

**Analysis of the Ccr4-Not  
deadenylase complex:  
a biochemical  
and computational approach**

**Dario Leonardo Balacco, MSc**

Thesis submitted to the University of Nottingham for  
the degree of  
Doctor of Philosophy

**October 2016**

To Biagio

*"Rare sono le persone che usano la mente,  
poche coloro che usano il cuore  
e uniche coloro che usano entrambi"*

(Rita Levi Montalcini)

## **ABSTRACT**

In eukaryotic cells, the degradation of the mRNA poly(A) tail (deadenylation) is a crucial step in the regulation of gene expression. The Ccr4-Not complex is the major deadenylase enzyme involved in the mRNA deadenylation. The complex is composed by two subunits with ribonuclease activity (Caf1 and Ccr4) and at least six non-catalytic subunits. In vertebrate cells, the duplication of the catalytic subunits Caf1 (encoded by CNOT7 and CNOT8) and Ccr4 (encoded by CNOT6 and CNOT6L) lead to the heterogeneity of the complex. The non-catalytic subunits are organised in modules, each with a specific function, allowing the recruitment of the complex on specific mRNAs. Regulatory proteins interact with the deadenylase complex tethering it on specific mRNAs and activating mRNA degradation pathways and down-regulating protein expression. This study discovered and explored the interactome of the catalytic subunits Caf1 and Ccr4 and detected new interacting partners that may recruit the deadenylase complex on specific mRNAs. Nowadays, genotyping patients using whole genome and next generation sequencing technologies, allows a wider but more accurate sight of the genomic context of a specific disease. The alteration of the function or the structure of the Ccr4-Not complex in cancer was assessed investigating the structure and function of the variants of the components of the NOT-module (CNOT1, CNOT2 and CNOT3) and the nuclease sub-complex (CNOT7, CNOT6L, CNOT1, and the regulatory protein BTG1) found in various types of cancers. Finally, phylogenetic analysis of 15 mammalian species identified different evolutionary rates acting on the paralogous deadenylase subunits.

## **ACKNOWLEDGMENTS**

I would first like to thank my supervisor, Dr Sebastiaan Winkler, for guiding me throughout my PhD, with patience, support, and trust.

My sincere gratitude goes to the University of Nottingham for awarding me the Vice-Chancellor's Scholarship for Research Excellence (European Union).

I would also like to thank Dr. Klaas Mulder, currently based at the Faculty of Science, Radboud Institute for Molecular Life Science, Radboud University, Netherlands, for providing me the SILAC-RIME data.

I must thank Hibah Almasmoum, "Habbush", for providing me the pBD-GAL4-BTG1 plasmid, but more importantly for providing me joy and hard times every single day of my PhD. I would also thank Lorenzo Pavanello for helping me with the cloning of the pCI $\lambda$ N-CNOT7 variants, for being my Italian bench-mate, and for always encouraging me.

I would like to thank the other members of the Winkler group: Maryati, Vanessa, Blessing and Ben.

A huge thank you to Professor David Heery and all the members of the Gene Regulation and RNA Biology Group: Hilary, Johnny, Bismoy, Anna and Barbara. A special thanks to my special Brazilian friends, the Monteiro sisters, Cintia and Patricia, for being my happy place in this years, for keeping me company during our ungodly hours, and for our great times in Ibiza with Corona. I would also like to thank Francesca, precious friends and the best housemates I could ever ask for.

A particular thank to Dr Anna Pacelli, a lifetime-long friend, for sharing the same dream as me, for supporting me in every choice of my life and for being an inspiration. An immense thank you to Leo, who stays next to me, supports me and takes care of me on a daily basis.

I must thank dad, my family, and friends for always believing in me.

I also thank my sister Tiziana for being the closest person in my life.

The biggest thanks goes to mum for believing in me since I was a gamete, for feeding me with knowledge and inspiring me to become the person I've chosen to be.

## List of content

<b>Abstract</b> .....	<b>III</b>
<b>Acknowledgments</b> .....	<b>IV</b>
<b>List of content</b> .....	<b>V</b>
<b>List of Figures</b> .....	<b>XI</b>
<b>List of Tables</b> .....	<b>XVI</b>
<b>List of Supporting Materials</b> .....	<b>XVIII</b>
<b>List of abbreviation</b> .....	<b>XX</b>
<b>Chapter 1. Introduction</b> .....	<b>2</b>
1.1 Eukaryotic mRNA synthesis, processing and function .....	2
1.1.1 Eukaryotic transcription.....	2
1.1.2 mRNA processing.....	6
1.1.3 Polyadenylation .....	7
1.1.4 mRNA nuclear export.....	9
1.2 Regulated mRNA decay .....	10
1.2.1 Deadenylase enzymes .....	14
1.2.2 Decapping complex and its regulation.....	16
1.2.3 Protein-mediated mRNA decay .....	17
1.2.4 microRNA mediated decay .....	20
1.3 Quality control of the mRNA .....	22
1.3.1 Nonsense-mediated decay (NMD).....	22
1.3.2 Nonstop mRNA decay (NSD) and No-go decay (NGD).....	23
1.4 The Pan2-Pan3 complex .....	25
1.4.1 Overview of the Pan2-Pan3 complex .....	25
1.4.2 Pan2 .....	26
1.4.3 Pan3 .....	26
1.5 The Ccr4-Not complex overview .....	29
1.5.1 The Ccr4 subunit .....	29
1.5.2 The Caf1 subunit.....	31
1.5.3 The redundancy of the catalytic subunits .....	32
1.5.4 The non-catalytic subunits.....	35
1.5.5 The additional roles of the Ccr4-Not complex.....	36

1.6 The BTG/TOB proteins.....	40
1.6.1 The role of the BTG/TOB proteins.....	41
1.7 Aims of the study.....	45
1.8 Experimental approaches.....	46
1.8.1 Stable Isotope Labelling using Amino acids in Cell culture (SILAC)- Rapid Immunoprecipitations and Mass spectrometry of Endogenous protein (RIME).....	46
1.8.2 Identification of residues critical for the NOT-module assembly .....	46
1.8.3 Analyses of sequence variants of the CNOT7 ribonuclease .....	47
1.8.4 Phylogenetic and variability analysis.....	47
<b>Chapter 2. Materials and Methods .....</b>	<b>49</b>
2.1 Bacterial growth and transformation .....	50
2.1.1 Reagents, stock solutions and buffers.....	50
2.1.2 Culture of <i>Escherichia coli</i> .....	51
2.1.3 Preparation of <i>E. coli</i> competent cells.....	51
2.1.4 Transformation of competent cells .....	51
2.2 Molecular Biology .....	53
2.2.1 Reagents, stocks solutions and buffers.....	53
2.2.2 Small scale plasmid preparation .....	53
2.2.3 Large scale plasmid preparation.....	53
2.2.4 Determination of DNA/RNA concentration.....	54
2.2.5 Agarose gel electrophoresis .....	54
2.2.6 DNA extraction and purification from agarose gels.....	54
2.2.7 Restriction enzyme digestion of DNA .....	55
2.2.8 Removal of 5' phosphate group from linearised plasmid DNA.....	55
2.2.9 Phosphorylation of blunt PCR products.....	56
2.2.10 Ligation of DNA fragments .....	56
2.2.11 Gibson cloning .....	56
2.2.12 Polymerase Chain Reaction (PCR) .....	57
2.2.13 Site-directed mutagenesis .....	58
2.2.14 List of Primer.....	59
2.3 Protein purification and analysis .....	62
2.3.1 Reagents, stock solutions and buffers.....	62
2.3.2 Expression culture of <i>E.coli</i> .....	62
2.3.3 Protein extraction from expression culture of <i>E. Coli</i> .....	63

2.3.4 Protein purification.....	63
2.3.5 Bradford assay to determine protein concentration.....	63
2.3.6 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS- PAGE).....	64
2.4 Yeast two-hybrid analysis .....	65
2.4.1 Reagents, stock solutions and buffers.....	66
2.4.2 Culture of <i>Saccharomyces cerevisiae</i> strain YRG2.....	66
2.4.3 Transformation of YRG2 cells.....	66
2.4.4 Yeast two-hybrid $\beta$ -galactosidase assay .....	67
2.4.5 Western blot analysis of yeast cell lysates .....	68
2.5 Cell culture .....	65
2.5.1 Reagents, stock solutions and buffers for use in tissue culture.....	65
2.5.2 Routine maintenance of cell lines .....	65
2.5.3 Freezing cells for long term storage .....	70
2.5.4 Retrieving cells from liquid nitrogen storage.....	70
2.6 Immunoprecipitation analyses .....	71
2.6.1 Reagents, stock solutions and buffers for use in immunoprecipitations .....	71
2.6.2 Preparation of protein lysates for immunoprecipitation.....	72
2.6.3 Preparation of antibody-agarose beads for immunoprecipitation .....	72
2.7 siRNA knockdown.....	73
2.7.1 Reagents, stock solutions and buffers for use in protein analysis.....	73
2.7.2 siRNA transfection.....	74
2.8 RNA tethering assay.....	75
2.8.1 Reagents, stock solutions and buffers for use in protein analysis.....	75
2.8.2 Cell seeding and transfection for RNA tethering assay .....	75
2.8.3 Luciferase reporter assay .....	76
2.8.4 Cellular lysates for western blot.....	76
2.9 Protein analysis.....	78
2.9.1 Reagents, stock solutions and buffers for use in protein analysis.....	78
2.9.2 Bradford assay to determine protein concentration.....	79
2.9.3 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS- PAGE).....	79
2.9.4 Western blot analysis .....	80
2.10 Stable isotope labelling in cell culture (SILAC)-rapid immunoprecipitations mass spectrometry of endogenous protein (RIME).....	82

2.10.1 Reagents, stock solutions and buffers for use in SILAC-RIME analysis .....	82
2.10.2 Preparation of magnetic beads.....	83
2.10.3 Preparation of cell lysates .....	83
2.10.4 Immunoprecipitation .....	84
2.11 Bioinformatics .....	85
2.11.1 Protein-protein interaction network generation .....	85
2.11.2 Expansion of intra-network protein-protein interactions .....	85
2.11.3 Clustering analysis .....	86
2.11.4 Generation of Residues interaction networks (RIN) .....	86
2.11.5 Analysis of protein variants using UCSF Chimera .....	86
2.12 Phylogenetic analysis .....	88
2.12.1 Retrieval of homologous proteins.....	88
2.12.2 Alignment of retrieved proteins .....	88
2.12.3 Phylogenetic reconstruction with Neighbour-joining method .....	89
2.12.4 Selection of the best substitution model .....	89
2.12.5 Phylogenetic reconstruction by Maximum-Likelihood method .....	89
2.12.6 Phylogenetic analysis by Maximum-Likelihood (PAML) .....	91
2.13 Statistical data analysis.....	93

## **Chapter 3. The interactome of the nuclease module of the Ccr4-Not complex ..... 94**

3.1 Introduction .....	95
3.2 Stable Isotope Labelling with Amino acids in Cell culture-Rapid Immunoprecipitation and Mass spectrometry of Endogenous proteins (SILAC-RIME) of the Ccr4-Not nuclease module .....	97
3.3 Visualisation of SILAC-RIME data .....	101
3.4 Combined protein-protein interaction network of the nuclease module .	106
3.4.1 Clustering analyses.....	110
3.4.1.1 Molecular complexes identified by MCODE.....	113
3.4.1.2 Molecular complexes identified by ClusterOne.....	117
3.5 Evaluation of Pan2 and Pan3 antibodies .....	124
3.6 Discussion .....	128
3.6.1 The interactome of the Ccr4-Not deadenylase module .....	128

3.6.2 Pan2 and Pan3.....	130
--------------------------	-----

## **Chapter 4. Analysis of the NOT module of the Ccr4-Not**

### **complex ..... 131**

4.1 Introduction .....	132
4.2 Collection of mutations and prediction of their impact on the function of the NOT-module.....	134
4.3 Structure analysis .....	140
4.3.1 Proteins, Interfaces, Structures and Assemblies (PISA) .....	140
4.3.2 RINalyzer .....	147
4.4 Protein-protein interaction analysis .....	149
4.4.1 Yeast two hybrid assay for CNOT1-CNOT2 interactions .....	149
4.4.2 Co-Immunoprecipitation for CNOT1-CNOT3 interaction.....	153
4.5 Discussion .....	156

## **Chapter 5. Analysis of sequence variants of the CNOT7**

### **ribonuclease..... 160**

5.1 Introduction .....	161
5.2 Collection of mutations and prediction of their impact on the function of CNOT7.....	167
5.3 Prediction of the impact of the mutations on the structure of CNOT7.....	172
5.4 The effect of CNOT7 mutations on protein-protein interactions in the nuclease sub-complex.....	178
5.4.1 Generation of plasmids for yeast two hybrid assays.....	178
5.4.2 Identification of CNOT7 residues required for the interaction with the MIF4G domain of CNOT1 .....	179
5.4.3 Identification of CNOT7 residues required for the interaction with CNOT6L.....	185
5.4.4 Identification of CNOT7 residues required for the interaction with BTG1 .....	190
5.4.5 Identification of residues required for the activity of CNOT7 .....	195
5.5 Discussion .....	202

## **Chapter 6. Phylogenetic and variability analysis of the Ccr4-**

### **Not complex..... 208**

6.1 Introduction .....	209
------------------------	-----

6.2 Phylogenetic and variability analysis of CNOT1.....	212
6.2.1 Phylogenetic reconstruction of CNOT1 .....	212
6.2.2 Variability analysis of the CNOT1 .....	213
6.3 Phylogenetic and variability analysis of CNOT2.....	218
6.3.1 Phylogenetic reconstruction of CNOT2 .....	218
6.3.2 Variability analysis of the CNOT2 component.....	218
6.4 Phylogenetic and variability analysis of CNOT3.....	223
6.4.1 Phylogenetic reconstruction of CNOT3 .....	223
6.4.2 Variability analysis of the CNOT3 .....	223
6.5 Phylogenetic and variability analysis of the Ccr4 subunits .....	228
6.5.1 Phylogenetic reconstruction of the Ccr4 subunits .....	228
6.5.2 Variability analysis of the Ccr4 homologues .....	228
6.6 Phylogenetic and variability analysis of the Caf1 subunit.....	235
6.6.1 Phylogenetic reconstruction of the Caf1 subunit.....	235
6.6.2 Variability analysis of the Caf1 homologues .....	236
6.7 Phylogenetic and variability analysis of CNOT9, CNOT10 and CNOT11....	241
6.7.1 Phylogenetic reconstruction of CNOT9, CNOT10 and CNOT11 .....	241
6.7.2 Variability analysis of the CNOT9, CNOT10 , and CNOT11 subunits.	242
6.7.3 Discussion .....	247

## **Chapter 7. Concluding Remarks and Future Outlook..... 252**

7.1 The interactome of the Ccr4 and Caf1 subunit revealed by SILAC-RIME..	253
7.2 Bioinformatic approaches to predict residues critical for the assembly and function of the NOT-module .....	255
7.3 <i>In silico</i> and <i>in vitro</i> approaches identified CNOT7 variants unable to interact with CNOT6L, CNOT1, and BTG1.....	256
7.4 Variability analyses identified different evolutionary rate on the orthologous protein of Ccr4 and Caf1.....	257
7.5 Future Outlook .....	259

## **BIBLIOGRAFY.....260**

## List of Figures

### Chapter 1

Figure 1.1 Overview of the eukaryotic gene regulation.....	3
Figure 1.2 Schematic representation of the pre-initiation complex (PIC) assembled on the RNAPII core promoter.....	5
Figure 1.3 Mammalian pre-mRNA 3' end processing machinery. (A) The nuclear cleavage and polyadenylation complex.....	8
Figure 1.4 Cytoplasmic mRNA degradation pathways.....	13
Figure 1.5 Protein-mediated recruitment of the Ccr4-Not complex. ....	19
Figure 1.6 Recruitment of the human deadenylase complexes by miRNA machinery.....	21
Figure 1.7 Nonsense mediated decay (NMD) and No-stop decay (NSD)....	24
Figure 1.8 Overview of Pan2-Pan3 deadenylase complex. ....	28
Figure 1.9 Architecture of the Ccr4-Not complex.....	30
Figure 1.10 The Caf1 subunit.....	33
Figure 1.11 The Caf1 nuclease component.....	34
Figure 1.12 Crystal structures of the NOT-module and the nuclease module.....	38
Figure 1.13 Models of the BTG/TOB protein family member. ....	43

### Chapter 2

Figure 2.1 Phylogenetic Analysis by Maximum-Likelihood (PAML) control file.....	86
--	----

### Chapter 3

Figure 3.1 Stable isotope labelling with amino acids in cell culture (SILAC) .....	99
Figure 3.2 SILAC-RIME approach to obtain Ccr4 and Caf1 interactome. .	100

Figure 3.3 Venn diagrams indicating overlap between SILAC-RIME data sets.....	102
Figure 3.4 SILAC-RIME results visualised by Cytoscape. ....	103
Figure 3.5 Correlation between CNOT6/CNOT6L co-immunoprecipitated proteins from the nuclear, cytoplasmic and combined fractions.. ....	104
Figure 3.6 Correlation between CNOT7/CNOT8 co-immunoprecipitated proteins from the nuclear, cytoplasmic and combined fractions. ....	105
Figure 3.7 Protein-protein interactions network.....	108
Figure 3.8 Nuclease module interactome networks.....	109
Figure 3.9 Clustering analyses of the Caf1-Ccr4 interactome. ....	111
Figure 3.10 Ccr4-NOT complex interactome. ....	112
Figure 3.11 Decapping complex identified by MCODE.....	114
Figure 3.12 Components of the LSm complex identified by MCODE. ....	115
Figure 3.13 Mini-chromosome maintenance complex identified by MCODE. ....	116
Figure 3.14 The Ccr4-Not complex identified using ClusterONE.....	119
Figure 3.15 LSm complex subunits and small nuclear ribonucleoproteins (SNRP) identified by ClusterONE. ....	120
Figure 3.16 Polypyrimidine binding proteins identified by ClusterONE. ....	121
Figure 3.17 Coatomer protein complex subunits identified by ClusterONE.. ....	122
Figure 3.18 Splicing factors identified by ClusterONE. ....	123
Figure 3.19 Pan2 and Pan3 antibodies test.....	125
Figure 3.20 Transfection efficiency of Interferin tested with AllStars Cell Death siRNA.....	126
Figure 3.21 Western blot of cellular lysates of MCF7 cells transfected with Pan2 and Pan3 siRNAs. Pan3 antibody.. ....	127

## Chapter 4

Figure 4.1 Impact of the NOT module mutations impact on the proteins function. ....	135
---	-----

Figure 4.2 Tissue localisation of mutations found in the NOT module. ....	137
Figure 4.3 Bioinformatic approaches to identify amino acid residues critical for the NOT module. ....	142
Figure 4.4 Residue Interaction Network of the human NOT module crystal structure. ....	148
Figure 4.5 Generation of pBD-GAL4-HA-CNOT1 and PAD-GAL4-2.1- CNOT2. ....	151
Figure 4.6 CNOT1 and CNOT2 yeast two hybrid assay. <i>S. cerevisiae</i> YRG2 was transformed with pBD-GAL4-HA-CNOT1 and pAD-GAL4-2.1- CNOT2. ....	152
Figure 4.7 Generation of pCMV5-GFP-CNOT1 and pCMV5-FLAG-CNOT3 plasmids. ....	154
Figure 4.8 CNOT1 and CNOT3 co-immunoprecipitation. GFP-CNOT1 and FLAG-CNOT3 were expressed in HEK293 cells. ....	155

## Chapter 5

Figure 5.1 Overview of the approaches used for the analysis of sequence variants of CNOT7. ....	164
Figure 5.2 Detection of CNOT7 interactions using yeast two hybrid analysis. ....	165
Figure 5.3 RNAtethering assay for functional studies of CNOT7 variants. .....	166
Figure 5.4 Overview of CNOT7 mutations annotated in COSMIC and Ensembl. ....	169
Figure 5.5 Map of selected CNOT7 variants. ....	171
Figure 5.6 Clashes caused by amino acid substitutions in CNOT7. ....	177
Figure 5.7 Generation of plasmids for yeast two hybrid analysis. ....	180
Figure 5.8 Yeast two-hybrid analysis of the CNOT1 <sub>MIF4G</sub> interaction with CNOT7 variants (Group I). ....	181
Figure 5.9 Yeast two-hybrid analysis of the CNOT1 <sub>MIF4G</sub> interaction with CNOT7 variants (Group II). ....	182

