,G,T,C
224	T	-0.357	7	66/66	G,T
225	T	0.645	1	66/66	G,T
226	T	1.725	1	66/66	G,T,C
227	C	0.355	3	66/66	A,C,T
228	T	-0.809	9	66/66	T
229	G	4.32	1	66/66	C,T,A,G
230	A	2.211	1	66/66	A,G,T,C
231	G	0.567	2	66/66	A,G
232	A	-0.005	5	66/66	T,A
233	A	-0.073	5	66/66	A,G,C
234	G	1.529	1	66/66	A,G,T
235	T	1.601	1	66/66	A,C,T
236	T	2.229	1	66/66	G,A,T,C
237	C	0.55	2	66/66	T,C
238	T	0.1	4	66/66	T,C
239	T	0.515	2	66/66	G,C,T
240	T	-0.361	7	66/66	C,T
241	T	0.164	4	66/66	T,C
242	G	0.087	5	66/66	G,A,T
243	G	1.035	1	66/66	A,G
244	T	-0.809	9	66/66	T
245	G	-0.809	9	66/66	G
246	C	-0.809	9	66/66	C
247	C	-0.809	9	66/66	C
248	A	-0.809	9	66/66	A
249	T	-0.809	9	66/66	T
250	G	-0.809	9	66/66	G
251	T	-0.809	9	66/66	T
252	T	-0.809	9	66/66	T
253	T	1.118	1	66/66	C,T
254	T	0.151	4	66/66	G,A,T
255	G	-0.397	7	66/66	A,G
256	T	-0.809	9	66/66	T
257	G	-0.809	9	66/66	G
258	G	-0.809	9	66/66	G
259	C	-0.809	9	66/66	C
260	T	-0.35	7	66/66	G,T
261	T	0.621	2	66/66	C,T
262	G	1.574	1	66/66	G,A
263	C	-0.119	6	66/66	C,A
264	A	-0.46	8	66/66	G,A
265	T	-0.809	9	66/66	T
266	C	-0.809	9	66/66	C
267	A	-0.459	8	66/66	T,A
268	A	-0.46	8	66/66	G,A

269	A	-0.452	8	66/66	A,C
270	A	-0.46	8	66/66	G,A
271	G	-0.398	7	66/66	G,A
272	A	-0.809	9	66/66	A
273	G	-0.397	7	66/66	A,G
274	G	-0.809	9	66/66	G
275	A	-0.809	9	66/66	A
276	G	0.012	5	66/66	G,A
277	T	-0.809	9	66/66	T
278	T	-0.809	9	66/66	T
279	T	-0.362	7	66/66	T,C
280	G	-0.809	9	66/66	G
281	T	-0.809	9	66/66	T
282	C	-0.809	9	66/66	C
283	T	-0.809	9	66/66	T
284	T	-0.809	9	66/66	T
285	C	-0.809	9	66/66	C
286	A	-0.809	9	66/66	A
287	T	-0.809	9	66/66	T
288	G	-0.809	9	66/66	G
289	A	-0.809	9	66/66	A
290	A	-0.809	9	66/66	A
291	G	-0.809	9	66/66	G
292	A	-0.809	9	66/66	A
293	T	-0.809	9	66/66	T
294	T	-0.809	9	66/66	T
295	C	-0.809	9	66/66	C
296	C	-0.809	9	66/66	C
297	T	-0.809	9	66/66	T
298	A	-0.809	9	66/66	A
299	A	-0.809	9	66/66	A
300	C	-0.472	8	66/66	C,T

Tab. 9.1.1 Raw Data of Consurf Analysis of Ex1_202up*. Different colors are used to underline different regions of the sequence analyzed. The upstream 10 nts of the Ex1_202up are coloredin yellow (1-10); the 25 nts upstream the TSS of Ex1_205 that belong to Ex1_202 are reportedin green (11-35); light grey is used to indicate the positions of Ex1_205 (36-123); dark pink are used to indicate the positions of the UTex1 (124-286); light pink is used to indicate the first positions belonged to the translated region of the exon 2, reported as reference of a conserved sequence. POS: The position of the Nucleic Acid on the sequence; SEQ: The Nucleic Acid; SCORE: The normalized conservation scores; COLOR: The color scale representing the conservation scores (9 - conserved, 1 - variable); MSA DATA: The number of alignedsequences having the nucleic acid (non-gapped) from the overall number of sequences at each position; RESIDUE VARIETY: The residues variety at each position of the multiple sequencealignment.*