Figure 5.10 Yeast two-hybrid analysis of the CNOT1 <sub>MIF4G</sub> interaction with CNOT7 variants (Group III). .....	183
Figure 5.11 Yeast two-hybrid analysis of the CNOT1 <sub>MIF4G</sub> interaction with CNOT7 variants (Group IV). .....	184
Figure 5.12 Yeast two-hybrid analysis of the CNOT6L interaction with CNOT7 variants (Group I). .....	186
Figure 5.13 Yeast two-hybrid analysis of the CNOT6L interaction with CNOT7 variants (Group II). .....	187
Figure 5.14 Yeast two-hybrid analysis of the CNOT6L interaction with CNOT7 variants (Group III). .....	188
Figure 5.15 Yeast two-hybrid analysis of the CNOT6L interaction with CNOT7 variants (Group IV). .....	189
Figure 5.16 Yeast two-hybrid analysis of the BTG1 interaction with CNOT7 variants (Group I). .....	191
Figure 5.17 Yeast two-hybrid analysis of the BTG1 interaction with CNOT7 variants (Group II). .....	192
Figure 5.18 Yeast two-hybrid analysis of the BTG1 interaction with CNOT7 variants (Group III). .....	193
Figure 5.19 Yeast two-hybrid analysis of the BTG1 interaction with CNOT7 variants (Group IV). .....	194
Figure 5.20 Generation of the $\lambda$ N-CNOT7 expression plasmids for RNA tethering assays. ....	197
Figure 5.21 Group I CNOT7 variants are functional in RNA degradation. ....	198
Figure 5.22 CNOT7 variant G163C displays reduced activity in RNA degradation. ....	199
Figure 5.23 Group III CNOT7 variants are functional in RNA degradation. ....	200
Figure 5.24 Group IV CNOT7 variants are functional in RNA degradation. ....	201
Figure 5.25 Overview of the CNOT7 variants and their impact on the interaction with CNOT6L, CNOT1, and BTG1. ....	207

## Chapter 6

Figure 6.1 CNOT1 Maximum Likelihood phylogenetic tree.....	214
Figure 6.2 CNOT1 posterior mean $\omega$ graph.....	216
Figure 6.3 CNOT2 Maximum Likelihood phylogenetic tree.....	220
Figure 6.4 CNOT2 posterior mean $\omega$ graph.....	221
Figure 6.5 CNOT3 Maximum Likelihood (ML) phylogenetic tree.....	225
Figure 6.6 CNOT3 posterior mean $\omega$ graph. ....	226
Figure 6.7 Ccr4 Maximum Likelihood phylogenetic tree. ....	230
Figure 6.8 CNOT6 and CNOT6L posterior $\omega$ graph. ....	234
Figure 6.9 Caf1 Maximum Likelihood phylogenetic tree.....	237
Figure 6.10 CNOT8 posterior mean $\omega$ graph.....	239
Figure 6.11 CNOT9 Maximum Likelihood (ML) phylogenetic tree.....	243
Figure 6.12 CNOT10 Maximum Likelihood (ML) phylogenetic tree. ....	244
Figure 6.13 CNOT11 Maximum Likelihood (ML) phylogenetic tree. ....	245

## List of Tables

### Chapter 1

Table 1.1 Factors involved in cytoplasmic mRNA degradation.....	12
---	----

### Chapter 2

Table 2.1 Restriction endonucleases.....	55
Table 2.2 Standard Phusion High-Fidelity DNA polymerase reaction.....	57
Table 2.3 Standard Taq DNA Polymerase reaction.....	58
Table 2.4 List of primers for amplification from cells.....	59
Table 2.5 List of primers used for cloning .....	59
Table 2.6 List of primers for CNOT7 site directed mutagenesis.....	60
Table 2.7 List of primers for CNOT7 site directed mutagenesis.....	61
Table 2.8 List of primers for CNOT7 Gibson cloning into pCI $\lambda$ N vector. ....	61
Table 2.9 Mix for yeast transformation.....	67
Table 2.10 Mammalian cell lines used.....	70
Table 2.11 List of antibody used for immunoprecipitation.....	73
Table 2.12 siRNA used for knockdown. ....	74
Table 2.13 Transfection setup for RNA tethering assay .....	76
Table 2.14 Preparation of SDS-PAGE gels.....	80
Table 2.15 Primary antibodies used for western blotting. A.....	81
Table 2.16 Secondary antibodies used for western blotting. ....	81
Table 2.17 Parameters of the models used for Phylogenetic Analysis by Maximum-Likelihood (PAML).....	91

### Chapter 4

Table 4.1 CNOT1 damaging mutations.....	138
Table 4.2 CNOT2 damaging mutations.....	139
Table 4.3 CNOT3 damaging mutations.....	139
Table 4.4 Interfacing monomers.....	143
Table 4.5 CNOT1 and CNOT3 contacts. ....	143
Table 4.6 CNOT1 and CNOT2 contacts. ....	144

Table 4.7 Hydrogen bonds contacts between CNOT3 and CNOT2.....	145
Table 4.8 Salt bridges between CNOT3 and CNOT2.....	146

## Chapter 5

Table 5.1 Selected CNOT7 variants.....	170
Table 5.2 Predicted stability of CNOT7 variants.....	173
Table 5.3 Accommodation of CNOT7 variants.....	176

## Chapter 6

Table 6.1 Species analysed .....	211
Table 6.2 CNOT1 Likelihood Ratio Tests.....	215
Table 6.3 CNOT1 positively selected sites in the primates lineage.....	217
Table 6.4 CNOT2 Likelihood Ratio Test table. ....	221
Table 6.5 CNOT2 positively selected sites in primates. ....	222
Table 6.6 CNOT3 Likelihood Ratio test table.....	226
Table 6.7 CNOT3 positively selected sites in the primates lineage.....	227
Table 6.8 Ccr4 Likelihood Ratio Test table.....	231
Table 6.9 CNOT6 positively selected sites. ....	232
Table 6.10 CNOT6 positively selected sites in the primate lineage.....	233
Table 6.11 Caf1 Likelihood Ratio Test table.....	238
Table 6.12 CNOT8 positively selected sites in the primate lineage.....	240
Table 6.13 CNOT9, CNOT10, and CNOT11 Likelihood Ratio test table....	246

## List of Supplementary Material

### Chapter 3

Supporting Table 3.1	SILAC-RIME data analysed using Proteome Discover.
Supporting Table 3.2	Clusters of proteins identified by MCODE.
Supporting Table 3.3	Clusters of proteins identified by ClusterONE.

### Chapter 4

Supporting Table 4.1	Impact prediction of CNOT1 mutation using SIFT.
Supporting Table 4.2	Impact prediction of CNOT2 mutation using SIFT.
Supporting Table 4.3	Impact prediction of CNOT3 mutation using SIFT.
Supporting Table 4.4	CNOT1 variants interactions using RINalyzer.
Supporting Table 4.5	CNOT2 variants interactions using RINalyzer.
Supporting Table 4.6	CNOT3 variants interactions using RINalyzer.

### Chapter 5

Supporting Table 5.1	Impact prediction of CNOT7 mutation using SIFT.
Supporting Table 5.2	Impact prediction of CNOT7 mutation using SuSPect.
Supporting Table 5.3	Effect of amino acid substitution on the CNOT7 structure.

### Chapter 6

Supporting Figure 6.1	CNOT1 phylogenetic tree reconstructing using the Neighbour-Joining method.
-----------------------	--

Supporting Figure 6.2	CNOT2 phylogenetic tree reconstructing using the Neighbour-Joining method.
Supporting Figure 6.3	CNOT3 phylogenetic tree reconstructing using the Neighbour-Joining method.
Supporting Figure 6.4	Ccr4 phylogenetic tree reconstructing using the Neighbour-Joining method.
Supporting Figure 6.5	Caf1 phylogenetic tree reconstructing using the Neighbour-Joining method.
Supporting Figure 6.6	CNOT9 phylogenetic tree reconstructing using the Neighbour-Joining method.
Supporting Figure 6.7	CNOT10 phylogenetic tree reconstructing using the Neighbour-Joining method.
Supporting Figure 6.8	CNOT11 phylogenetic tree reconstructing using the Neighbour-Joining method.
Supporting Table 6.1	CNOT1 homologous proteins retrieved by DELTA-BLAST.
Supporting Table 6.2	CNOT2 homologous proteins retrieved by DELTA-BLAST.
Supporting Table 6.3	CNOT3 homologous proteins retrieved by DELTA-BLAST.
Supporting Table 6.4	Ccr4 homologous proteins retrieved by DELTA-BLAST.
Supporting Table 6.5	Caf1 homologous proteins retrieved by DELTA-BLAST.
Supporting Table 6.6	CNOT9 homologous proteins retrieved by DELTA-BLAST.
Supporting Table 6.9	CNOT10 homologous proteins retrieved by DELTA-BLAST.
Supporting Table 6.10	CNOT11 homologous proteins retrieved by DELTA-BLAST.

## List of Abbreviation

<b>m<sup>7</sup>G</b>	7-methyl-guanosine
<b>AIC</b>	Akaike Information Criterion
<b>APS</b>	Ammonium Persulphate
<b>Ago</b>	Argonaute
<b>ARE</b>	AU-Rich Elements
<b>BEB</b>	Bayes Empirical Bayes
<b>CED</b>	C-terminal Effector Domain
<b>CK</b>	C-terminal Knob domain
<b>COSMIC</b>	Catalogue of Somatic Mutations In Cancer
<b>CSPF</b>	Cleavage and Polyadenylation Specificity Factor
<b>CC</b>	Coiled Coil
<b>DNA</b>	Deoxyribonucleic Acid
<b>DMSO</b>	Dimethylsulfoxyde
<b>SuSPect</b>	Disease-Susceptibility-based SAV Phenotype Prediction
<b>DTT</b>	Dithiothreitol
<b>DELTA- BLAST</b>	Domain Enhanced Lookup Time Accelerated-Basic Local Alignment Tool
<b>DSE</b>	Downstream Element
<b>DPE</b>	Downstream Promoter Element
<b>DMEM</b>	Dulbecco's Modified Eagle Medium
<b>EDC</b>	Enhancer of Decapping
<b>EDTA</b>	Ethylenediaminetetraacetic Acid
<b>EJC</b>	Exon Junction Complex
<b>EEP</b>	Exonuclease-Endonuclease-Phosphatase
<b>FBS</b>	Fetal Bovine Serum
<b>GTR</b>	General Time Reversible
<b>GTF</b>	General Transcription Factor
<b>HBS</b>	Hepes-Buffered Saline
<b>Hb</b>	Hydrogen bond
<b>IPTG</b>	Isopropyl $\beta$ -D-1-thiogalactopyranoside

<b>LRR</b>	Leucine-Rich Repeats
<b>LRT</b>	Likelihood Ratio Test
<b>LB</b>	Lysogeny Broth
<b>ML</b>	Maximum Likelihood
<b>mRNA</b>	Messenger RNA
<b>miRNA</b>	MicroRNA
<b>miRISC</b>	MicroRNA Induced Silencing Complex
<b>NED</b>	N-terminal Effector Domain
<b>Nos</b>	Nanos domain
<b>NNI</b>	Nearest-Neighbour Interchange
<b>NEF</b>	Negative Elongation Factor
<b>NJ</b>	Neighbour-Joining
<b>NGD</b>	No-go Decay
<b>NMD</b>	Non-sense Mediated Decay
<b>NSD</b>	Nonstop Decay
<b>NAR</b>	Not Anchoring Region
<b>NES</b>	Nuclear Export Sequence
<b>NLS</b>	Nuclear Localisation Sequence
<b>NPC</b>	Nuclear Pore Complex
<b>OD</b>	Optical Density
<b>PAM2</b>	PABP-interacting Motif 2
<b>PBS</b>	Phosphate Buffered Saline
<b>PAML</b>	Phylogenetic Analysis by Maximum-Likelihood
<b>Poly(A)</b>	Poly adenosine
<b>PABP</b>	Poly(A) Binding Protein
<b>PAP</b>	Poly(A) Polymerase
<b>PAN</b>	Poly(A) specific Nuclease
<b>PARN</b>	Poly(A) specific Ribonuclease
<b>PAS</b>	Polyadenylation Site
<b>PEG</b>	Polyethylene glycol
<b>PCR</b>	Polymerase Chain Reaction
<b>PNK</b>	Polynucleotide Kinase

<b>b-PTF</b>	Positive Transcription Factor b
<b>PTM</b>	Post-Translational Modification
<b>PIC</b>	Pre-Initiation Complex
<b>pre-mRNA</b>	Pre-messenger RNA
<b>PTC</b>	Premature Termination Codon
<b>PDB</b>	Protein Data Bank
<b>PISA</b>	Proteins, Interfaces, Surfaces, Assemblies
<b>PUM</b>	Pumilio
<b>RIME</b>	Rapid Immunoprecipitation Mass Spectrometry of Endogenous Protein
<b>RIN</b>	Residue Interaction Network
<b>RNA</b>	Ribonucleic Acid
<b>rRNA</b>	Ribosomal RNA
<b>RBP</b>	RNA Binding Protein
<b>RNAP</b>	RNA Polymerase
<b>RRM</b>	RNA Recognition Motif
<b>Sb</b>	Salt bridge
<b>SH-like</b>	Shimodaira-Hasegawa-like
<b>siRNA</b>	Short Interference RNA
<b>SAV</b>	Single Amino acid Variant
<b>ssDNA</b>	Single Stranded DNA
<b>SDM</b>	Site Directed Mutator
<b>snRNA</b>	Small Nuclear RNA
<b>snoRNA</b>	Small Nucleolar RNA
<b>SDS-PAGE</b>	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
<b>SIFT</b>	Sorting Intolerant From Tolerant
<b>SILAC</b>	Stable Isotope Labelling of Amino acid in Cell culture
<b>SPR</b>	Subtree-Pruning-Regrafting
<b>SMG</b>	Suppressor with Morphogenetic effect on Genitalia
<b>TrN</b>	Tamura-Nei nucleotide substitution model
<b>TBP</b>	TATA Binding Protein
<b>TAF</b>	TBP Associated Factor

<b>TREX</b>	Transcription Export Complex
<b>TF</b>	Transcription Factor
<b>Inr</b>	Transcription Initiator sequence
<b>tRNA</b>	Transfer RNA
<b>TIM1</b>	Transition nucleotide substitution model
<b>TBST</b>	Tris buffer saline supplemented with 0.05% Tween-20
<b>TTP</b>	Tristetraprolin
<b>UTR</b>	Untranslated Region
<b>UPF</b>	UP-frameshift
<b>YPD</b>	Yeast extract Peptone Dextrose medium
<b>YS</b>	Yeast Selective complete medium

# **Chapter 1**

## **Introduction**

## **Chapter 1. Introduction**

### **1.1 Eukaryotic mRNA synthesis, processing and function**

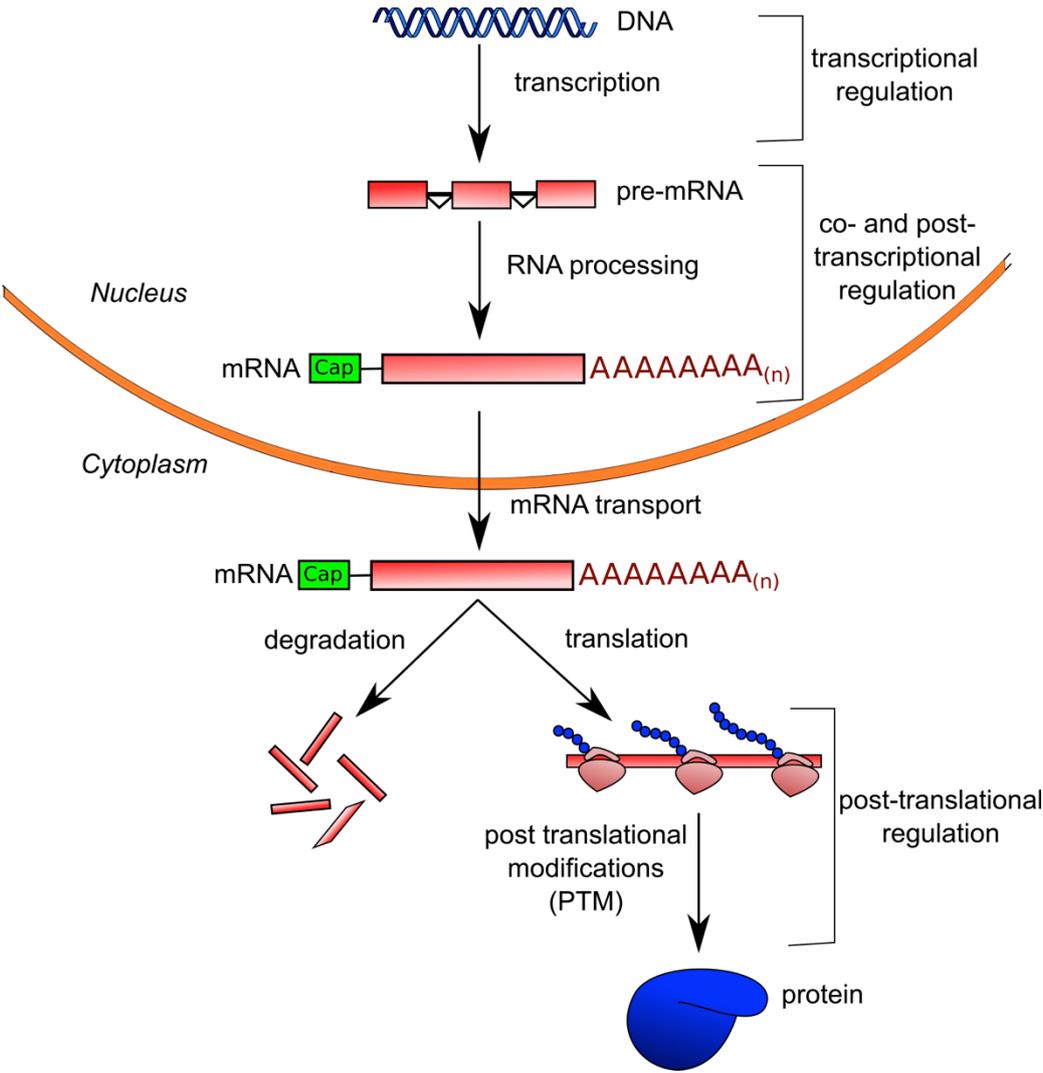
Eukaryotic cells adopt several processes to finely regulate genes and proteins expression. There are three main regulatory levels: transcriptional, co- and post-transcriptional, and post-translational regulation. An alteration of this balance caused by loss of control in one or more of the regulatory steps could cause several damages in cells leading to cell death or onset of disease e.g. cancer (Yeoh et al., 2002; Morris et al., 2010; Lee and Young, 2013; Eisenberg-Lerner et al., 2016).

The transcription is the first nuclear process of the gene expression and consists in copying the DNA double strand into a pre-messenger RNA (pre-mRNA). This single stranded molecule is subjected to modifications such as capping, splicing, polyadenylation, and editing to be converted into messenger RNA (mRNA). The mRNA is then transported from the nucleus to the cytoplasm where it undergoes translation or to mRNA degradation. Newly synthesised proteins are subjected to covalent and enzymatic alterations named post-translational modifications (PTM) (Figure 1.1).

#### **1.1.1 Eukaryotic transcription**

Transcription consists in copying DNA sequences into complementary molecules of RNA by RNA polymerase enzymes (RNAPs). There are five RNAPs in eukaryotic cells transcribing specific RNAs. The RNAP I transcribes the 35S precursor of ribosomal RNAs (rRNAs); RNAP II transcribes protein coding mRNAs, small nuclear RNAs (snRNAs) and microRNAs; RNAP III transcribes the short rRNA 5 S, U6 snRNA, and transfer RNA (tRNA) (Arimbasseri and Maraia, 2016). Plants have two additional unorthodox multi-subunits RNA polymerases that transcribe short interference RNAs (siRNAs), RNAP IV and RNAP V (Daxinger et al., 2009; Landick, 2009).

The transcription of mRNA is a dynamic and multi-step process performed by the RNAPII consisting in initiation, elongation, and termination (Cheung and Cramer, 2012).

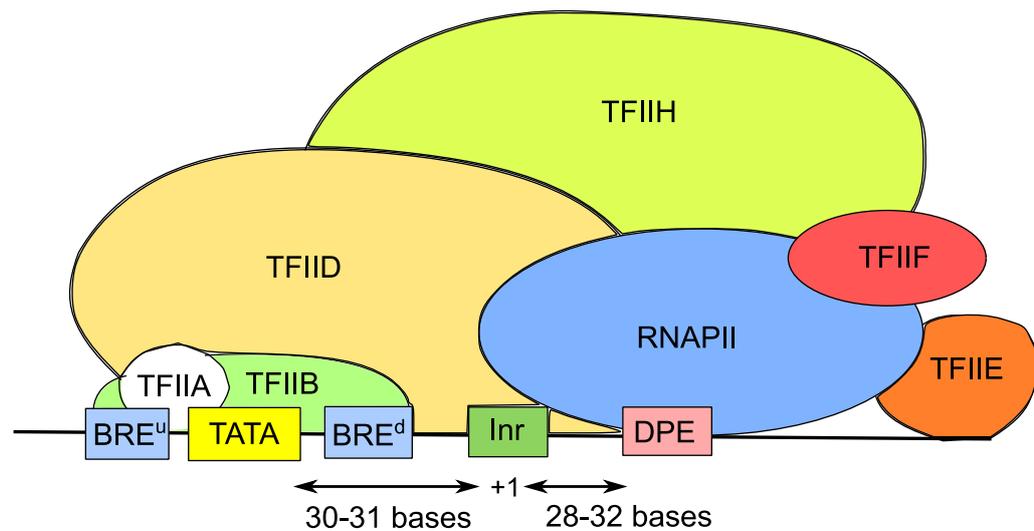


**Figure 1.1 Overview of eukaryotic gene regulation.** In eukaryotic cells, gene expression is regulated on three major levels: transcriptional, co- and post-transcriptional (mRNA capping, splicing, polyadenylation, RNA editing), and post-translational (amino-acidic chain modifications by attachment of functional groups) regulation.