A



B

Taxon		Species
Order	ARTIODACTYLA	Lagenorhynchus obliquidens
Order	ARTIODACTYLA	Monodon monoceros
Order	ARTIODACTYLA	Camelus dromedarius
Order	ARTIODACTYLA	Globicephala melas
Order	ARTIODACTYLA	Tursiops truncatus
Order	ARTIODACTYLA	Ovis aries
Order	ARTIODACTYLA	Capra hircus
Order	ARTIODACTYLA	Delphinapterus leucas
Order	ARTIODACTYLA	Bos indicus
Order	ARTIODACTYLA	Bos taurus
Order	ARTIODACTYLA	Bubalus bubalis
Order	CARNIVORA	Lynx canadiensis
Order	CARNIVORA	Acinonyx jubatus
Order	CARNIVORA	Canis lupus dingo
Order	CARNIVORA	Vulpes vulpes
Order	CARNIVORA	Ailuropoda melanoleuca
Order	CARNIVORA	Eumetopias jubatus
Order	CARNIVORA	Leptonychotes weddellii

Order	CARNIVORA	Panthera pardus
Order	CARNIVORA	Suricata suricatta
Order	CARNIVORA	Felis catus
Order	CARNIVORA	Enhydra lutris
Order	CARNIVORA	Zalophus californianus
Order	CARNIVORA	Canis lupus familiaris
Order	CARNIVORA	Ursus arctos horribilis
Order	CHIROPTERA	Pteropus vampyrus
Order	CHIROPTERA	Eptesicus fuscus
Order	CHIROPTERA	Pteropus alecto
Order	CHIROPTERA	Hipposideros armiger
Order	CHIROPTERA	Rousettus aegyptiacus
Order	CHIROPTERA	Myotis lucifugus
Order	CINGULATA	Dasypus novemcinctus
Order	PRIMATES	Homo sapiens
Order	PRIMATES	Macaca mulatta
Order	PRIMATES	Callithrix jacchus
Order	PRIMATES	Aotus nancymaae
Order	PRIMATES	Macaca nemestrina
Order	PRIMATES	Theropithecus gelada
Order	PRIMATES	Pan troglodytes
Order	PRIMATES	Pongo abelii
Order	PRIMATES	Nomascus leucogenys
Order	PRIMATES	Papio anubis
Order	PRIMATES	Otolemur garnettii
Order	PRIMATES	Cebus imitator
Order	PRIMATES	Microcebus murinus
Order	PRIMATES	Rhinopithecus roxellana
Order	PRIMATES	Piliocolobus tephrosceles
Order	PRIMATES	Gorilla gorilla
Order	PERISSODACTYLA	Equus caballus
Order	RODENTIA	Mus musculus
Order	RODENTIA	Ictidomys tridecemlineatus
Order	RODENTIA	Nannospalax galili
Order	RODENTIA	Mastomys coucha
Order	RODENTIA	Cricetulus griseus
Order	RODENTIA	Heterocephalus glaber
Order	RODENTIA	Mus pahari
Order	RODENTIA	Peromyscus leucopus
Order	RODENTIA	Rattus norvegicus
Order	RODENTIA	Marmota flaviventris
Order	RODENTIA	Mus caroli
Order	RODENTIA	Cavia porcellus
Order	RODENTIA	Mesocricetus auratus
Infraclass	MARSUPIALIA	Monodelphis domestica
Infraclass	MARSUPIALIA	Sarcophilus harrisii
Infraclass	MARSUPIALIA	Vombatus ursinus
Infraclass	MARSUPIALIA	Trichosurus vulpecula