The RNAP II is unable to recognise the minimum segment of DNA that is necessary to direct accurately the initiation of the transcription, the RNAP II core promoter. The polymerase therefore requires additional transcription factors and co-factors to localise on the core promoter (Juven-Gershon et al., 2008; Kadonaga, 2012). To initiate transcription, the RNAP II together with six general transcription factors (GTFs) forms the pre-initiation complex (PIC) (Thomas and Chiang, 2006). This is achieved by the GTFs recognition of elements of the core promoter. The assembly of the PIC is an important step for the transcription initiation specifying the transcription start site. Several models have been suggested for the assembly of the pre-initiation complex. A step-wise model of transcription proposes that transcription factor TFIID, composed by TATA binding protein (TBP) and TBP associated factors (TAFs), recognises a sequence of the core promoter, which is rich in T and A termed TATA box, the initiator element (Inr), and the downstream promoter element. This interaction is stabilised by TFIIA, followed by the recruitment of TFIIB, which in turns recruits the RNAPII-TFIIF sub-assembly. TFIIH and TFIIE are then recruited stimulating the separation of the two DNA helices (promoter melting) and the RNAPII escape from the promoter (Thomas and Chiang, 2006). Another model suggests that the RNAPII and some GTFs bind to the promoter as part of a pre-assembled holoenzyme (Wilson et al., 1996; Gupta et al., 2016; Zhang et al., 2016) Figure 1.2.

Most of the promoters require additional factors called activators, to initiate a regulated transcription (active transcription). The activators recognise specific DNA sequences often located upstream from the core promoter and alter the local chromatin environment facilitating the assembly of the PIC (Venters and Pugh, 2009).

Once the PIC is assembled on the DNA, the phosphorylation of Ser2 of the C-terminal domain of the RNAPII is required for the release of the Negative Elongation Factor (NEF), for the exchange of protein factors and for the escape of the polymerase from the promoter and initiate the elongation (Buratowski, 2009; Kostrewa et al., 2009; Eick and Geyer, 2013).



**Figure 1.2 Schematic representation of the pre-initiation complex (PIC) assembled on the RNAPII core promoter.** The typical RNAP II core promoter is composed by regulatory elements located up and down stream of the transcription initiator (Inr) site. The downstream promoter element (DPE) together with the Inr recruits the transcription factor TFIID and the RNA polymerase II (RNAP II) to the initiator site. The TATA box, and the flanking upstream and downstream B recognition element (BRE<sup>u</sup> and BRE<sup>d</sup>) located upstream of the Inr, recruit further transcription factor that stimulate the RNAP II to start an efficient transcription.

In the early stage of the elongation, the RNAPII undergoes an abortive initiation stage until the transcript length reaches 10-12 nucleotides. The recruitment of the Positive Transcription Factor b (b-PTF) and other Elongation Factors such as elongins, DSIF, Stp6, TFIIS, and the transcription export complex (TREX) promote the synthesis of long transcripts and finely regulate the elongation (Marshall and Price, 1995; Sims et al., 2004; Peterlin and Price, 2006; Sharma, 2016). The immature mRNA molecule synthesised by the RNA polymerase II (RNAPII) is termed pre-mRNA.

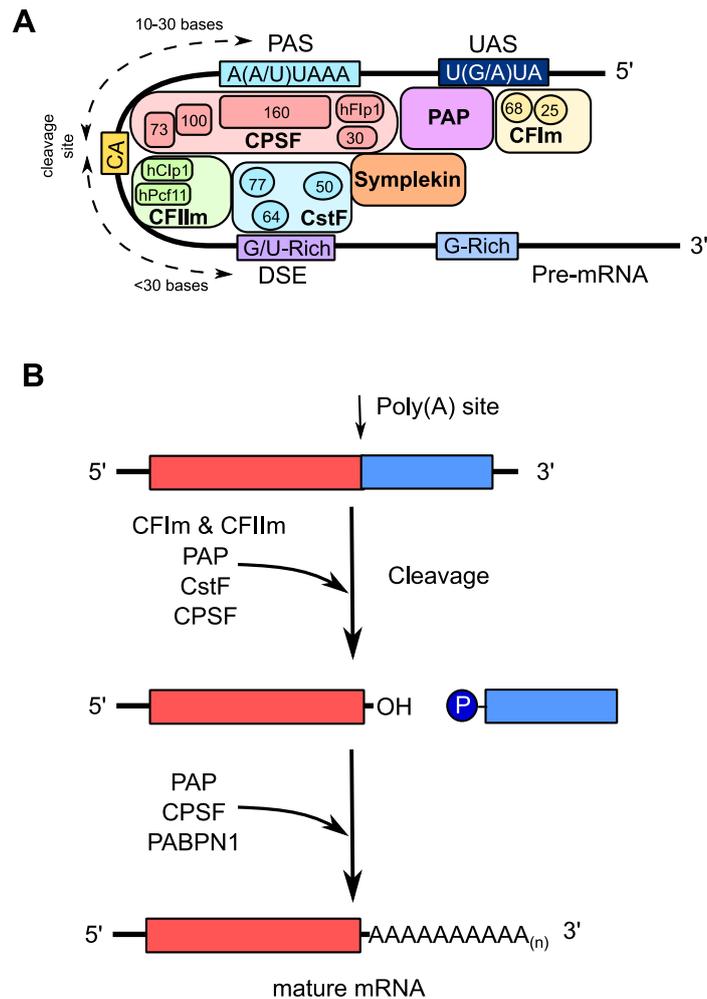
### **1.1.2 mRNA processing**

The pre-mRNA is subjected to three modifications in order to become mature mRNA. Transcripts are modified at their 5' ends by the addition of the 7-methylguanosine cap, internal removal of intronic sequences, polyadenylation at the 3' end, and editing (Topisirovic et al., 2011; Girard et al., 2012; Proudfoot, 2011). Shortly after the 5' end of the nascent RNA molecule emerges, the capping enzyme complex is recruited. This complex catalyses an unusual 5' to 5' triphosphate bridge attaching the guanidine residue at the end of the transcript that is then methylated in position 7 by a methyltransferase forming the 5' m<sup>7</sup>G cap. This cap stabilises the RNA against the attack by exonucleases and promotes transcription, splicing, and polyadenylation and is required for the translation (Cowling, 2009; Braun and Young, 2014). Therefore, transcripts undergo a further process termed splicing. Introns are excised from the RNA molecule leaving the protein coding sequences. Alternative splicing joins gene exons in different ways to produce alternative transcripts called splice variants (Johnson et al., 2003; Johnson and Vilardeell, 2012; Hossain and Johnson, 2014). This process is catalysed by the ribonucleoprotein complex spliceosome, composed by proteins and small nuclear RNA (snRNA) molecules (McManus and Graveley, 2011; Sperling, 2016).

### 1.1.3 Polyadenylation

The final step to produce a mature mRNA is the polyadenylation at the nascent RNA 3' end. The poly(A) tail has multiple roles: it promotes translation, nuclear export, and confers stability to the RNA, protecting it from ribonuclease attack. The polyadenylation flags the termination of the transcription and the release of the mature mRNA. Transcripts with long poly(A) tail were shown to be more stable and have increased half-lives. The control of the length of the poly(A) tail is finely regulated both in the nucleus for nascent mRNAs and in the cytoplasm after the export from the nucleus (Colgan and Manley, 1997; Eckmann et al., 2011; Proudfoot, 2011). The process is a two-steps reaction consisting in endonucleolytic cleavage and synthesis of a non-templated 3' poly adenosine tail, which is approximately 200-300 adenosine residues long (Sheets and Wickens, 1989; Zhang et al., 2010). The cleavage/polyadenylation machinery requires specific signals in the pre-mRNA sequence. The polyadenylation site consists in three primary sequence elements the cleavage site, the downstream element (DSE), and the polyadenylation site (PAS). Mammalian PASs are AU-rich and are a highly conserved in their nucleotide composition (Tian and Graber, 2012; Shi, 2012). The A(A/U)UAAA sequence is recognised by the cleavage and polyadenylation specificity factor (CSPF) via CSF-160, its largest subunit. The CSPF complex binds to the pre-mRNA cleavage site, typically a CA sequence, via the CPSF-73 subunit. The downstream element DSE is an U/GU-rich region recognised by the CstF complex. The CSPF, CstF, and CFI bind to directly to the RNA and recruit in turn other factors such as the CFII, Symplekin, and the poly(A) polymerase (PAP) (Shi et al., 2009; Chan et al., 2011) (Figure 1.3). The co-transcriptional assembly of all these factors on the PASs is facilitated by the C-terminal domain of the RNAPII (Bentley, 2005; Shi and Manley, 2015).

The accumulation of processing factors at the 3' end triggers the termination of the transcription and the release of the RNAPII is stimulated through XRN2 exonuclease degradation downstream the PAS signal (West et al., 2008).



**Figure 1.3 Mammalian pre-mRNA 3' end processing machinery. (A)**

The nuclear cleavage and polyadenylation complex. The CPSF complex recognise the poly(A) signal (PAS) located upstream of the cleavage site rich in CA. The cleavage factor CFIIm recognises an upstream sequence (UAS) whereas the CstF complex recognises a G/U rich downstream element (DSE). The interaction of CPSF, CStF, Symplekin, CFIIm and CFIIm results in the recruitment of the poly(A) polymerase (PAP). **(B)** The cleavage and polyadenylation process to form mature mRNA. Mammalian processing machinery cleave the pre-RNA at the 3' end followed by addition of non-templated poly(A) tail.

#### **1.1.4 mRNA nuclear export**

Following the RNA processing events, the mRNA is transported into the cytoplasm where it can be translated into proteins or subjected to mRNA degradation. The exchange between the nucleus and cytoplasm is carried out by the Nuclear Pore Complex (NPC) composed by multiple copies of nucleoporins. Phenylalanine-glycine (FG) peptides within the pore channel establish a selective barrier, enabling the passive diffusion of small molecules or the selective transport of molecules greater than ~40 kDa. Nuclear localisation or nuclear export (NLS/NES) sequences are recognised by transport receptors (importins/exportins) and can interact with the FG repeats (Terry and Wentz, 2009; Wentz and Rout, 2010; Floch et al., 2014). A requisite for the mRNA export is the completion of the pre-mRNA processing reactions and the recruitment of mRNA export receptor that interacts with the FG-nucleoporins at the NPC. During the transcription, the RNAPII C-terminal domain, the splicing exon junction complex (EJC) and the 5'cap recruit the transcription export complex (TREX). The TREX complex is composed by the THO complex and other export factors including Yra1p, ALY, POLDIP3, and DEAD-box RNA helicase UAP56 (Carmody and Wentz, 2009; Folco et al., 2012; Heath et al., 2016). This complex facilitates the association of the mRNA with export receptor NXF1 and NXT1 via ALY and UAP56 (Hautbergue et al., 2008; Taniguchi and Ohno, 2008). Another pathway for the mRNA export is Crm1 dependent. This receptor cannot bind directly to the mRNA, but recognises distinct mRNA specific NES-containing adaptors such as NXF3 or LRPPRC (Cullen, 2003; Culjkovic-Kraljacic and Borden, 2013; Natalizio and Wentz, 2013).

Once in the cytoplasm, the mRNA undergoes translation to form proteins or mRNA degradation.

## 1.2 Regulated mRNA decay

Regulation of the gene expression in eukaryotic cells occurs at multiple levels. In addition to the synthesis, processing, and nuclear export rates, the RNA decay rate in the cytoplasm plays a vital role in endorsing transcriptional regulation (Wu and Brewer, 2012). Interestingly, mRNA decay rates differ notably between transcripts in eukaryotic organisms; mRNAs coding for housekeeping proteins tend to have longer half-lives compared to those encoding for regulatory proteins (Yang et al., 2003; Chen et al., 2008). The rate of mRNA degradation may change in response to stimuli, giving the cell a rapid method of altering gene expression.

The stability of the mRNA is influenced by the 5' m<sup>7</sup>G cap and the poly(A) tail at the 3'. The 5' cap interacts with the eIF4E protein and the poly(A) tail interacts with the poly(A) binding proteins (PABP); eIF4E and PABPs interact via eIF4F forming the closed loop configuration, necessary for an efficient protein synthesis. These structures stabilise the mRNA protecting it from exonuclease attack. To initiate the mRNA degradation, one of these components must be perturbed, or the mRNA must be cleaved internally, losing the closed loop configuration (Garneau et al., 2007; Mitchell et al., 2010; Łabno et al., 2016). There are three classes of RNA-degrading enzymes: endonucleases that cleave the RNA internally, 5' exonucleases that degrade the RNA from the 5' end, and 3' exonucleases that degrade the RNA from the 3' end.

In most cases, the precursor of the mRNA degradation is the removal of the poly(A) tail which must be shortened to 10 nucleotides (Couttet et al., 1997; Meyer et al., 2004; Houseley and Tollervey, 2009).

This process is the initial step in cytoplasmic mRNA degradation involving several deadenylases (Parker and Song, 2004; Garneau et al., 2007; Wahle and Winkler, 2013).

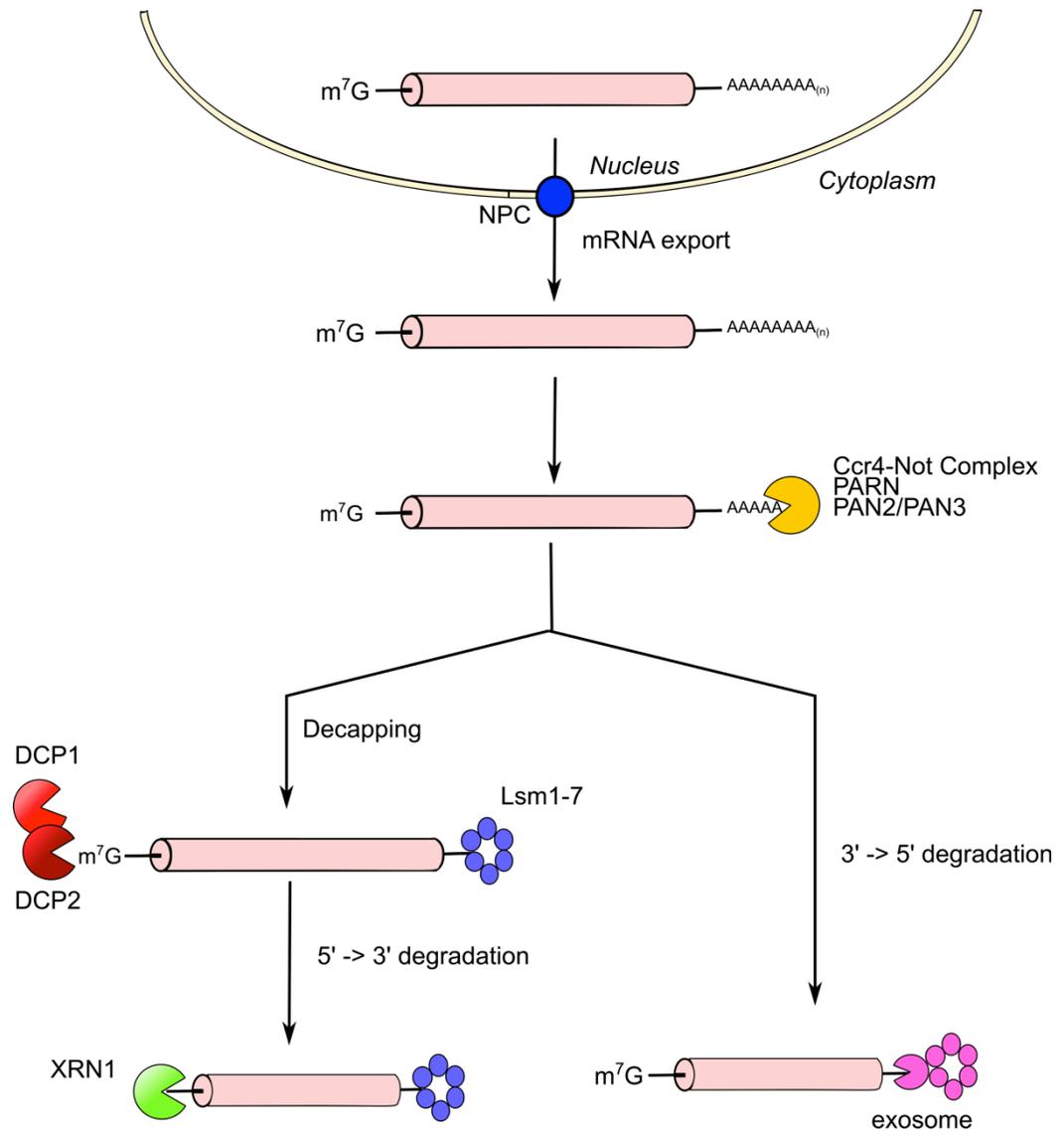
Two pathways have been described for the eukaryotic cytoplasmic deadenylation-dependent mRNA degradation (Figure 1.4). In the first pathway, following the deadenylation, the decapping complex removes the 5' m<sup>7</sup>G cap. Lsm1-7 proteins bind to the 3' end of the deadenylated transcript to assist the decapping enzymes Dcp1-Dcp2 in the removal of

the 5' m<sup>7</sup>G cap, resulting in the exposure of the mRNA to digestion by the 5'-3' exonuclease XRN1 (Couttet et al., 1997; Ingelfinger et al., 2002; Yamashita et al., 2005; Behm-Ansmant et al., 2006).

In the second pathway, following deadenylation, the 3' end of the transcript is directly exposed to the digestion by the 3'-5' exosome multi-subunit complex (Parker and Song, 2004; Garneau et al., 2007; Goldstrohm and Wickens, 2008; Wu and Brewer, 2012; Łabno et al., 2016) (Figure 1.4). Factors involved in the cytoplasmic mRNA decay pathways are shown in Table 1.1.

<b>Complex</b>	<b>Human name</b>	<b>Function</b>
<b>Decapping factors</b>	Dcp1a, Dcp1b	Member of Dcp decapping complex
	Dcp2	Catalytic subunit of the decapping complex
	Lsm1-7	Lsm 3' RNA binding complex
	Pat1	Required for mRNA decapping
<b>Deadenylase enzymes</b>	Nocturnin	EEP-type deadenylase
	ANGEL	EEP-type deadenylase
	ANGEL 2	EEP-type deadenylase
	CNOT6 (Ccr4a)	EEP-type deadenylase subunit of the Ccr4-Not complex
	CNOT6L (Ccr4b)	EEP-type deadenylase subunit of the Ccr4-Not complex
	TOE1 (Caf1z)	DEDD-type deadenylase
	PARN	DEDD-type deadenylase
	PAN2	DEDD-type deadenylase that interacts with PAN3
	CNOT8 (Caf1b)	DEDD-type deadenylase subunit of the Ccr4-Not complex
	CNOT7 (Caf1a)	DEDD-type deadenylase subunit of the Ccr4-Not complex
<b>Exoribonuclease enzymes</b>	XRN1	Cytoplasmic 5'-3' exoribonuclease
	Exosome	3'-5' exoribonuclease complex

**Table 1.1 Factors involved in cytoplasmic mRNA degradation.**



**Figure 1.4 Cytoplasmic mRNA degradation pathways.** The mRNA is exported from the nucleus by the nuclear pore complex (NPC). In the cytoplasm deadenylase enzymes (Ccr4-Not complex, PARN, Pan2-Pan3) shorten the poly(A) tails. Deadenylation stimulates the degradation of the mRNA by recruitment of the decapping complex DCP1/DCP2 followed by 5' → 3' exonucleolytic digestion by XRN1. Alternatively, following deadenylation, the exosome degrades the transcript in direction 3' → 5'.

### 1.2.1 Deadenylase enzymes

Deadenylases are  $Mg^{2+}$  dependent exoribonucleases that hydrolyse the degradation of the poly(A) tail in 3'-5' direction. Studies identified ten different deadenylase enzymes in human cells (Goldstrohm and Wickens, 2008). It is established that the majority of the cytoplasmic deadenylase activity is associated with the Ccr4-Not complex, whose subunits are Ccr4a/Ccr4b and Caf1a/Caf1b, and with Pan2-Pan3 complex (Wahle and Winkler, 2013). Additional deadenylases enzymes expand the repertoire of poly(A) tail length regulatory elements in eukaryotic cells (Godwin et al., 2013). These enzymes are categorised into two groups depending on their nuclease domain. The exonuclease-endonuclease-phosphatase (EEP) family consists in nucleases containing conserved Asp and His residues coordinating  $Mg^{2+}$  ions. Ccr4a/CNOT6, Ccr4b/CNOT6L, Nocturnin, and ANGEL are member of this family. The second group of deadenylase enzymes consists of DEDD-type nucleases, which contain Asp and Glu residues in the exonuclease motifs. The DEDD family exonuclease includes Caf1a/CNOT7, Caf1b/CNOT8, poly(A)-specific ribonuclease (PARN), and PAN2 (Table 1.1) (Dlakić, 2000; Zuo and Deutscher, 2001).

Poly(A) specific ribonuclease (PARN) is a nuclear and cytoplasmic exonuclease that specifically degrades the poly(A) tail and is present as a dimer (Körner and Wahle, 1997; Wu et al., 2005). Interestingly, in addition to its nuclease activity, this protein interacts with the 5' m<sup>7</sup>G cap stimulating the deadenylase activity and the enzymatic processivity (Dehlin et al., 2000; Martinez et al., 2000, 2001). PARN is composed of at least three domains: the DEDD-type catalytic domain, the R3H domain and the RNA recognition motif (RRM). PARN requires two divalent  $Mg^{2+}$  ions in its catalytic domain. The two ions are coordinated by the residue Asp-28, Asp-292 and Asp-282, with Asp-28 being the most important for the  $Mg^{2+}$  binding (Ren et al., 2002a, 2002b, 2004; Wu et al., 2005; He et al., 2011). The RRM domain, in addition to its role in the RNA binding, is responsible for the interaction with the m<sup>7</sup>G cap via Trp 475 (Monecke et al., 2008; Nagata et al., 2008). The RH3 domain can bind to the RNA and interacts

with the active site of the other subunit (Wu et al., 2005). PARN is involved in the maturation of the 3' end of the small nucleolar RNA (snoRNAs), in the response to the DNA damage, and in the nonsense mediated mRNA decay (Lejeune et al., 2003; Cevher et al., 2010; Berndt et al., 2012).

Nocturnin, Angel 1 and Angel 2 are distant homologues of the Ccr4 proteins, belonging to the EEP deadenylase group, but lack the leucine-rich repeats domain responsible for the interaction with the Ccr4-Not complex (Wahle and Winkler, 2013). Nocturnin deadenylase activity was reported only in few organisms. It is suggested to be a circadian protein, and to be involved in inflammatory responses and nutrient metabolism (Godwin et al., 2013). Angel 1 and Angel 2 bind to a distant homolog of Caf1, Caf1z/TOE. This complex has deadenylase activity, but its biological function is not clear (Wagner et al., 2007).

Poly(A) specific nuclease (PAN) was firstly identified in *S. cerevisiae* as a DEDD type nuclease. It is composed by two subunits, Pan2 (127 kDa) and Pan3 (76 kDa) and is mostly found in cytoplasm (Boeck et al., 1996a; Garneau et al., 2007). Pan2 enzyme releases 5'-AMP and require a 3'-OH group in its substrate. Pan2-Pan3 recruitment on the mRNA is PABP dependent. Pan3 interacts with the poly(A) binding protein 1 PABPC1 via the PAM2 motif (PABP-interacting motif 2) and stimulates Pan2 deadenylase activity (Wolf and Passmore, 2014). Pan2-Pan3 complex shortens the poly (A) tail to approximately 110 nt. Deadenylation carried by Pan2-Pan3 complex is followed by degradation of the poly(A) tail by the Ccr4-Not complex. (Yamashita et al., 2005; Bartlam and Yamamoto, 2010).

The Ccr4-Not complex is the major deadenylase complex and is highly conserved among organisms. Two deadenylase enzymes are the catalytic subunits of the complex, the EEP-type Caf1 and the DEDD-type Ccr4. These enzymes regulate deadenylation and mRNA turnover (Tucker et al., 2001). Vertebrates have two homologues per protein, CNOT6 and CNOT6L are

Ccr4 homologues and CNOT7 and CNOT8 are Caf1 homologues, but only one DEDD-type and one EEP-type can assemble to the complex at a time (Lau et al., 2009; Winkler and Balacco, 2013). Ccr4 subunits have a leucine-rich repeat domain (LRR) responsible for the interaction with the Caf1 subunits (Basquin et al., 2012). Deadenylation is a finely regulated process that takes place both in the nucleus and in the cytoplasm. Nuclear deadenylation is thought to restrict the length of newly added mRNA poly(A) tail, while the cytoplasmic shortening of the poly(A) tail initiates repression and degradation of mRNA (Bianchin et al., 2005; Goldstrohm and Wickens, 2008; Bartlam and Yamamoto, 2010).

### **1.2.2 Decapping complex and its regulation**

The major pathway of the mRNA degradation occurs through deadenylation, decapping, and exonucleolytic digestion of the mRNA body. The removal of the 5' m<sup>7</sup>G cap is a critical process in the degradation of the mRNA. The most well-characterised and highly conserved eukaryotic decapping enzyme is Dcp2, originally identified in *Saccharomyces cerevisiae* (Dunckley and Parker, 1999). This protein contains a Nudix domain responsible for the hydrolysis of methylated capped-mRNA with the release of m<sup>7</sup>GDP (Bessman et al., 1996; van Dijk et al., 2002). Dcp2 interacts directly with the essential cofactor Dcp1 that lacks intrinsic decapping activity and acts by stimulating Dcp2 activity (LaGrandeur et al., 1998; She et al., 2008). Studies showed that Dcp2 prefers substrates longer than 25 nucleotides (Piccirillo et al., 2003; Steiger et al., 2003).

The activity of the decapping Dcp1-Dcp2 complex is stimulated by several enhancers. Studies showed that enhancer of decapping 1 (EDC1) and 2 (EDC2) directly bind to the RNA stimulating the decapping *in vitro* (Schwartz et al., 2003). The highly conserved EDC3 interacts directly with the Dcp1-Dcp2 complex and studies revealed that EDC3 deficient yeast strains have reduced decapping rates (Kshirsagar and Parker, 2004; Fromm et al., 2012). Metazoan present an additional enhancer EDC4

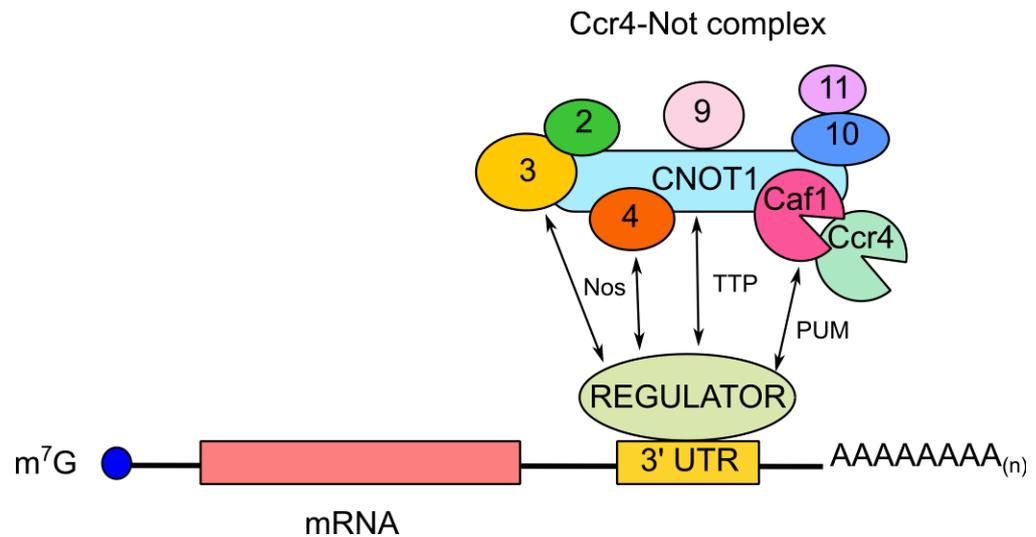
which stimulates association between Dcp1 and Dcp2 (Fenger-Grøn et al., 2005).

Furthermore, another decapping enhancer is Pat1, which interacts directly with the Dcp1-Dcp2 complex and stimulates its activity. This protein was found in association with other enhancer of decapping factors such as the Lsm1-7 complex. The Lsm1-7 complex promotes decapping by interacting with the 3' end of deadenylated mRNA (Tharun et al., 2000; Tharun and Parker, 2001; Chowdhury et al., 2007; Nissan et al., 2010). The DEAD box superfamily 2 helicase Dhh1 (called DDX6 or Rck/p54) acts as an enhancer of Dcp1-Dcp2 and was found to interact with both decapping and deadenylase complexes (Fischer et al., 2002; Nissan et al., 2010; Arribas-Layton et al., 2013).

### 1.2.3 Protein-mediated mRNA decay

The mRNA decay can be triggered by proteins that mediate the recruitment of the deadenylase complexes via regulatory elements in the mRNA sequences and act as translational repressor (Figure 1.5) (Inada and Makino, 2014). Tristetraprolin (TTP) is a RNA binding protein that mediates the recruitment of the Ccr4-Not complex and triggers the degradation of mRNA containing AU-rich elements (AREs) (Sandler et al., 2011; Brooks and Blackshear; Fabian et al., 2013). Messenger RNA containing AU-rich elements constitute 5-8% of the human transcriptome (Chen and Shyu, 1994; Halees et al., 2008). These elements typically contain one or more UAUUUAU consensus heptamers, are mostly found in the 3'UTR and modulate mRNA translation, stability, and degradation. AREs are particularly abundant in mRNAs encoding for interleukins, proto-oncogenes (c-Fos, c-Myc) and cytokines (GM-CSF, tumour necrosis factor  $TNF\ \alpha$ ) (Wu and Brewer, 2012; Brooks and Blackshear; Hamid and Makeyev, 2016). The recruitment of the Ccr4-Not complex by TTP occurs via the CONT1 central scaffold of the deadenylation complex (Sandler and Stoecklin, 2008; Sandler et al., 2011). Protein mediated recruitments can occur via direct interaction with deadenylase subunits. In *S. cerevisiae* the mRNA of the DNA endonuclease for mating-type switching *HO* is targeted by the Pumilio family

protein Mpt5. This protein binds to the 3' UTR of the *HO* mRNA and recruits the Ccr4-Not complex via the CAF1 subunit (Goldstrohm et al., 2006, 2007). Studies in *Drosophila melanogaster* identified the interaction of Pumilio (PUM) proteins with CAF1/POP2, and NANOS (Nos) with the not stably associated NOT4 subunit recruiting the Ccr4-Not complex on the Cyclin B and *hunchback* mRNA (Kadyrova et al., 2007; Joly et al., 2013). Recently, studies in *D. melanogaster* characterised the recruitment with the Ccr4-Not complex by interaction of Nanos and the NOT1-NOT3 subunits of the deadenylase complex highlighting a redundancy in Nanos domains (Raisch et al., 2016). Moreover, Pumilio and Nanos interact with each other. Nanos proteins play an essential role in embryonic germ line specification and maintenance, and neuronal homeostasis (Jaruzelska et al., 2003; Baines, 2005; Lai and King, 2013).

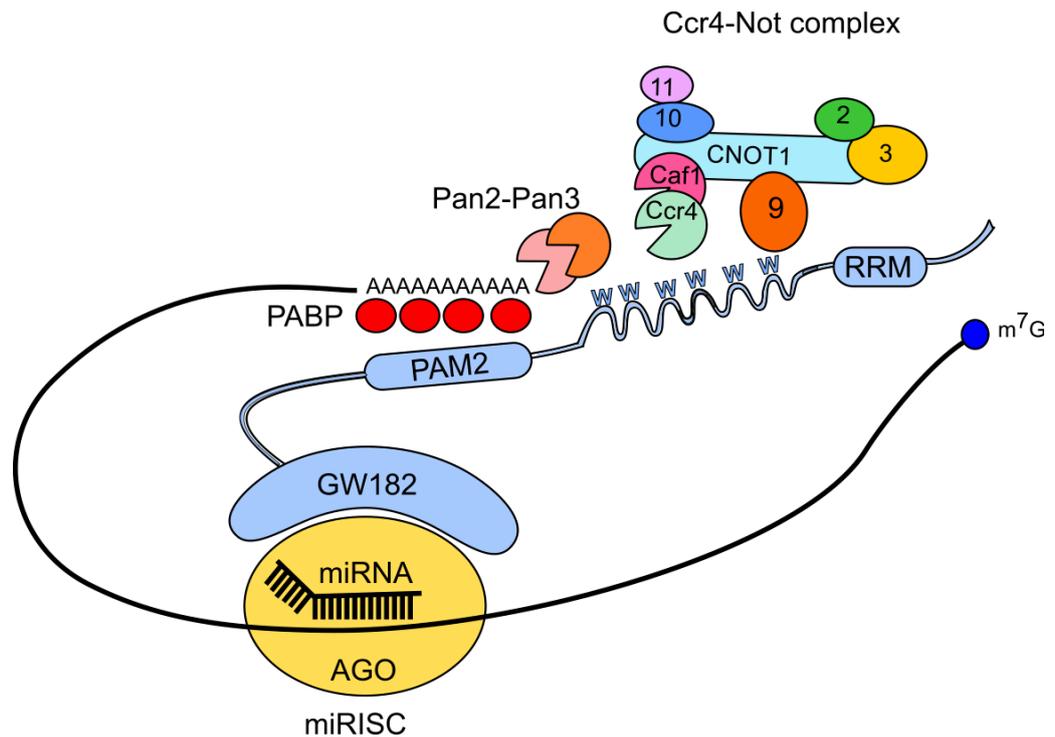


**Figure 1.5 Protein-mediated recruitment of the Ccr4-Not complex.**

The mRNA decay can be triggered by proteins that mediate the recruitment of the deadenylase complexes via regulatory elements that interact directly with components of the Ccr4-Not complex. Pumilio proteins (PUM) interact with the Caf1 subunit, Tristetraprolin (TTP) with CNOT1, and NANOS (Nos) proteins interact with CNOT1 and CNOT3 (3) or with CNOT4 (4), stable member of the yeast complex. Indicated are the remaining components of the complex CNOT2 (2), CNOT9 (9), CNOT10 (10), CNOT11 (11), and Ccr4.

#### 1.2.4 microRNA mediated decay

MicroRNAs (miRNAs) are short (~ 21 nucleotides) non-coding RNAs that regulate gene expression post-transcriptionally. These short molecules regulate several cell processes such as cell differentiation, cell proliferation, apoptosis, metabolism, development (Eulalio et al., 2008). Micro RNAs repress protein synthesis by base pairing imperfectly to the 3'UTR of target mRNA, resulting in the co-translational protein degradation, inhibition of translation elongation, premature dissociation of ribosomes, inhibition of translation initiation or promoting mRNA degradation (Bushati and Cohen, 2007; Chen and Shyu, 2011; Huntzinger and Izaurralde, 2011). During its biogenesis, miRNAs are loaded by the miRNA induced silencing complex (miRISC) onto Argonaute proteins (AGO) forming a ribonucleoprotein complex. The complex recruits GW182 proteins in order to play its role (Chendrimada et al., 2005; Pfaff and Meister, 2013; Catalanotto et al., 2016). GW182 proteins are characterised by a N-terminal effector domain (NED) rich in G/W (glycine-tryptophan) repeats, a Q-rich domain and a C-terminal effector domain (CED) containing PAM2 motifs and a RNA recognition motif (RRM) (Eulalio et al., 2009; Zekri et al., 2009; Huntzinger et al., 2010). It was found that mammals encode three GW182 paralogues TNRC6A, TNRC6B, and TNRC6C. The TNRC6 proteins bind to AGO through GW repeats in their NED (Nishi et al., 2015). The CED is considered to be the “silencing domain” and is an unstructured bipartite region formed by a Mid region, containing M1 and M2 motifs flanking PAM2 motif, and the C-terminus. Through the PAM2 motif, TNRC6 proteins interact with PABP allowing them to recruit Pan2-Pan3 to the miRNA silencing complex (miRISC) and to promote mRNA degradation (Jinek et al., 2010; Fabian et al., 2012; Huntzinger et al., 2013). Moreover, GW182 proteins can interact with W-binding pockets of the CNOT9/RCD1/RQCD1 subunit of the Ccr4-Not complex (Figure 1.6) (Chen et al., 2014; Kuzuoğlu-Öztürk et al., 2016).



**Figure 1.6 Recruitment of the human deadenylase complexes by miRNA machinery.** MicroRNA bound to Argonaute protein (AGO) recruits GW182 through GW repeats in GW182. In turn, GW182 proteins recognise the RNA by the RNA recognition motif and interacts with poly(A)-binding protein (PABP) via its PAM2 motif recruiting the Pan2-Pan3 deadenylase complex. Moreover, GW182 proteins can interact with W-binding pockets of the CNOT9 subunit of the Ccr4-Not complex.

### 1.3 Quality control of the mRNA

Surveillance of the mRNA protects cells from aberrant mRNAs that would produce potentially defective proteins. The quality control of the mRNA leads to the degradation of not properly processed mRNAs. There are three mechanisms present in eukaryotic cells: non-sense mediated decay (NMD), nonstop decay (NSD), and no-go decay (NGD) (Doma and Parker, 2007; Isken and Maquat, 2007; Wu and Brewer, 2012).

#### 1.3.1 Nonsense-mediated decay (NMD)

Aberrant mRNAs containing in-frame premature termination codons (PTCs) are recognised by the nonsense-mediated decay (NMD) machinery, preventing the production and accumulation of potentially deleterious truncated proteins in cells (Smith and Baker, 2015; Karousis et al., 2016). PTCs recognition results in an accelerated degradation triggered by deadenylation and endonucleolytic activity (Bhuvanagiri et al., 2010). Three trans-acting factors are required for the recognition of PTC containing mRNAs: UPF1 (UP-frameshift 1), UPF2, and UPF3. In addition to these core proteins, Suppressor with Morphogenetic effect on Genitalia (SMG) proteins are required for the nonsense-mediated decay (Behm-Ansmant et al., 2006). SMG proteins are phosphatidylinositol 3-kinase-related kinases; SMG1 phosphorylates UPF1 increasing its affinity for mRNA decay factors such as decapping, deadenylase, and endonuclease enzymes (Lejeune et al., 2003; Wu and Brewer, 2012). Two models were proposed to describe the recruitment of the NMD machinery on mRNAs containing PTCs. In the first model, the exon-exon junction complex (EJC) plays a key role in the recognition of the PTC. During translation, if the ribosome encounters a PTC upstream of the EJC, the ribosome stalls and the SURF complex, composed by SMG1, UPF1 and eRFs, is recruited on the mRNA leading to the activation of the NMD pathway (Figure 1.7 A) (Lejeune and Maquat, 2005). In the "faux 3' UTR" model, the premature stop codon and poly(A) tail are distant resulting in the inability of PABPC1 to interact with the terminating ribosome. This leads to a less efficient termination, ribosome pausing and enhanced interaction between UPF1 and translation release factor eRFs that promotes the NMD and

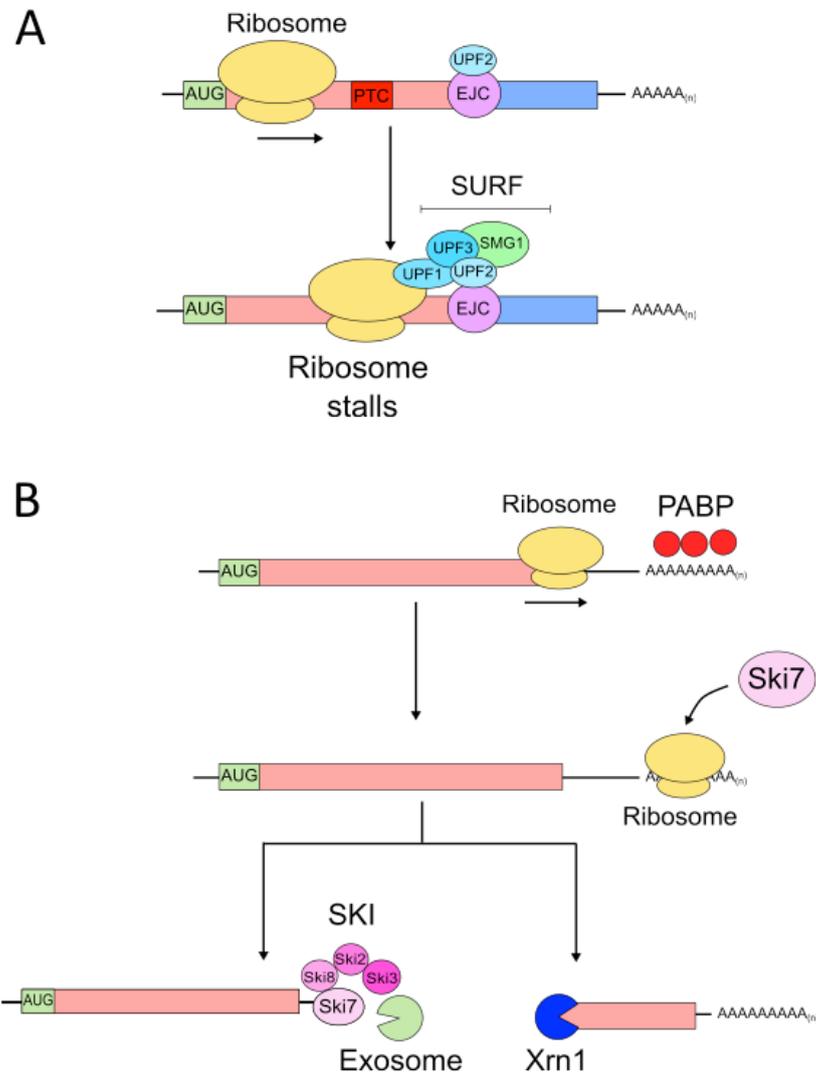
the activation of downstream mRNA decay events (Amrani et al., 2004; Eberle et al., 2008; Singh et al., 2008). In mammalian cells, SMG5 and SMG7 proteins mediate the rapid degradation of the PTC-containing mRNAs recruiting the Ccr4-Not complex by direct interaction with the Caf1 deadenylase subunit (Loh et al., 2013).