Tab. 9.1.2 The Sequences used for the alignment. *(A) The 66 sequences used for Consurf analysis of Ex1_202up belong to placentals. 6% of sequences is from the Infraclass of Marsupialia, while the remaining 94% is from the class of Mammalia (Orders: Artiodactyla 17%; Carnivora 14%; Chiroptera 9%; Primates 24%; Rodentia 20%; Other mammalians (Cingulata and Perissodactyla) 3%). (B) List of the species from which are taken the orthologous sequences reporting taxon details and scientific name.*

9.2 Predicted disruptive SNPs from RNAsnp

RNAsnp was used to assess the hypothetic disruptive effect of the Evans's SNP. However, we predicted also the putative most disruptive SNPs affecting the structure of CDKL5 205 5'UTR of interest. Here, we report this analysis that can be useful for future experiments. Interestingly, most of the predicted disruptive SNPs lie in the Ex1_205. This RNAsnp analysis was performed under the mode of operation "3", predicting the most impacting SNPs using both global (Mode 1) and local (mode 2) folding methods.

	Mode 2		Mode 1		
SNP	**Interval**	**p-value1**	**Folding Window**	**Local region**	**p-value2**
U3A	1-42	0.0946	1-203	1-50	0.0332
G11C	1-39	0.0895	1-211	2-54	0.0365
C12A	1-52	0.0068	1-212	1-50	0.0257
C12U	1-39	0.0100	1-212	1-50	0.0384
G14C	1-52	0.0197	1-214	1-50	0.0330
G14U	1-52	0.0361	1-214	1-50	0.0127
G29C	1-39	0.0744	1-229	2-51	0.0317
C30A	1-39	0.0402	1-230	1-50	0.0124
C30G	1-44	0.0712	1-230	1-50	0.0145
C33A	1-36	0.0473	1-233	1-50	0.0396
G58C	52-93	0.0004	1-250	56-105	0.0045
G76A	52-91	0.0013	1-250	56-105	0.0219
G76C	52-91	0.0010	1-250	56-105	0.0182
G76U	52-91	0.0008	1-250	56-105	0.0154
G77A	52-91	0.0008	1-250	56-105	0.0176
G77C	52-91	0.0005	1-250	56-105	0.0134
G77U	52-91	0.0005	1-250	56-105	0.0120
G78A	52-91	0.0015	1-250	56-105	0.0212
G78C	53-91	0.0001	1-250	56-105	0.0056
G78U	52-91	0.0038	1-250	56-105	0.0449
A115C	85-135	0.0802	1-250	86-135	0.0461
A147G	129-152	0.0563	1-250	128-182	0.0222
A229C	196-239	0.0167	29-250	200-249	0.0429
A229U	196-242	0.0278	29-250	140-237	0.0219
A232C	196-240	0.0216	32-250	140-237	0.0057
A232U	196-238	0.0427	32-250	140-237	0.0207

Tab. 9.2 Predicted disruptive SNPs obtained from RNAsnp. *Mode 2 Interval: length and positions of the local structure that is presumably disrupted by the predicted SNP; p-value1: P-value threshold to filter SNPs that are predicted using Mode 2 (threshold: 0.2); Mode 1*

Interval: window of the predicted disrupted global folding of the structure interested by the predicted SNP. Local region: Minimum length of flanking regions of the SNP interested in the disruption of the structure, predicted based on global folding. P-value2: P-value threshold to filter SNPs that are predicted using Mode 1 (threshold: 0.2).