### **1.3.2 Nonstop mRNA decay (NSD) and No-go decay (NGD)**

Loss of termination codon in mRNAs permits the ribosome to reach the very 3' end leading to the degradation of the nucleic acid (van Hoof et al., 2002). In *S. cerevisiae* two mechanisms of nonstop mRNA decay (NSD) were described: the Ski7 dependent and the non-Ski7 dependent pathway. Ski7 is a molecular mimic of the GTPase domains of the elongation factor EF1  $\alpha$  and the release factor eRF3. When the ribosome traverses the 3'-UTR and poly(A) tail, stalls and allows Ski7 to recognise the unoccupied A-site of the ribosome (van Hoof et al., 2002; Maquat, 2002; Isken and Maquat, 2007). Therefore, Ski7 recruits the SKI complex, composed by Ski2, Ski3, and Ski8. This complex recruits in turn the exosome for the degradation of the mRNA. A less efficient non-Ski7 dependent mechanism involves sequential decapping and degradation by Xrn1 exonuclease (Wu and Brewer, 2012; Horikawa et al., 2016). Ski7 is yeast specific and is not present in mammalian cells. Studies in HeLa cells found that the Hbs1-Dom34 complex binds to the stalled ribosome and recruits the exosome-SKI complex leading to the aberrant mRNA degradation (Saito et al., 2013).

The Hbs1-Dom34 complex acts as a ribosome-recycling factor and plays a key role also in the No-go decay (NGD). Ribosomes stalled on mRNAs with ribosome barriers, such as strong secondary structures or mRNAs that cannot properly elongate, recruit the Hbs1-Dom34 complex (Doma and Parker, 2006, 2007; Houseley and Tollervey, 2009). Hbs1 and Dom34 are two GTPase family member endoribonucleases (Pisarev et al., 2007). The interaction of the Hbs1-Dom34 complex with the stalled ribosome stimulates the peptide-tRNA hydrolysis, the peptide release and the mRNA cleavage. Xrn1 and exosome are then recruited to degrade the mRNA in 5'

and 3' directions (Doma and Parker, 2006; Lee et al., 2007; Graille et al., 2008).



**Figure 1.7 Nonsense mediated decay (NMD) and No-stop decay (NSD).**

**(A)** Upon translation of mRNA containing premature termination codons (PTCs), the ribosome stalls on a PTC upstream of the exon junction complex (EJC) and the SURF complex, composed by SMG1, UPF1, UPF2, and UPF3 is recruited on the mRNA leading to the activation of the NMD pathway. **(B)** In mRNA lacking stop codons, the ribosome is allowed to reach the very 3' end and poly(A) tail of the mRNA. Ski7 senses the stalled ribosome and recruits the SKI complex composed by Ski2, Ski3, Ski8 that recruits in turn the exosome. A less efficient non-Ski7 dependent pathway involves sequential decapping and degradation by Xrn1 exonuclease.

## 1.4 The Pan2-Pan3 complex

### 1.4.1 Overview of the Pan2-Pan3 complex

In eukaryotic cells, deadenylation is a process with biphasic kinetics (Yamashita et al., 2005). The first phase is a slow deadenylation led by Pan2-Pan3 complex that shortens poly(A) tail to approximately 110 nucleotides. The second step involves the Ccr4-Not complex that shortens the poly(A) tail to 10 nucleotides. The poly(A)-specific nuclease (PAN) complex was firstly identified in *S. cerevisiae* (Sachs and Deardorff, 1992). A 3' to 5' exonuclease activity that degraded the poly(A) but not the mRNA body was found, and was dependent on two proteins: Pan2 and Pan3. The stoichiometry of the complex is still not clear due to conflicting findings. These two proteins were considered to form a 1:1 complex; however, Pan3 dimerisation evidence was obtained in yeast two-hybrid experiments and in structural studies suggesting a 1:2 complex (Figure 1.8 B) (Sachs and Deardorff, 1992; Mangus et al., 2004; Christie et al., 2013; Wolf et al., 2014). Early works showed that Pan2-Pan3 nuclease activity is poly(A)-specific and magnesium dependent, requiring a 3'-OH group and releasing 5'AMP as product of reaction (Lowell et al., 1992; Uchida et al., 2004). Moreover, the deadenylase activity of the complex depends on the cytoplasmic poly(A) binding protein PABPC in mammals or its orthologue Pab1p in yeast (Wahle and Winkler, 2013). Pan2-Pan3 are only active on a poly(A)-PABP complex being unable to remove the final 20-25 nucleotides of the poly(A) tail (Lowell et al., 1992; Uchida et al., 2004; Wolf and Passmore, 2014). The deadenylation by Pan2-Pan3 is the first and limiting step of mRNA degradation; the complex removes the poly(A) tail until the PABP can no longer bind (Decker and Parker, 1993; Brown and Sachs, 1998).

### 1.4.2 Pan2

Poly(A) Specific Ribonuclease Subunit Homolog 2 (PAN2), also known as Ubiquitin Specific Peptidase 52 (UPS52), was firstly identified in *S. cerevisiae* as a 127 kDa nuclease (Boeck et al., 1996b). Pan2 is the subunit designated to the catalytic activity; it belongs to the DEDD-type exonuclease family containing a DEDD/RNaseD domain in its C-terminus. In addition to this catalytic domain, Pan2 C-terminus contains an Ubiquitin C-terminal hydrolase domain predicted to be inactive, lacking active site residues (UCH-like) (Quesada et al., 2004). Pan2 N-terminus contains a WD40 domain. These domains usually mediate protein-protein interactions, but no interacting partner for Pan2 has been identified yet. Moreover, between the UCH like and WD40 domains a linker region was identified as the domain responsible for the interaction with Pan3 (PID/linker) (Figure 1.8 A) (Wolf et al., 2014). Point mutation in Asp 1083, one of the amino acid residues of Pan2 DEDD domain responsible for the coordination with the two divalent metal ions, inactivates the catalytic activity (Lowell et al., 1992; Uchida et al., 2004). Pan2 was demonstrated to be unable to bind to RNA when isolated and a very low deadenylation activity was observed in the absence of Pan3 (Wolf et al., 2014).

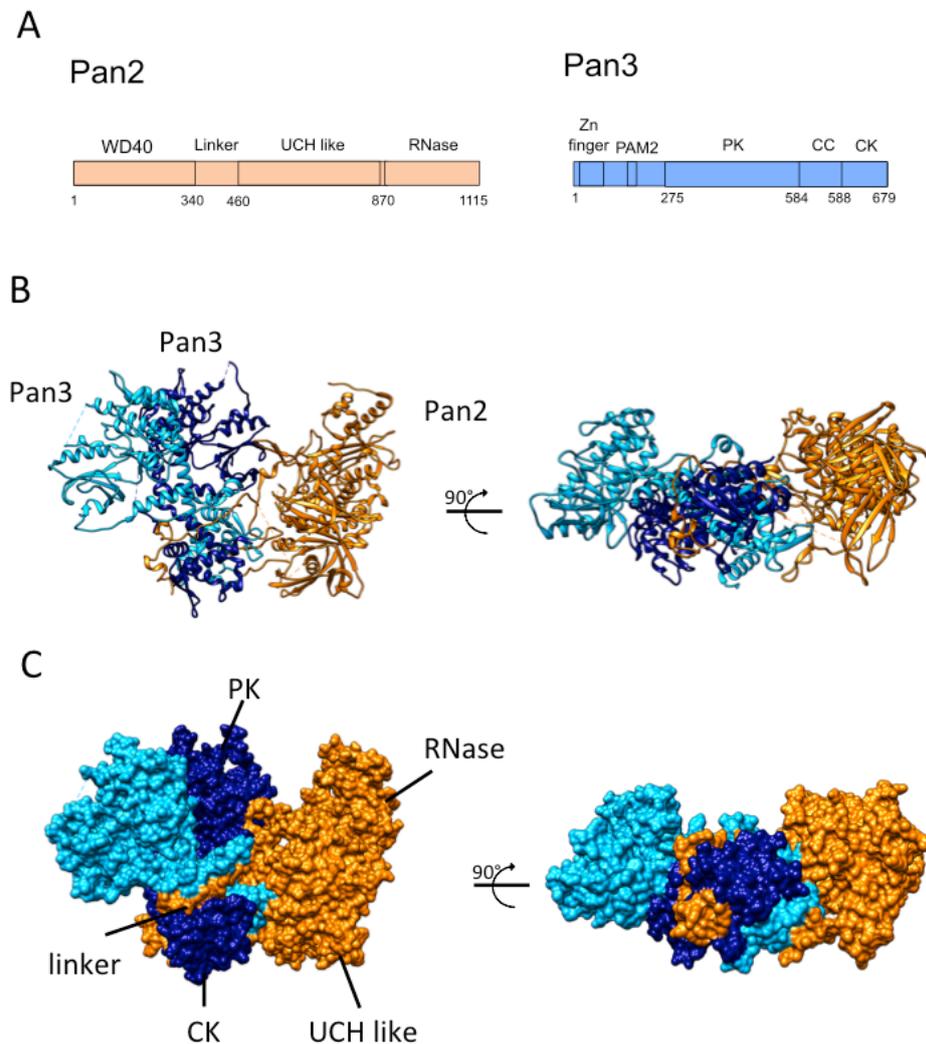
### 1.4.3 Pan3

Pan3 is a 76 kDa protein that interacts with Pan2 and can form an homodimer (Brown et al., 1996; Christie et al., 2013). Pan3 presents a highly conserved C-terminal domain, a pseudokinase domain and a N-terminal zinc finger. Moreover, a linker region between the zinc finger and the pseudokinase domains contain a PABP-interacting motif 2 (PAM2) that binds to the C-terminus of poly(A) binding proteins (Siddiqui et al., 2007). The C-terminal domain is involved in the interaction with Pan2 and present coiled coil (CC) and C-terminal knob (CK) domains (Figure 1.8 A) (Schäfer et al., 2014). The N-terminal zinc finger confers specificity of the complex to poly(A) RNA preferentially over other polyribonucleotides. The C-terminal pseudokinase domain also binds to RNA, but not in a poly(A) specific manner. Moreover, the PAM2 motif, in addition to its role

in the interaction with PABP proteins, can act synergistically with the zinc finger and the pseudokinase domain to channel the poly(A) tail into Pan2 active site (Wolf et al., 2014; Schäfer et al., 2014).

Pan3 is the non-catalytic subunit of the complex and is responsible for the interaction with other proteins. The major recruitment of the Pan2-Pan3 complex occurs via the Pan3-PABP interaction. Crystal structure of the yeast Pan2-Pan3 core trimeric complex showed how Pan2 RNase D domain latches onto Pan3 and a linker extended region anchors the protein on the homodimer (PDB accession number 4XR7) (Figure 1.8 B and C) (Schäfer et al., 2014). Phosphorylation of a cluster of serine and threonine residues adjacent the PAM2 motif may regulate this interaction (Siddiqui et al., 2007; Huang et al., 2013). Pan2-Pan3 complex is involved in the miRNA-mediated deadenylation. Recruitment of the complex can occur by interaction of Pan3 with GW182/TRNC6 via PABP that acts as bridge (Chekulaeva et al., 2011; Fabian et al., 2012). Furthermore, studies in *Caenorhabditis elegans* found the direct interaction between the GW182/AIN-1 and the Pan3 pseudokinase domain (Kuzuoğlu-Öztürk et al., 2012).

The recruitment of the Pan2-Pan3 deadenylase complex was described in response to replication stress. Interestingly, Pan3 interaction with the Dun1 kinase was found in yeast. These proteins function together to regulate the mRNA of Rad5, a protein involved in error-free post-replication repair pathways (Hammet et al., 2002). These mechanisms allow the recruitment of the Pan2-Pan3 complex on specific mRNAs. The Pan2-Pan3 poly(A) shortening is completed by the recruitment of the Ccr4-Not complex in the second phase of the deadenylation process.



**Figure 1.8 Overview of Pan2-Pan3 deadenylase complex. (A)** Schematic representation of the Pan2 and Pan3 proteins. Pan2 is a DEDD-type exonuclease containing the DEED/RNase domain in its C-terminus. Pan2 contains an inactive ubiquitin C-terminal hydrolase-like (UCH-like) domain and a WD40 domain in the N-terminal. A linker region between the WD40 and UCH-like domain is responsible for the interaction with Pan3. Pan3 presents a conserved C-terminus containing coiled coil (CC) and C-terminal knob (CK) domains, a pseudokinase, a PAM2, and a zinc-finger. **(B)** Crystal structure of the yeast Pan2-Pan3 core complex (PDB accession number 4XR7). **(C)** Surface visualization of the Pan2 (orange)-Pan3 (blue and light blue) core complex. Indicated are the RNase domain, UCH like domain, and the linker region of Pan2 and the pseudokinase domain (PK) and the C-terminal knob domain (CK) of Pan3. Structures were visualized using the UCSF Chimera package.

## 1.5 The Ccr4-Not complex overview

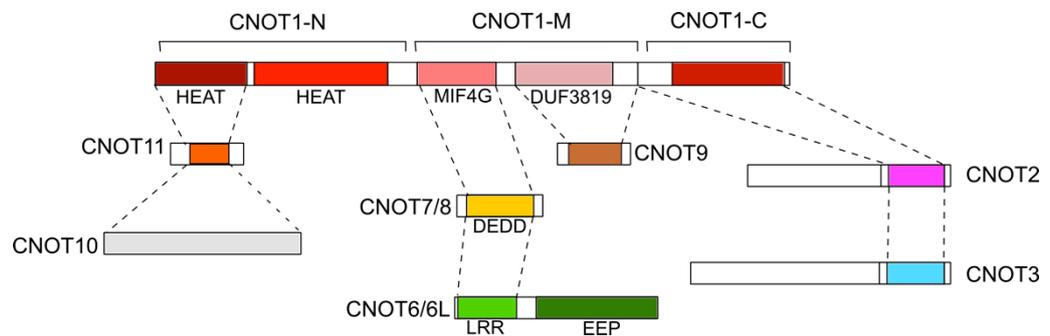
The Ccr4-Not complex is composed by catalytic and non-catalytic subunits. Two are the components associated with its deadenylase activity: a Ccr4 subunit containing an EEP domain, and a Caf1 component, characterised by a DEDD domain. In addition to these deadenylase subunits, several components have been identified to assemble in the complex (Figure 1.9). NOT1, NOT2, NOT3 and NOT4 were firstly identified in *S. cerevisiae* (Collart and Struhl, 1994). In yeasts, this complex exists in two forms of 1.0 and 1.9 MDa, and consists of five NOT proteins (Not1-Not5), the two catalytic enzymes Caf1 and Ccr4, and additional subunits Caf40 and the yeast specific Caf130 (Liu et al., 1997, 1998; Chen et al., 2001).

Orthologues for most of the yeast Ccr4-Not complex components were identified in human. (Albert et al., 2000). Two deadenylase enzymes, Caf1a (CNOT7) or Caf1b (CNOT8) and Ccr4a (CNOT6) or Ccr4b (CNOT6L), and several non-catalytic subunits assemble in the Ccr4-NOT complex with an estimated molecular weight of 1.2 MDa (Winkler and Balacco, 2013). CNOT3 is the only human homologue for both yeast Not3 and Not5. Moreover, Not4 human orthologue, CNOT4, resides in a separate complex with a molecular weight of 200 kDa (Albert et al., 2000; Lau et al., 2009).

### 1.5.1 The Ccr4 subunit

The Carbon catabolite repression (*ccr*) 4 was originally identified in yeast as regulatory of the alcohol dehydrogenase II (ADHII) (Denis, 1984). This catalytic subunit is characterised by the presence of a leucine-rich repeat (LRR) domain in its N-terminus and an endonuclease-exonuclease phosphatase (EEP) catalytic domain in its C-terminus. The yeast Ccr4 presents an additional N-terminal glu-rich domain that interacts intramolecularly with the LRR domain. The EEP domain is associated to its Mg<sup>2+</sup> dependent ribonuclease activity and deadenylase activity assays showed that the subunit is strictly poly(A) specific (Chen et al., 2002; Viswanathan et al., 2003; Wang et al., 2010). Structural studies of the Ccr4 subunit revealed that the C-terminus is very similar to that of apurinic/aprimidinic endonuclease (APE) 1 forming a  $\alpha / \beta$ -sandwich

fold (Wang et al., 2010).



**Figure 1.9 Architecture of the CCR4-Not complex.** Schematic overview of the modular architecture of the human Ccr4-Not complex (adapted from Bawankar et al., 2013). The CNOT10/CNOT11 module is anchored on the amino terminal domain of CNOT1 (CNOT1-N), the deadenylase module, composed by CNOT6/CNOT6L and CNOT7/CNOT8 is bound to the MIF4G domain in the middle region of CNOT1 (CNOT1-M). CNOT9 interacts with the DUF3819 domain of CNOT1-M. The NOT module composed by CNOT2 and CNOT3 is anchored on the CNOT1 C-terminal domain (CNOT1-C). For each protein the domains are indicated and coloured differently.

The two  $Mg^{2+}$  ions, required for the hydrolysis, are coordinated to an asparagine, glutamate, two aspartates and a histidine residue and each of these residues is required for the enzymatic activity. In *H. sapiens* the residues coordinating the divalent ions are Asn195, Glu240, Asp410, Asp489 and His529 (Chen et al., 2002; Wang et al., 2010). The amino LRR repeats domain is composed of alternating of  $\alpha$ -helices and  $\beta$ -sheets. This domain is responsible for the interaction with the other catalytic subunits of the complex, Caf1 (Malvar et al., 1992; Dupressoir et al., 2001; Clark et al., 2004; Basquin et al., 2012). This subunit is highly conserved in eukaryotes and vertebrates present two paralogues named CNOT6 and CNOT6L (Dupressoir et al., 2001; Morita et al., 2007; Cooke et al., 2010). The two paralogues are 78% identical and 88% similar at amino acid level (Winkler and Balacco, 2013). Both paralogues can associate with the Ccr4-Not complex, but cannot coexist in the same complex (Lau et al., 2009; Wahle and Winkler, 2013). Interestingly, plants such as *Oryza sativa* and *Arabidopsis thaliana* present multiple homologues, none of which contain the typical LRR amino terminal domain, suggesting a fundamentally different architecture (Figure 1.10 C) (Winkler and Balacco, 2013).

### 1.5.2 The Caf1 subunit

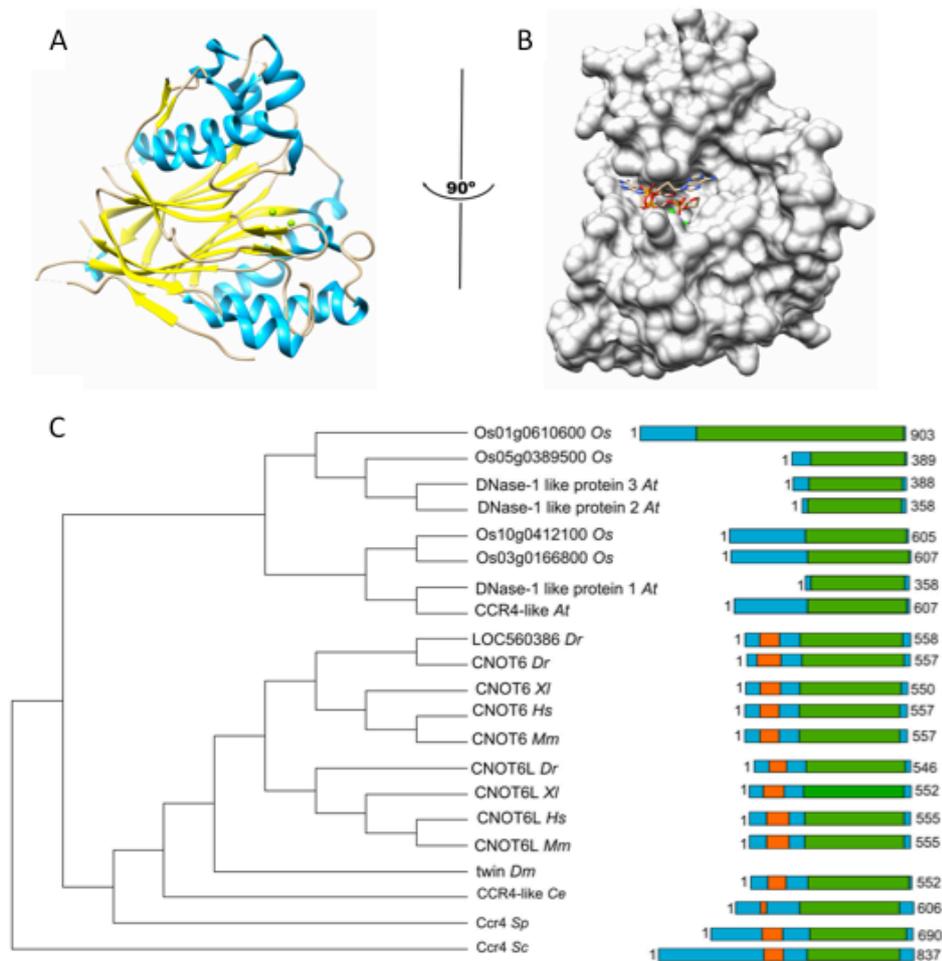
The Ccr4-associated factor (Caf) 1 was originally identified in *S. cerevisiae* (Sakai et al., 1992; Draper et al., 1995). The Caf1 subunit is characterised by the presence of an RNase D domain, which belongs to the DEDD (Asp-Glu-Asp-Asp) type exonuclease family. A single glutamate and three aspartate residues coordinate two  $Mg^{2+}$  ions and substitution of any of these residues abolishes the catalytic activity (Thore et al., 2003; Jonstrup et al., 2007; Horiuchi et al., 2009). Yeast Caf1 is unusual compared to its orthologues. In its N-terminus, it contains a longer extension, the metal-binding region contains a non-canonical sequence and its ribonuclease activity is not specific for the poly(A) (Thore et al., 2003).

*Saccharomyces pombe* Caf1 homologue has highly conserved active site residues and its activity is specific for the poly(A) as in *H. sapiens* orthologue (Bianchin et al., 2005; Jonstrup et al., 2007; Horiuchi et al.,

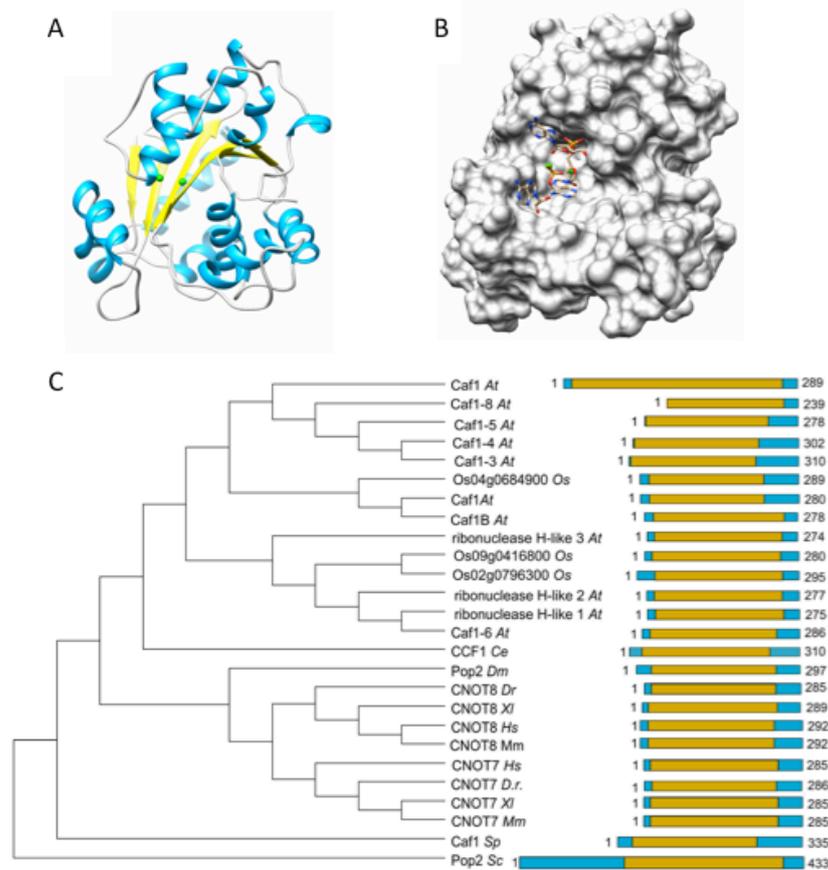
2009). Caf1 crystal structures of the *S. cerevisiae* Caf1 were determined showing a kidney-shaped subunit (Jonstrup et al., 2007; Andersen et al., 2009). Compared to the Ccr4 poly(A) binding pocket, the Caf1 binding site has a different shape and is wider (Figure 1.11 A). Crystal structures of the deadenylase subunit complexed with the RNA are not available. A model for the recognition of the RNA was proposed based on the superimposition of DEDD domains of Caf1 and PARN complexed with poly(A) leading to the identification of leucine for selectivity and serine for processivity (Figure 1.11 B) (Wu et al., 2005; Jonstrup et al., 2007; Andersen et al., 2009). Caf1 is conserved among eukaryotes and in vertebrates is present as two paralogues CNOT7 and CNOT8 (78% identity and 88% similarity at amino acid level). As for the case of the Ccr4 subunits CNOT6 and CNOT6L, the Caf1 homologous proteins are mutually exclusive. In *A. thaliana* and *O. sativa* there are multiple homologous of Caf1, suggesting a gene expansion in plants (Figure 1.11 C) (Winkler and Balacco, 2013; Chou et al., 2014).

### 1.5.3 The redundancy of the catalytic subunits

The gene duplication for the Caf1 and Ccr4 subunits in vertebrata suggests that the paralogues might have different roles in mRNA degradation. The human Caf1 paralogues, encoded by CNOT7 and CNOT8 have 76% identity and 89% similarity at amino acid level. The human Ccr4 paralogues encoded by CNOT6 and CNOT6L have 78% identity and 88% similarity. The surface of the active sites of both Caf1 and Ccr4 paralogues are highly conserved (Albert et al., 2000; Dupressoir et al., 2001; Winkler and Balacco, 2013). Redundant roles were described for CNOT7 and CNOT8, as well as CNOT6 and CNOT6L, but the proteins may have cell-type specific functions (Bianchin et al., 2005; Aslam et al., 2009; Lau et al., 2009; Mittal et al., 2011). CNOT6L and CNOT6 are required for the proliferation of human MCF7 breast carcinoma cells, but only CNOT6L is required for proliferation of mouse 3T3 fibroblast cells (Morita et al., 2007; Mittal et al., 2011).



**Figure 1.10 The Ccr4 subunit.** (A) Structure of the human Ccr4/CNOT6L catalytic domain (PDB accession number 3NGO). The nuclease domain forms an  $\alpha/\beta$  sandwich typical for the endonuclease-exonuclease-phosphatase (EEP) domain. The  $Mg^{2+}$  ions are indicated in green;  $\alpha$ -helical regions in blue, and  $\beta$ -sheets in yellow. (B) Binding of poly(A) by human Ccr4/CNOT6L. The  $Mg^{2+}$  ions located in the active site are indicated in green. (C) Evolutionary conservation of Ccr4 among *Saccharomyces cerevisiae* (Sc), *Schizosaccharomyces pombe* (Sp), *Caenorhabditis elegans* (Ce), *Drosophila melanogaster* (Dm), *Danio rerio* (Dr), *Xenopus laevis* (Xl), *Mus musculus* (Mm), *Homo sapiens* (Hs), *Arabidopsis thaliana* (At), and *Oryza sativa* (Os). The EEP domain is indicated in green; the leucine-rich repeat (LRR) domain is highlighted in orange. as originally published in Winkler and Balacco (2013) Heterogeneity and complexity within the nuclease module of the Ccr4-Not complex. Front Genet. 4:296. doi:10.3389/fgene.2013.0029. The figure was reproduced under the terms of the Creative Commons Attribution License, (CC-BY 3,0).



**Figure 1.11 The Caf1 nuclease component.** (A) Structure of the human Caf1/CNOT7 catalytic domain (PDB accession number 4GMJ). The Mg<sup>2+</sup> ions are indicated in green,  $\alpha$ -helical regions in blue, and  $\beta$ -sheets in yellow. (B) Model of poly(A) binding by human Caf1/CNOT7. The model was derived from superposition the structure of the PARN deadenylase in complex with RNA (PDB accession number 2A1R). The Mg<sup>2+</sup> ions located in the active site are indicated in green. (C) Evolutionary conservation of Caf1 in *Saccharomyces cerevisiae* (Sc) Caf1 were selected from the following species: *Schizosaccharomyces pombe* (Sp), *Caenorhabditis elegans* (Ce), *Drosophila melanogaster* (Dm), *Danio rerio* (Dr), *Xenopus laevis* (XI), *Mus musculus* (Mm), *Homo sapiens* (Hs), *Arabidopsis thaliana* (At), and *Oryza sativa* (Os). The DEDD domain is highlighted in yellow. Figure as originally published in Winkler and Balacco (2013)

Heterogeneity and complexity within the nuclease module of the Ccr4-Not complex. *Front Genet.* 4:296. doi: 10.3389/fgene.2013.0029. The figure was reproduced under the terms of the Creative Commons Attribution License, (CC-BY 3.0).

#### 1.5.4 The non-catalytic subunits

The Ccr4-Not complex is a multi-subunit complex composed by Not1, Not2, Not3, Not5, Caf40, Ccr4, and Caf1. Not4 is a core component of the complex in *S. cerevisiae*, is conserved in eukaryotes but is not a core component of the human complex (Bhaskar et al., 2015). Moreover, in *H. sapiens* the species specific CNOT10 and CNOT11 are components of the complex (Villanyi and Collart, 2015). Several modules are anchored on the large CNOT1 subunit that act as a central scaffold of the complex and tether the Ccr4-Not complex to mRNAs (Bai et al., 1999). The nuclease sub-complex composed by CCr4 and Caf1 subunits are anchored on the central CNOT1 MIF4G domain. This domain is composed by several  $\alpha$ -helices and binds directly to Caf1 (Bai et al., 1999; Basquin et al., 2012; Petit et al., 2012). The interaction between Caf1 and Ccr4 via the LRR domain is essential for the stable assembly of Ccr4 on the complex (Figure 1.12) (Draper et al., 1995; Dupressoir et al., 2001; Mittal et al., 2011; Fromm et al., 2012). In vertebrates, the presence of two paralogues of the deadenylase subunits suggest that the Ccr4-Not complex is heterogeneous because the nuclease module can contain a single Caf1 and Ccr4 subunit bound to CNOT1 and may result in the formation of four highly related complexes (Winkler and Balacco, 2013). The amino-terminus of CNOT1 contains HEAT repeats and anchors the CNOT11/CNOT10 module (Basquin et al., 2012; Bawankar et al., 2013; Mauxion et al., 2013). The central region of CNOT1 contains a DUF3819 domain that interacts with the RQCD1/RCD1/Caf40/CNOT9 component. This subunit binds to single and double stranded nucleic acids with an affinity for sequences containing G/C/T greater than for oligo(A) (Garces et al., 2007; Bawankar et al., 2013).

The C-terminus of CNOT1 anchors the NOT-module components CNOT2 and CNOT3. These two proteins contain a conserved region in their C-terminus responsible for their interaction named Not-box. The C-terminus also present a region responsible for the interaction with CNOT1 and termed Not Anchoring Region (NAR) (Figure 1.12 A) (Boland et al., 2013). The yeast Ccr4-Not complex contains the highly related

subunits Not2, and the duplicated subunits Not3 and Not5 (Liu et al., 1998; Oberholzer and Collart, 1998; Wahle and Winkler, 2013). Crystal structure of the Not module revealed that the module forms a surface that serves as platform for molecular interactions and contains tracts of positively charged residues on the surface able to bind to poly(U) RNA *in vitro* (Bhaskar et al., 2013; Boland et al., 2013).

The overall structural organisation of the complex was resolved by electron microscopy and cryo-electron microscopy 3D reconstruction of the yeast CCR4-NOT complex indicates a L-shaped structure (Nasertorabi et al., 2011; Ukleja et al., 2016). The crystal structures of the yeast nuclease module and human Caf1 interacting with the MIF4G domain of CNOT1 confirm that Ccr4 interacts with Caf1 via the LRR domain and associate with Not1 independently from the other proteins (Figure 1.12 B, C and D) (PDB accession number 4B8C) (Draper et al., 1995; Bai et al., 1999; Basquin et al., 2012; Petit et al., 2012).

Not4 is a functional ubiquitin ligase and is not considered a conserved subunit of the Ccr4-Not complex in mammals (Albert et al., 2002; Temme et al., 2004; Lau et al., 2009). Functional studies of Not4 do not support a role in the mRNA degradation, but evidence suggests its involvement in the degradation of the products of "non-stop" translation. It can be speculated that Not4 ubiquitylates non-stop products and may link their degradation with the nonstop mRNA decay (NMD) (Dimitrova et al., 2009; Wahle and Winkler, 2013).

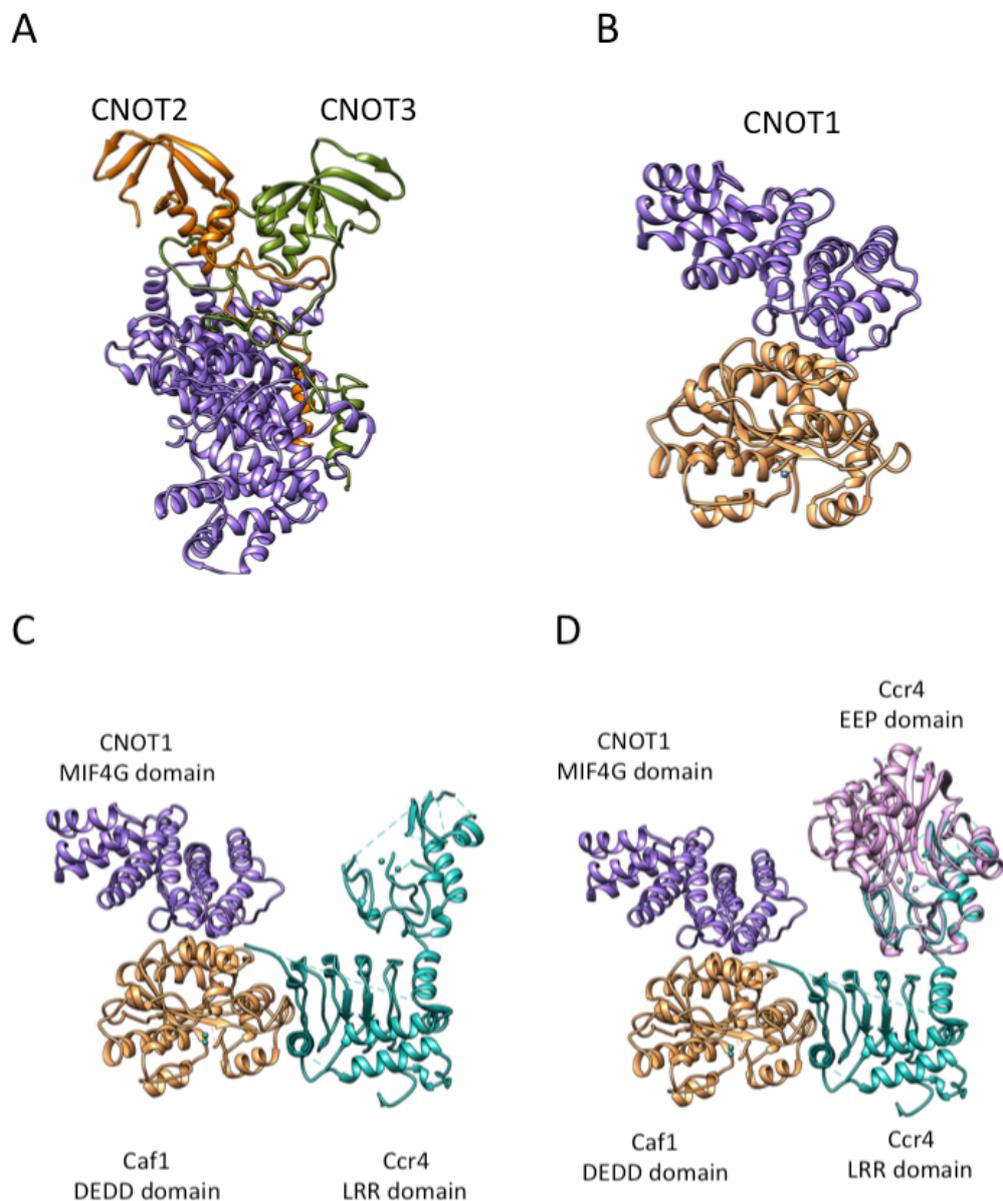
### **1.5.5 The additional roles of the Ccr4-Not complex**

In addition to its role in the mRNA decay pathways, the Ccr4-Not complex was found to be involved in the regulation of the acetylation status of histones H3 and H4. The chromatin status is regulated by the yeast homologues of Not3/Not5 and the yeast Not4 subunit. Not5 alter the nucleosome acetylation ability of the histone acetyltransferase (HAT) Gcn5, whereas Not4 is responsible for the turnover of the Jhd2 histone H3 demethylase (Laribee et al., 2007; Mulder et al., 2007; Peng et al., 2008; Mersman et al., 2009). Moreover, the yeast Not5 subunit plays an

important role in translation. It is required to localise the co-chaperone R2TP to ribosomes for an efficient synthesis of the Rpb1 subunit or RNAPII (Boulon et al., 2012; Villanyi et al., 2014).

The Ccr4-Not complex is involved in the regulation of the transcription initiation. The interaction of the deadenylase complex with transcription factors represses their ability to activate the transcription. Interactions of the human CNOT1 subunit with the retinoid acid X receptor in a ligand dependent manner and the interaction of the murine Caf40 with the transcriptional activators c-Myb and AP-1 were found (Haas et al., 2004; Winkler et al., 2006). Moreover, the ubiquitination of yeast Yap1 and mammalian PAF complex transcription activator by Not4 was observed (Gulshan et al., 2012; Sun et al., 2015). The Ccr4-Not complex is involved in the maturation of mRNPs, which is coupled to their export in the cytoplasm. The role of the deadenylase complex is achieved by interacting with poly(A) binding proteins such Nab2 or Hbp1 of the THO complex, or the Mlp1 of the nuclear pore complex (NPC) (Kerr et al., 2011). RNA binding proteins (RBP) play a pivotal role in recruiting the Ccr4-Not complex. The interaction between the RBP Mpt5 via Caf1 activate translation repression in yeast cells lacking Ccr4, the catalytically active subunit (Hook et al., 2007). Caf1 was shown to repress translation in *D. melanogaster* and *Xenopus laevis* via RBPs. CNOT1, recruited on the mRNA via Caf1-RBP interaction, recruits in turn the eIF4E transporter protein 4E-T that links the mRNA decay machinery to the 5' cap by interacting with eIF4E (Nishimura et al., 2015; Ozgur et al., 2015b; Waghray et al., 2015). The Ccr4-Not complex regulates various physiological processes. It was observed that the complex is involved in cell growth and survival (Ito et al., 2011a; Morita et al., 2007; Mittal et al., 2011; Fabian et al., 2012). Analyses of CNOT7(-/-) mice showed that the deadenylase subunit has a role in regulation of bone mass (Washio-Oikawa et al., 2007). Studies demonstrated that depletion of CNOT1 compromises deadenylase activity and induces apoptosis of HeLa cells (Ito et al., 2011b). Moreover, CNOT3(+/-) mice resulted to be obesity resistant

highlighting a role of the protein in the metabolic regulation (Morita et al., 2011).