9.3 CAGEr Workflow

Here we reported the workflow of the CAGEr peak analysis used to obtain the results presented in **par. 5.7**.

````
title: "Human CAGE"
output: html_document
---
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE)
```
```{r}
library(CAGEr)
library(TxDb.Hsapiens.UCSC.hg19.knownGene)
library(org.Hs.eg.db)
```
# retrieve FANTOM5 samples annotation
```{r}
data(FANTOM5humanSamples)
```
# download files
```{r}
body_tissue_samples <- c(
'liver_adult__pool1',
'kidney_adult_pool1',
'heart_adult__pool1',
'testis__adult__pool1',
'lung__adult__pool1',
'brain__adult__pool1',
'brain_fetal_pool1'
)
```
```{r}
FANTOM5humanSamples[FANTOM5humanSamples$sample %in% body_tissue_samples,1:4]
```
```{r, eval=FALSE}
for (sample_name in body_tissue_samples) {
  url <- subset(FANTOM5humanSamples, sample == sample_name)$data_url
download.file(url, paste0('bed_in/',gsub('_-_','_',gsub('-','_',gsub('%252[0c]', '_',
basename(url))))))
}
# then gunzip files from file system
```
# start from local files
```{r}
inputFiles <- list.files(pattern = 'ctss.bed$', path = 'bed_in', full.names = TRUE)
````
````

```
sampleNames <- sapply(strsplit(basename(inputFiles), split = '\\.'), '[[', 1)
sampleNames <- sub('-','',sampleNames)
ce <- CAGEexp( genomeName     = "BSgenome.Hsapiens.UCSC.hg19"
           , inputFiles     = inputFiles
           , inputFilesType = "bedScore"
           , sampleLabels  = sampleNames
)
getCTSS(ce)
CTSStagCountSE(ce)
```
Genome annotations
```{r}
hg19_exons <- exonsBy(TxDb.Hsapiens.UCSC.hg19.knownGene, by = 'gene')
hg19_transcripts <- transcriptsBy(TxDb.Hsapiens.UCSC.hg19.knownGene, by = 'gene')
hg19_genes <- reduce(hg19_transcripts)
#
genesymbols <- mapIds(org.Hs.eg.db, names(hg19_exons), 'SYMBOL', 'ENTREZID')
ix <- !is.na(genesymbols)
hg19_exons <- hg19_exons[ix]
names(hg19_exons) <- genesymbols[ix]
hg19_exons <- unlist(hg19_exons)
hg19_exons$exon_id <- NULL
hg19_exons$exon_name <- NULL
hg19_exons$gene_name <- names(hg19_exons)
hg19_exons$type <- 'exon'
# transcripts
genesymbols <- mapIds(org.Hs.eg.db, names(hg19_transcripts), 'SYMBOL', 'ENTREZID')
ix <- !is.na(genesymbols)
hg19_transcripts <- hg19_transcripts[ix]
names(hg19_transcripts) <- genesymbols[ix]
hg19_transcripts <- unlist(hg19_transcripts)
hg19_transcripts$tx_id <- NULL
hg19_transcripts$tx_name <- NULL
hg19_transcripts$gene_name <- names(hg19_transcripts)
hg19_transcripts$type <- 'transcript'
# genes
hg19_genes <- unlist(hg19_genes[elementNROWS(hg19_genes) == 1])
genesymbols <- mapIds(org.Hs.eg.db, names(hg19_genes), 'SYMBOL', 'ENTREZID')
ix <- !is.na(genesymbols)
hg19_genes <- hg19_genes[ix]
names(hg19_genes) <- genesymbols[ix]
hg19_genes$gene_name <- names(hg19_genes)
hg19_genes$type <- 'gene'
#
names(hg19_genes) <- NULL
names(hg19_transcripts) <- NULL
names(hg19_exons) <- NULL
hg19_ann <- c(hg19_genes, hg19_transcripts, hg19_exons)
```
```{r}
annotateCTSS(ce, hg19_ann)
```
```{r}
plotAnnot(ce, "counts")
```
```{r}
corr.m <- plotCorrelation2( ce, samples = "all"
```