**Figure 1.12 Crystal structures of the NOT-module and the nuclease module.** **(A)** Crystal structure of the NOT-module (PDB accession number 4C0D). Indicated are CNOT1 (purple), CNOT2 (orange) and CNOT3 (green). **(B)** Crystal structure of the CNOT1 MIF4G domain (purple) and Caf1/CNOT7 (sandy brown) (PDB accession number 4B8C). **(C)** Crystal structure of the yeast CNOT1 MIF4G domain (purple) associated with Caf1 DEDD domain (sandy brown) and Ccr4 LRR domain (sky blue) (PDB accession number 4B8C). **(D)** Structure of the nuclease sub-complex.

Indicated are the CNOT1 MIF4G domain (purple), the Ccr4 EEP domain (pink), the Ccr4 LRR domain (sky blue), and the Caf1 DEDD domain (sandy brown). The model was generated by alignment of the crystal structure of the MIF4G domain of human CNOT1 in complex with CNOT7 (PDB accession number 4GMJ) and the EEP nuclease domain of human CNOT6L (PDB accession number 3NGO) on the structure of the yeast Not1-Caf-Ccr4 complex (PDB accession number 4B8C).

## 1.6 The BTG/TOB proteins

The BTG/TOB proteins are the best characterised proteins that interact with the vertebrates Caf1 subunit (Mauxion et al., 2009; Winkler, 2010). Unicellular organisms do not present genes encoding for any BTG/TOB proteins, whereas proteins are present in *C. elegans* (1), *D. melanogaster* (2), mouse (6) and human cells (6) (Winkler and Balacco, 2013). In human cells, the BTG/TOB family consists of six members: BTG1, BTG2/PC3/Tis21, BTG3/Ana, BTG4/PCRB, TOB/Tob1, and Tob2 (Winkler, 2010). The proteins present a conserved N-terminal domain, termed BTG domain, which covers 104-106 amino acids responsible for the majority of the BTG/TOB protein interactions. In humans, all six members of the BTG/TOB family, apart from BTG4, have been shown to interact with Caf1a CNOT7 and CNOT8 through the BTG domain (Prévôt et al., 2001; Horiuchi et al., 2009). BTG4 is expected to interact with the deadenylase subunit due to the high similarity of its N-terminus with the other protein family members. Sequence alignment showed that TOB1 and TOB2, as well as BTG1 and BTG2, are highly similar, whereas BTG3 and BTG4 are more distant.

Moreover, in their C-terminus, Tob1 and Tob2 contain a highly conserved PAM2 motif, responsible for the interaction with PABP. Tob1 can bind at the same time to PABP and Caf1, suggesting that TOB1 can act as an adaptor for the recruitment of the Ccr4-Not complex (Ezzeddine et al., 2007). Furthermore, studies showed that Tob1 regulates the recruitment of the Ccr4-Not complex by binding the RBP cytoplasmic polyadenylation element-binding protein 3 (CPEB3) via its C-terminus, resulting in mRNA destabilisation (Ezzeddine et al., 2007; Hosoda et al., 2011). BTG1 and BTG2 contain a shorter C-terminus and interact with Caf1 via the BTG domain. Crystal structures were resolved for the human BTG2 (Figure 1.13 A), the mouse paralogue TIS21, and TOB1 in complex with the CNOT7 subunit (Yang et al., 2008; Horiuchi et al., 2009).

The analyses of the crystal structure of TOB1-CNOT7 (PDB accession number 2Z15) indicated that the association of the two proteins is mediated by two regions named Box A and Box B (Horiuchi et al., 2009).

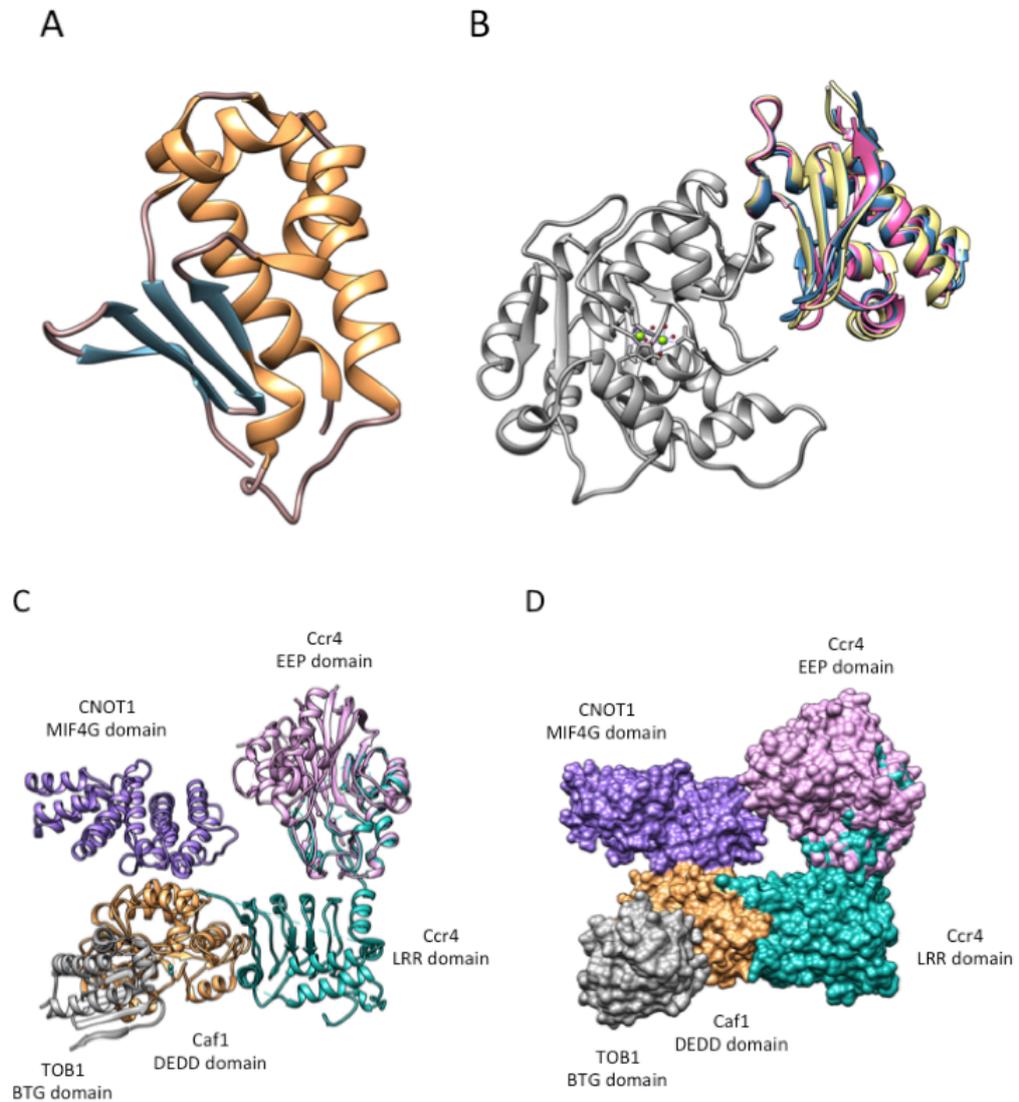
BTG1 and BTG2 have an additional region named Box C as shown in the human BTG2 crystal structure (Figure 1.13 A; PDB accession 3DJU) (Yang et al., 2008). Residues in both Box A and Box B of BTG2 were demonstrated to be essential for the interaction with CNOT7 by mutagenesis analyses (Yang et al., 2008; Horiuchi et al., 2009). Superimposition of the BTG2 and Caf1-Tob1 crystals showed a high conservation of the BTG domain between BTG2 and TOB1 (Figure 1.13 B). Interaction of the BTG2/TOB protein with CNOT7 recruits the Ccr4-Not complex. Figure 1.13 C and D shows the model of the nuclease sub-complex module in association with BTG2, obtained by superimposition of the crystal structures.

### 1.6.1 The role of the BTG/TOB proteins

The members of the BTG/TOB protein family share a conserved anti-proliferative activity (Matsuda et al., 2001; Tirone, 2001; Winkler, 2010). *In vitro* analyses indicated that BTG2 and TOB1 inhibit the activity of purified Caf1, although BTG is a general activator of the mRNA decay. Moreover the BTG domain of TOB1 does not affect the deadenylase activity (Mauxion et al., 2008; Miyasaka et al., 2008; Yang et al., 2008; Horiuchi et al., 2009).

Functional studies based on reporter genes showed that BTG2, TOB1, and TOB2 promote deadenylation and mRNA degradation (Ezzeddine et al., 2007; Mauxion et al., 2008). In addition to their role in deadenylation, BTG/TOB proteins act as transcriptional regulator of the gene expression by interacting with DNA-binding transcription factors (Matsuda et al., 2001; Tirone, 2001; Winkler, 2010). The BTG/TOB proteins are important for the prevention of tumorigenesis. For instance, BTG2 is a direct target of the tumour suppressor p53 and Tob1 has a critical role in Ras-mediated transformation (Rouault et al., 1996; Suzuki et al., 2002; Boiko et al., 2006). Mice Tob(-/-) developed liver tumour to a higher rate than wild type, showing increased cyclin D1 mRNA levels (Yoshida et al., 2003). Moreover, reduced level of BTG/TOB proteins were found in tumours. Tob1 levels are reduced and induction of its expression by peritoneal

injection suppressed pancreatic cancer peritonitis (Yanagie et al., 2009). A reduced expression of BTG3 was found in lung cancer cell lines and BTG2 reduced levels were found in kidney and breast carcinomas (Boiko et al., 2006; Yoneda et al., 2009). BTG1 and BTG2 were found mutated in non-Hodgkin lymphomas (Morin et al., 2011; Waanders et al., 2012).



**Figure 1.13 Models of the BTG/TOB protein family member. (A)** Model of the human BTG2 (PDB accession 3DJU). In sandy brown are highlighted the  $\alpha$ -helices and in blue the  $\beta$ -sheets. **(B)** Model of the Caf1-BTG2 complex. The model was generated by alignment of the Caf1 (gray) in complex with the BTG domain of Tob1 (pink; PDB accession number 2D5R), Tob1 BTG domain (blue; PDB accession number 2Z15), and the human BTG2 (yellow, PDB accession number 3DJU). **(C-D)** Structure and surface of the model of the nuclease sub-complex in association with the BTG domain of Tob1. Indicated are the CNOT1 MIF4G domain (purple), Caf1 DEDD domain (sandy brown), Ccr4 LRR domain (sky blue), Ccr4 EEP domain (pink) and TOB1 BTG domain (gray). The model was obtained by alignment of the Tob-Caf1 structure, the MIF4G

domain in complex with CNOT7) (PDB accession number 4GMJ), and the EEP nuclease domain of the human CNOT6L (PDB accession number 3NGO) on the structure of the yeast Not1-Caf1-Ccr4 complex (PDB accession number 4B8C).

## 1.7 Aims of the study

The Ccr4-Not complex plays a critical role in the regulation of the gene expression. The complex is the major deadenylase enzyme involved in the shortening of the poly(A) tail and is composed of two catalytic subunits, Ccr4 (CNOT6/CNOT6L) and Caf1 (CNOT7/CNOT8), and at least six non-catalytic subunits organised in modules (Wahle and Winkler, 2013). Regulatory proteins interact with the deadenylase complex tethering it on specific mRNAs activating mRNA degradation pathways and down-regulating protein expression (Goldstrohm et al., 2006, 2007; Sandler and Stoecklin, 2008; Joly et al., 2013; Raisch et al., 2016)

The main aim of this study is to understand in more detail the role of the subunits of Ccr4-Not complex in the recruitment of specific mRNAs and in the maintenance of the integrity the complex assembly, pivotal for its function. Specifically:

1. The study aims to discover and explore the interactome of the catalytic subunits Caf1 and Ccr4 to detect new interacting partners that may recruit the deadenylase complex on specific mRNAs.
2. The study aims to understand whether the function or structure of the Ccr4-Not complex is compromised in cancer by assessing the structure and function of Ccr4-Not variants found in various types of cancer. Specifically, this would also lead to the identification of amino acid residues critical for the integrity and function of the components of the NOT-module (CNOT1, CNOT2 and CNOT3) and the nuclease sub-complex (CNOT7, CNOT6L, CNOT1, and the regulatory protein BTG1).
3. The study aims to investigate the variability of the Ccr4-Not components using phylogenetic methods to understand the presence of paralogous genes encoding the nuclease subunits, which are only found in vertebrates.

## 1.8 Experimental approaches

### 1.8.1 Stable Isotope Labelling using Amino acids in Cell culture (SILAC)-Rapid Immunoprecipitations and Mass spectrometry of Endogenous protein (RIME).

The discovery and exploration of the interactome of the deadenylase subunits of the complex Caf1 (CNOT7/CNOT8) and Ccr4 (CNOT6/CNOT6L) will be performed conjugating *in vitro* and *in silico* approaches. Stable Isotope Labelling Amino acid in Cell culture (SILAC)-Rapid Immunoprecipitations Mass Spectrometry of Endogenous protein (RIME), based on the *in vivo* labelling of endogenous proteins by incorporation of amino acids with substituted stable isotopic nuclei, coupled with RIME immunoprecipitations, will identify proteins directly or indirectly interacting with Caf1 and Ccr4. The software Cytoscape (Shannon et al., 2003) will be used for the generation of protein-protein interaction network and to explore the interactome obtained by SILAC-RIME. Cytoscape plug-ins MCODE (Bader et al., 2003) and ClusterONE (Zhang et al., 2014) will identify biological protein clusters allowing detection of the involvement of the Caf1 and Ccr4, hence the Ccr4-Not complex, in known and new biological pathways.

### 1.8.2 Identification of residues critical for the NOT-module assembly

The identification of residues critical for the assembly and integrity of the NOT-module will be approached with bioinformatic predictive methods. Mutations found in cancer, as annotated in the Catalogue Of Somatic Mutation found In Cancer (COSMIC) (Bamford et al., 2004; Forbes et al., 2008, 2011) and Ensemble (Chen et al., 2010; Flicek et al., 2014) databases, will be collected and investigated using the Sort Intolerant From Tolerant (SIFT) software (Ng and Henikoff, 2003) to point residues whose substitution can be deleterious for the function and assembly of the module. Chemical and thermodynamic properties, and chemical contacts between the residues involved in the interaction with the components of the Not-module will be investigated by the Proteins, Interfaces, Surfaces and Assemblies (PISA) (Krissinel and Henrick, 2007) software and

RINalyzer (Doncheva et al., 2011). This software will be used to construct a bi-dimensional map of the crystal structure to visualise it as a residue-residue interaction network using Cytoscape.

### **1.8.3 Analyses of sequence variants of the CNOT7 ribonuclease**

CNOT7 variants found in cancer will be studied using a similar approach to the identification of the residues critical for the assembly of the NOT-module. This investigation will identify residues critical for the deadenylase function of CNOT7 and its interaction with the other proteins of the nuclease module CNOT6L and CNOT1, and with the regulatory protein BTG1. The CNOT7 variants found in cancer will be collected from the COSMIC and Ensemble databases and their effect on the function of the protein will be predicted using SIFT and Disease-Susceptibility-based SAV Phenotype Prediction software (SuSPect) (Ng and Henikoff, 2003; Yates et al., 2014). The effect of the deleterious mutations on the stability of the protein will be investigated using Site Directed Mutator (Worth et al., 2011). UCSF Chimera will be used in order to investigate whether the mutations can be accommodated into the CNOT7 structure model without major rearrangements of the protein. Further investigation on the effects of the amino acid substitution will involve *in vitro* approaches. Yeast two hybrid assay will assess the impact of CNOT7 mutations on the interactions with the other protein by the generation of hybrid proteins GAL4-AD-CNOT7, GAL4-AD-CNOT7 variants, GAL4-BD-CNOT1<sub>MIF4G</sub>, GAL4-BD-CNOT6L, and GAL4-BD-BTG1.

RNA tethering assay will investigate the impact of the mutations on the activity of CNOT7 independently to its capacity of binding to the RNA.

### **1.8.4 Phylogenetic and variability analysis**

A phylogenetic reconstruction for each component of the Ccr4-Not complex will be performed using and comparing the Neighbour-Joining (NJ) and Maximum Likelihood (ML) methods.

Variability analysis will be performed to detect selective pressures acting on the proteins. To detect selective pressures, two kinds of models will be used: the site and branch-site models. The site model will allow the

parameter describing the selective pressure to vary among codons  
(Nielsen and Yang 1998; Yang, 2007))

. The branch-site model will allow selective pressures to vary both among codons in the protein and across branches of the tree, with the intent of detecting positive selection acting on few sites along particular lineages.

# **Chapter 2**

## **Materials and methods**

## Chapter 2. Materials and Methods

### 2.1 Bacterial growth and transformation

#### 2.1.1 Reagents, stock solutions and buffers

*Lysogeny broth (LB) medium and agar:* 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, pH 7.2 (NaOH), for LB agar add 15 g/L bacteriological agar. The medium was sterilised using a Prestige medical 2100 classic bench-top autoclave and stored at room temperature.

*Ampicillin 1000× stock solution:* 100 mg/mL in 50% ethanol/water. The solution was sterilised using 0.22 µm pore size filters. Stored at -20°C.

*Chloramphenicol 1000× stock solution:* 34 mg/L in 100% ethanol. The solution was sterilised using 0.22 µm pore size filters. Stored at -20°C.

*Kanamycin 1000× stock solution:* 50mg/L in H<sub>2</sub>O. The solution was sterilised using 0.22 µm pore size filters. Stored at -20°C.

*1M IPTG:* Isopropyl β-2-1-thiogalactopyranoside, in H<sub>2</sub>O. The solution was sterilised using 0.22 µm pore size filters. Stored at -20°C.

*X-Gal solution:* 5-bromo-4-chloro-indolyl-β-galactopyranoside (50mg/ml in dimethylformimide) was purchased from Promega. The solution was stored at -20°C.

### 2.1.2 Culture of *Escherichia coli*

*Escherichia coli* strain DH5 $\alpha$  and BL21-RIL-DE3 were used for different aims. The bacterial strain *E. coli* DH5 $\alpha$  was used for DNA manipulation *i.e.* cloning and mutagenesis; the bacterial strain *E. coli* BL21-RIL-DE3 was used for protein expression. *E. coli* was streaked onto agar plate containing the appropriate antibiotic and grown overnight at 37°C. The plates were stored at 4°C for few weeks. To prepare a liquid culture, a single colony was grown overnight in liquid LB medium containing the appropriate antibiotic, incubated in the shaker at 37°C, 200 rpm. For large scale cultures preparation glass conical flasks were used.

### 2.1.3 Preparation of *E. coli* competent cells

*E. coli* DH5 $\alpha$  and BL21-RIL-DE3 strains were made competent using magnesium chloride/calcium chloride. A single colony of *E. coli* from an Luria-Bertani agar plate was used to inoculate 3 ml of LB medium and grown at 37°C, 200 rpm over-night. The following day the culture was diluted 1:100 in LB medium and grown at 37°C, 200 rpm until the optical density at 600 nm (OD<sub>600</sub>) reached 0.5. The cells were then centrifuged at 4000 rpm for 10 minutes at 4°C. The supernatant was discarded and the pellet re-suspended in 50 ml of MgCl<sub>2</sub>. Cells were centrifuged, re-suspended in 50 ml CaCl<sub>2</sub> and incubated on ice for 20 min. Following another centrifugation, cells were re-suspended in 8.6ml of CaCl<sub>2</sub>. After adding 1.4 mL of glycerol, the cells were aliquoted in 0.5 ml tubes and snap frozen in liquid nitrogen. Competent cells were stored at -80°C.

### 2.1.4 Transformation of competent cells

*E. coli* competent cells, stored at -80°C, were thawed and mixed with DNA in pre-chilled round bottomed universal tubes and incubated on ice 5-20 minutes on ice. Less than 500 ng of DNA was used to transform 50-100  $\mu$ L of competent cells. Up to 7.5  $\mu$ L of ligation mix solution were used to transform 50  $\mu$ L of competent cells to obtain a good yield. The solution was heat shocked for 90 seconds in water bath at 42°C and then placed on ice for 2 minutes. 1 mL of LB medium was added to the cells followed by incubation at 37°C for 1 hour shaking. *E. coli* DH5 $\alpha$  transformed with

ligation mix were pelleted and re-suspended in 200  $\mu$ l. 100  $\mu$ l of the suspension were plated on LB agar plates with the appropriate selection antibiotic and then incubated at 37°C overnight.

## 2.2 Molecular Biology

### 2.2.1 Reagents, stocks solutions and buffers

*5× TBE*: 40 mM Tris base, 40 mM boric acid, 1 mM EDTA pH 8.0, stored at room temperature.

*6× Ficoll loading dye*: 16% Ficoll PM400 (purchased by Sigma-Aldrich) 0.2% orange G in H<sub>2</sub>O, stored at room temperature.

*Oligonucleotides*: Lyophilised primers were dissolved in TE to obtain a 100 mM solution and stored at -20°C (purchased by Sigma-Aldrich).