```
                    , tagCountThreshold = 1, applyThresholdBoth = FALSE
                    , method = "pearson")
```
```{r}
librarySizes(ce)
```
```{r, fig.height=8, fig.width=10}
plotReverseCumulatives(ce, fitInRange = c(1, 1e+06), onePlot = TRUE,
                 xlim = c(1, 1e+08), ylim = c(1, 1e+07))
```
```{r}
normalizeTagCount(ce, method = "powerLaw", fitInRange = c(1, 1e+06),
              alpha = 1.2, T = 1*10^7)
ce[["tagCountMatrix"]]
```
```{r}
clusterCTSS( object = ce
        , threshold = 1
        , thresholdIsTpm = TRUE
        , nrPassThreshold = 1
        , method = "distclu"
        , maxDist = 20
        , removeSingletons = TRUE
        , keepSingletonsAbove = 5)
```
```{r}
cumulativeCTSSdistribution(ce, clusters = "tagClusters",
                     useMulticore = T, nrCores = 8)
```
```{r}
quantilePositions(ce, clusters = "tagClusters", qLow = 0.1, qUp = 0.9,
              useMulticore = TRUE, nrCores = 8)
```
```{r}
aggregateTagClusters(ce, tpmThreshold = 5, qLow = 0.1, qUp = 0.9, maxDist = 100,
                useMulticore = TRUE, nrCores = 8)
```
```{r}
annotateConsensusClusters(ce, hg19_ann)
cumulativeCTSSdistribution(ce, clusters = "consensusClusters",
                     useMulticore = TRUE, nrCores = 8)
quantilePositions(ce, clusters = "consensusClusters", qLow = 0.1, qUp = 0.9,
              useMulticore = TRUE, nrCores = 8)
```
```{r}
saveRDS(ce, 'ceDataset.rds')
```
```{r}
exportCTSStoBedGraph(ce, values = "normalized", format = "BigWig")
exportToBed(object = ce, what = "tagClusters", qLow = 0.1, qUp = 0.9, oneFile = TRUE)
exportToBed(object = ce, what = "tagClusters", qLow = 0.1, qUp = 0.9, oneFile = FALSE)
```

9.4 Luciferase Reporter Vectors

We reported a more detailed insight on the two vector systems used in **par. 5.8**: pBRm2L and pBATmod2. The cloning strategy of pBRm2L luciferase reporter vector (7180 bp) is reported in the paper De Pietri Tonelli et al., 2004. It is a two-promoter vector that allows the simultaneous translation of the Renilla (124-1059 nts) and Firefly (1229-2881 nts) luciferases genes, each one under the control of its own T7 promoter (1-19 nts and 1183-1201 nts). Renilla luciferase is translated through cap-dependent initiation, having an optimized, short 5'UTR (that contains also a Ribosome Binding Site for bacterial expression; RBS, 56-78 nts). On the other hand, Firefly luciferase is under the translational control of an optimized short 5'UTR sequence, that is 27 nts long and that contains the polylinker region in which the restriction site of SalI and NcoI are placed. At the end of the Fluc coding sequence there is the unique T7 terminator (3012-3059 nts). Other elements present in pBR2mL are lac operator for bacterial expression (19-43 nts), Ori sequence (6351-6939 nts), Amp promoter (5215-5319 nts) and Amp gene (5320-6180 nts). The vector map is reported in **Fig. 9.4 A**.

pBATmod2 luciferase reporter vector (7641 bp) was cloned as described in De Pietri Tonelli et al., 2003. It is a dicistronic vector allowing the simultaneous translation of the Renilla and the Firefly luciferase genes. Renilla luciferase cistron (124-1059 nts) is translated through cap-dependent initiation, having an optimized, short 5'UTR (as described before) after the T7 promoter (1-19 nts). Firefly luciferase cistron (1690-3342 nts) is placed under the sequence of the EMCV IRES (1103-1689 nts), downstream to the end of the sequence of Renilla luciferase. Thus, Firefly luciferase is translated trough cap-independent initiation. At the end of the Fluc coding sequence there is the T7 terminator (3473-3520 nts). Other elements present in pBATmod2 are lac operator for bacterial expression (19-43 nts), Ori sequence (6812-7400 nts), Amp promoter (5676-5780 nts) and Amp gene (5781-6641 nts). The vector map is reported in **Fig. 9.4 B**.

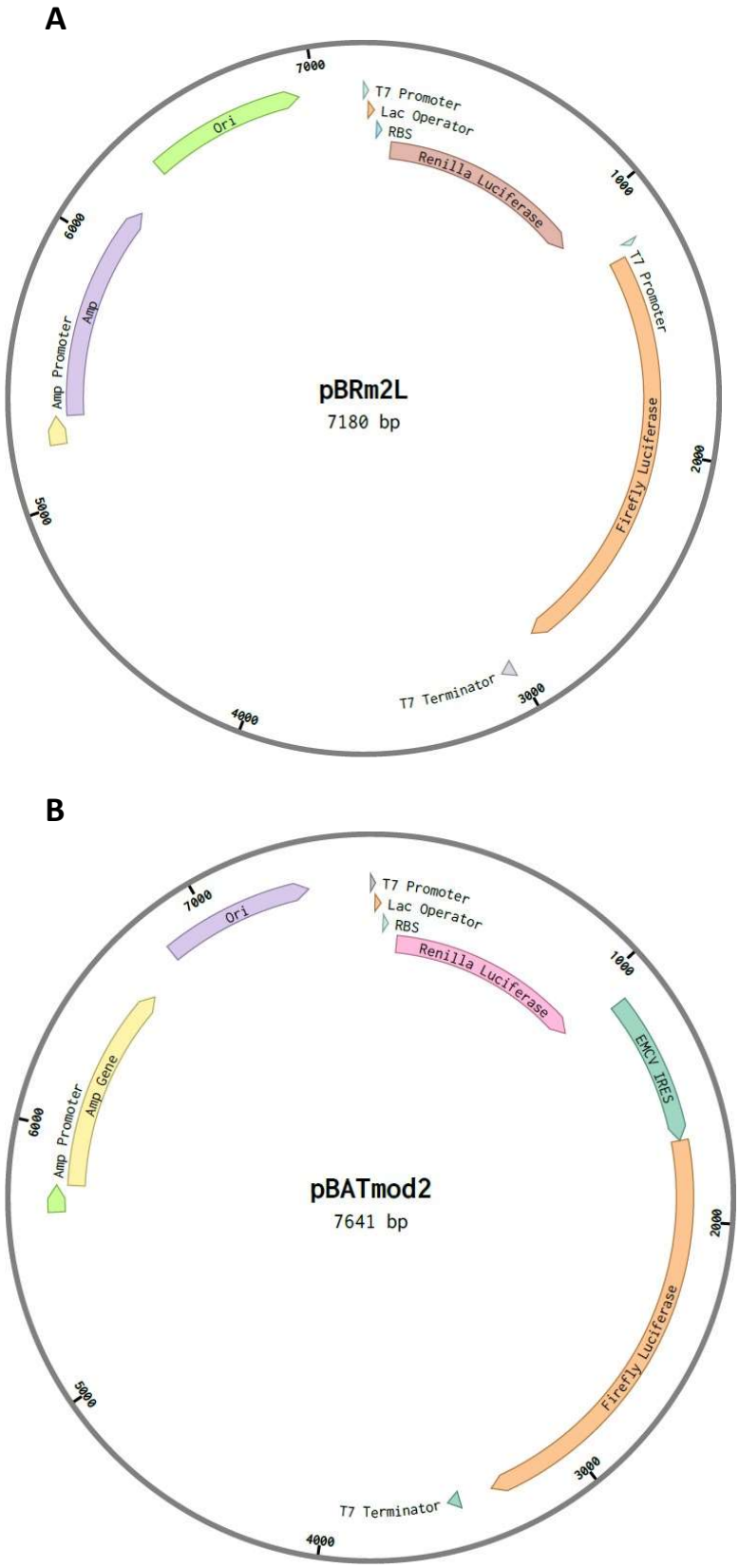


Fig. 9.4 Luciferase Reporter Vectors Maps. *(A) Map of the vector pBRm2L (De Pietri Tonelli et al., 2004). (B) pBATmod2 vector (De Pietri Tonelli et al., 2003)*