*TE buffer*: 10 mM Tris-HCl pH 8.0, 1 mM EDTA, stored at room temperature.

### 2.2.2 Small scale plasmid preparation

For the preparation of small scale of plasmid DNA, the Macherey-Nagel mini prep kit was used. A single bacterial colony was used to inoculate 3 to 5 mL of LB medium containing the appropriate antibiotic (ampicillin 100 µg/mL, chloramphenicol 34 µg/mL, and kanamycin 50 µg/mL) and grown overnight in an orbital shaker at 37°C, 200 rpm. All of the overnight culture was then used to extract the plasmid DNA according to the manufacturer's instruction. Plasmid DNA was eluted in 50 µl of elution buffer. The plasmid DNA concentration and purity was determined using Nanodrop ND-100 spectrophotometer before being stored at -20°C.

### 2.2.3 Large scale plasmid preparation

Large scale preparation of high quality plasmid DNA was carried out using the Sigma GenElute HP maxi prep kit. A single colony was used to inoculate 5 ml of LB medium containing the appropriate antibiotic (ampicillin 100 µg/mL, chloramphenicol 34 µg/mL, and kanamycin 50 µg/mL) and grown for 5-6 hours in an orbital shaker at 37°C, 200 rpm. All of the culture was then used to inoculate 200 mL of LB containing the

appropriate antibiotic and grown overnight in an orbital shaker at 37°C, 200 rpm. The overnight culture was pelleted by centrifugation at 4000 g for 15 minutes and plasmid DNA was extracted following the manufacturer's instruction. The DNA was eluted in 500 µL of elution buffer and stored at -20°C.

#### **2.2.4 Determination of DNA/RNA concentration**

Nucleic acids concentration and purity were determined using a Nanodrop ND-1000 UV-Vis Spectrophotometer (NanoDrop Technologies). The  $A_{260/280}$  ratio was used to estimate sample purity. A ratio between 1.8 and 2.0 was expected for DNA samples, while a ratio between 2.0 and 2.2 was considered acceptable for RNA samples.

#### **2.2.5 Agarose gel electrophoresis**

Depending on the size of the DNA to resolve, 0.5-2% agarose gel (w/v) was run in 0.5% TBE buffer. To the gel, 0.5 µg/ml of ethidium bromide was added. DNA samples were loaded in 1× Ficoll loading dye. DNA agarose gels were run at 80V for 45-90 minutes depending on the sample in 0.5X TBE buffer. Gels were visualised using transillumination with ultraviolet light and documented using Gel-Doc 2000 (Bio-Rad) and Bio-Rad Quantity One computer software.

#### **2.2.6 DNA extraction and purification from agarose gels**

To purify DNA from agarose gels, DNA bands were excised using a surgical scalpel. Purification was conducted following the manufacturer's instructions provided in the Macherey-Nagel Gel/PCR purification kit. For small DNA fragments (<500 bp) used in cloning, the DNA was eluted in 30 µl of pre-heated at 60°C elution buffer. To increase the DNA purified yield, two consecutive elutions with 15-20 µl were performed. For large fragments (>2 kb), the DNA was eluted with two separate elutions in 40 µl final volume.

### 2.2.7 Restriction enzyme digestion of DNA

Plasmid DNA was digested with restriction endonucleases to generate compatible ends for cloning and verify created plasmid. All restriction endonucleases and corresponding buffers were purchased from New England Biolabs Inc. (NEB) and used according the manufacturer's instructions (Table 2.1).

To verify newly created plasmids prior to sequencing, 100 ng of plasmid DNA was digested in 10  $\mu$ l with the appropriate restriction enzymes; to digest large quantities of DNA for cloning and ligations 1  $\mu$ g of DNA was digested in 10  $\mu$ l volume reactions. DNA was digested for 1-2 hours according to the amount of DNA and at the optimal temperature for the enzyme used.

### 2.2.8 Removal of 5' phosphate group from linearised plasmid DNA

In order to prevent linearised plasmids to self-ligate in ligation reactions, phosphate groups were removed from 5' ends using Antarctic Phosphatase. 1  $\mu$ g of plasmid DNA was treated with 1  $\mu$ l of the enzyme and incubated at 37°C for 30 minutes. A further 1  $\mu$ l of the enzyme was added to the sample and incubated at 37°C for more 30 minutes. The enzyme was then deactivated by incubation at 65°C for 5 minutes.

<b>Restriction Endonuclease</b>	<b>Restriction site 5' -&gt; 3'</b>
EcorI-HF	GAATTC
Sall-HF	GTCGAC
NheI-HF	GCTAGC
XhoI	CTCGAG
BamHI-HF	GGATCC
XbaI	TCTAGA

**Table 2.1 Restriction endonucleases.** All restriction endonucleases were purchased from New England Biolabs Inc. (NEB) and were used according to the manufacturer's instruction

### 2.2.9 Phosphorylation of blunt PCR products

Phosphorylation of blunt PCR products was performed using the T4 Polynucleotide Kinase (T4 PNK) before ligation. One microliter of the enzyme was added to the to the DNA and incubated at 37°C for 30 minutes.

### 2.2.10 Ligation of DNA fragments

DNA fragments and digested plasmid were ligated using the T4 DNA ligase in order to create recombinant plasmids. Required mass of the insert was calculated depending on the desired insert/vector molar ratio, the mass of the vector and the lengths of both plasmid and insert as shown in Formula 2.1.

*Required mass insert (g) = desired insert/vector molar ratio × mass of vector (g) × ratio of insert to vector lengths*

Formula 2.1

Ligation reactions were set up using 100 ng of the digested vector, the calculated mass of insert, 1 µl T4 DNA ligase and 1× T4 ligase buffer in 10 µl final volume and left overnight at room temperature.

### 2.2.11 Gibson cloning

In order to sub-clone DNA fragments in plasmids, the NEBuilder HiFi DNA Assembly Cloning Kit was used. Twenty nanograms of the DNA to insert and 30ng of digested vector DNA were added for total volume up to 1 µl, and 1 µl of NEBuilder HiFi DNA Assembly Master mix was added to the reaction. Total volume of the reaction was 3 µl. Samples were incubated in a thermocycler at 50°C for 20 minutes. Following incubation, samples were stored at -20°C or used for transformation. *E. coli* DH5α were transformed with 3 µl of the assembly reaction.

### 2.2.12 Polymerase Chain Reaction (PCR)

Polymerase Chain Reaction (PCR) was performed for the generation of DNA fragments for cloning and the screening of recombinant plasmids. The peqlab Primus 96 advanced thermal cycler was used.

For the amplification of DNA fragments to clone into vectors, Phusion High-Fidelity DNA polymerase was used. A standard reaction was set up as shown in Table 2.2. Samples were initially denatured at 98°C for 30 seconds followed by 25-30 cycles of a denaturing step of 98°C for 10 seconds, annealing step depending on the primers melting temperatures for 30 seconds, then elongation at 72°C for 30 sec/1 kb. A final elongation step at 72°C for 10 minutes was included at the end of the amplification. To check if the amplicons were the correct size, 10 µl of the reaction were run on agarose gel.

For the screening of newly generated plasmids prior to enzyme digestion and sequencing, colony PCR was performed. One single clone from a ligation mix was picked and used to inoculate 500 µl of LB media and grown for 4-6 hours in an orbital shaker at 37°C, 200 rpm. A standard Taq Polymerase PCR reaction was set up as shown in Table 2.3. Sample were initially denatured at 95°C for 3 minutes followed by 25 cycles consisting in a denaturing step at 95°C for 30 seconds, an annealing step at 55°C for 30 seconds and elongation at 68° 1 min/kb. A further elongation step at 68°C for 5 minutes was included. To check the amplicons size, 10 µl of the reaction was run on agarose gel.

Volume	Component/Reagent
x µL	100 ng DNA template
1.0 µL	10 µM forward primer (Sigma-Aldrich)
1.0 µL	10 µM reverse primer (Sigma-Aldrich)
10.0 µL	5× Phusion HF buffer (NEB)
1 µL	10 mM dNTP mix
0.5 µL	Phusion High-Fidelity DNA Polymerase (NEB)
Up to 50 µL	nuclease-free H <sub>2</sub> O

**Table 2.2 Standard Phusion High-Fidelity DNA polymerase reaction.**

Volume	Component/Reagent
x $\mu\text{L}$	variable DNA template
1.0 $\mu\text{L}$	10 $\mu\text{M}$ forward primer (Sigma-Aldrich)
1.0 $\mu\text{L}$	10 $\mu\text{M}$ reverse primer (Sigma-Aldrich)
5.0 $\mu\text{L}$	10 $\times$ Standard Taq reaction buffer (NEB)
1 $\mu\text{L}$	10 mM dNTP mix
0.25 $\mu\text{L}$	Taq DNA Polymerase (NEB)
Up to 50 $\mu\text{L}$	nuclease-free $\text{H}_2\text{O}$

**Table 2.3 Standard Taq DNA Polymerase reaction.**

### 2.2.13 Site-directed mutagenesis

Primers for the site-directed mutagenesis were designed using the NEBaseChanger™ v1.2.5 tool (<http://nebasechanger.neb.com/>) (Table 2.6 and Table 2.7) Reactions were set up using 1 ng template DNA. The standard protocol for Phusion DNA Polymerase was used. Following amplification, the PCR product was incubated with the DpnI enzyme in order to remove the template DNA. One microliter of DpnI was added directly to the 50  $\mu\text{l}$  of mutagenesis reaction and incubated at 37°C for 30 minutes. All the reaction volume was run on 0.8% agarose gel and the band excised for Gel purification and eluted in 30  $\mu\text{l}$ . Thereafter, 1  $\mu\text{l}$  of the T4 PNK and 1 $\times$  T4 reaction buffer were added to the sample and incubated at 37°C for 30 minutes for the phosphorylation of blunt ends. A further 1  $\mu\text{l}$  of T4 DNA ligase was added to the sample and incubated overnight at room temperature. The following day, 50 ng of the sample was used to transform 50  $\mu\text{l}$  *E. coli* DH5 $\alpha$  before DNA isolation and sequencing to confirm the presence of the desired mutation.

## 2.2.14 List of Primers

Primer name	Sequence 5'- 3'
Pan2 5UTR	GGTTGGGTCAGGCATCTAAA
Pan2 3UTR	GGTGGGCTAGGACCTTCAA
Pan3 5UTR	AGCGTCTTCCTTTCTCCC
Pan3 3UTR	CCTGTTCGTTTCCCAAATGT

Table 2.4 List of primers for amplification from cells

Primer name	Sequence 5'- 3'
YTHDF1 FW	AAAAAGGATCCGGGCGTGTGTTTCATCATCAAG
YTHDF1 RV	AAAAAGTCGACGATAATTTTCAGCACTTGCTTGGC
FW_CN0T1_EcoRI	AAAAAAGAATTCATGTTGGCCTGTCTGCAAGCTTG
RV_CN0T1_SalI	AAAAAAGTCGACCAGTGTCCAAGTCCACAGCCA
CN0T1_1020_1317_EcoRI_FW	AAAAAAGAATTCATGGCCCAGGCTCAGGCCAGGTC
CN0T1_1020_1317_SalI_RV	AAAAAAGTCGACGCAGAGAGTTGCTCATCTAAATT C
CN0T2_NehI_FW	AAAAAAGCTAGCATGGTGAGGACTGATGGACATAC
CN0T2_XhoI_RV	AAAAAACTCGAGTTAGAAGGCTTGCTGAGCAGG
pAD_CN0T7_XhoI_FW	AAAAAACTCGAGATGCCAGCGGCAACTGTAG
pAD_CN0T7_XhoI_RV	AAAAAACTCGAGTCATGACTGCTTGTTGGCTTC
pBD_CN0T6L_Eco_FW	AAAAAAGAATTCATGAGACTAATAGGGATG
pBD_CN0T6L_SalI	AAAAAAGTCGACCTACCTCCGATTAGGCAAG
CN0T3_XbaI_FW	AAAAAATCTAGAATGGCGGACAAGCGCAAAC
CN0T3_BamHI_RV	AAAAAAGGATCCTCACTGGAGGTCCCGGTC

Table 2.5 List of primers used for cloning

<b>Name</b>	<b>Sequence 5'- 3'</b>
CNOT7_N19I_RV2	TGGGCTTGCATCTTGGATGAAG
CNOT7_N19I_RV_3	AACTTCACAAATTCTTTGGC
CNOT7_M24I_FW_2	ATGAAGAGATTAAGAAAATTCGTCAAGTTAT
CNOT7_M24I_RV_2	CCAAGTTGCAAGCCCAAAC
CNOT7_R20C_FW	AGATCATAGCTAAAGAATTTGTGAAGTTTGG
CNOT7_R20C_RV	ACAGTTGCCGCTGGCATT
CNOT7_N19S_FW	TGGGCTTGCAGCTTGGATGAAG
CNOT7_N19S_RV	AACTTCACAAATTCTTTGGCTATG
CNOT7_R28C_FW	AGATCATAGCTAAAGAATTTGTGAAGTTTGG
CNOT7_R28C_RV	ACAGTTGCCGCTGGCATT
CNOT7_G45C_FW	GAGTTTCCAGTTGTGGTTGCAAG
CNOT7_G45C_RV	GGTGTCCATAGCAACGTA
CNOT7_A58V_FW_2	AGGAGCAATGTTGACTATCAATAC
CNOT7_A58V_RV_2	GAATTCTCCAATGGGTCTTG
CNOT7_L79Q_FW_2	CAGCTAGGACAGACATTTATGAATG
CNOT7_L79Q_RV_2	AATTATCTTTAACAAGTCTACATTAC
CNOT7_E87Q_FW_2	GAATGAACAGGGTCAATATCCGCCGGG
CNOT7_E87Q_RV_2	CCCGCGGATATTGACCCTGTTTCATTC
CNOT7_W153C_FW_2	GGGTCAAATGCTTGTCATTTTCATAG
CNOT7_W153C_RV_2	CTTCACAGAGGACCACTC
CNOT7_H157Y_FW_2	GTTGTCATTTTATAGCGGTTAC
CNOT7_H157Y_RV_2	CATTTGACCCCTTCACAG
CNOT7_D161N_FW_2	TAGCGGTTACAACCTTTGGCTAC
CNOT7_D161N_RV_2	TGAAATGACAACCATTTGAC
CNOT7_G163C_FW_2	TTACGACTTTTGCTACTTAATCAAATC
CNOT7_G163C_RV_2	CCGCTATGAAATGACAAC

**Table 2.6 List of primers for CNOT7 site directed mutagenesis.**

<b>Name</b>	<b>Sequence 5'- 3'</b>
CNOT7_G208V_FW_2	CTCAAAGGTGTATTACAGGAGG
CNOT7_G208V_RV_2	ATTTTTGCAGCTCTTCATG
CNOT7_L209F_FW_2	AAGGTGGATTCAGGAGGTGG
CNOT7_L209F_RV_2	TGAGATTTTTGCAGCTCTTC
CNOT7_E217Q_FW_2	AGAACAGTTACAGCTGGAACGGATAG
CNOT7_E217Q_RV_2	GCCACCTCCTGTAATCCAC
CNOT7_R220L_FW_2	GAGCTGGAAGTATAGGACCAC
CNOT7_R220L_RV_2	TAACTGTTCTGCCACCTC
CNOT7_R220W_FW	AGAGCTGGAATGGATAGGACC
CNOT7_R220W_RV	AACTGTTCTGCCACCTCC
CNOT7_G228E_FW_2	CATCAGGCAGAATCTGATTCATTG
CNOT7_G228E_RV_2	TTGTGGTCCTATCCGTTC
CNOT7_L232S_FW_2	TCTGATTCATCGCTCACAGGAATG
CNOT7_L232S_RV_2	TCCTGCCTGATGTTGTGG
CNOT7_F239I_FW_2	AATGGCCTTTATCAAAATGAGAG
CNOT7_F239I_RV_2	CCTGTGAGCAATGAATCAG
CNOT7_E243G_FW_2	AAAATGAGAGGAATGTTCTTTGAAG
CNOT7_E243G_RV_2	GAAAAAGGCCATTCCTGTG

**Table 2.7 List of primers for CNOT7 site directed mutagenesis.**

<b>Name</b>	<b>Sequence 5'- 3'</b>
Gibson Cnot7 Fw	CCACTGCTGGGCCTGGACAGCACCCCTCGAGATGCCAGCGG CAACTGTAGA
Gibson Cnot7 Rv	TCTAGAGGTACCACGCGTGAATTCTCGACTCATGACTGCT TGTTGGCTTC

**Table 2.8 List of primers for CNOT7 Gibson cloning into pCI $\lambda$ N vector.**

## 2.3 Protein purification and analysis

### 2.3.1 Reagents, stock solutions and buffers

*Extraction Buffer*: 20 mM Tris-HCl pH 7.8, 500 mM NaCl, 10% glycerol, 1 mM  $\beta$ -mercaptoethanol.

*Wash Buffer*: 20 mM Tris-HCl pH 7.8, 500 mM NaCl, 10% glycerol, 1 mM  $\beta$ -mercaptoethanol, 20 mM imidazole.

*Elution Buffer*: 20 mM Tris-HCl pH 7.8, 500 mM NaCl, 10% glycerol, 1mM  $\beta$ -mercaptoethanol, 250 mM imidazole.

*Upper Buffer (4x)*: 0.5 M Tris Base, 0.4% SDS, pH 6.8.

*Lower Buffer (4x)*: 1.5 M Tris Base, 1.0% SDS, pH 8.8.

*Running Buffer (10x)*: 0.25 M Tris Base, 1.0% SDS, 1.92 glycine

*Reducing SDS loading buffer (4x)*: 2.4 mL 1M Tris byffer, 4 mL 100% glycerol, 0.5 mL  $\beta$ -mercaptoethanol, 0.8 g SDS, 4 mg Bromophenolblue, and 3.1 mL water, stored at -20°C.

### 2.3.2 Expression culture of *E. coli*

A single colony of *E. coli* BL21 RIL containing plasmid was used to inoculate a starter culture (1 ml LB with 50-100  $\mu$ g/ml ampicillin), then incubated at 37 °C for 6–8 hours in an orbital shaker. The obtained starter culture was then used to inoculate the pre-culture (100 ml LB containing 50  $\mu$ g/ml ampicillin) and incubated overnight at 37 °C. The pre-culture was added to 1L of LB/ampicillin (50  $\mu$ g/ml) and grown at 37 °C. Once the OD<sub>600nm</sub> was between

0.6-0.8, the expression was induced by IPTG at a final concentration of 0.2 mM. The expression culture was incubated at 30°C for 3 hours.

### **2.3.3 Protein extraction from expression culture of E. Coli**

*E. coli* bacteria were harvested by centrifugation at 5500 rpm at 4°C for 15 minutes using Sorvall SLC-6000 SUPER LITE rotor. The supernatant was discarded and the bacterial pellet re-suspended in protein purification extraction buffer as appropriate. Then the suspension was frozen at -80 °C. The bacteria suspension was thawed and bacteria lysed by sonication on ice using Qsonica XL2000 for 5 minutes (30 sec on/30 sec off). The crude lysate was then centrifuged in a SS-34 rotor at 10,000 rpm at 4°C for 30 minutes. The pellet was discarded and the supernatant recovered. The suspension was then stored at -80 °C.

### **2.3.4 Protein purification**

Cobalt-agarose beads (Sigma) were washed twice with extraction buffer and then incubated overnight at 4 °C with the soluble lysate. 5 ml of extraction buffer were, then, added to the beads bound to proteins containing 6His-tag and the samples were applied to the column. After eluting, 10 ml of wash buffer was added, the wash was recovered for SDS-PAGE analysis. Proteins were eluted using Elution buffer and fractions were collected each 0.5 mL. The protein purification was followed by Bradford and SDS-PAGE analysis.

### **2.3.5 Bradford assay to determine protein concentration**

A Bradford assay was used to determine protein concentration of collected fractions or cell lysates. The protein assay reagent (Bio-rad, 500-0006) was diluted 1:5 in H<sub>2</sub>O before use. A standard curve was produced using BSA

dissolved in water and preparing 6 standards: 0.1mg/ml, 0.2 mg/ml, 0.4 mg/ml, 0.6 mg/ml, 0.8 mg/ml and 1.0 mg/ml. 10  $\mu$ l of protein sample was added to 990  $\mu$ l of 1x Bradford reagent. Samples were transferred to 1 ml cuvettes and the absorbance read at OD<sub>600</sub>.

### **2.3.6 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)**

In order to perform SDS-PAGE, 10% gels were prepared and cast in gel cassettes (Invitrogen). The Invitrogen X-Cell SureLock Mini Cell system was used. Before use the comb was removed and the wells washed with 1x running buffer. Protein samples were denatured for 5 min in 1x SDS sample buffer prior to loading. Gels were run in 1x running buffer at 180V for 90 min.

## 2.4 Yeast two-hybrid analysis

### 2.4.1 Reagents, stock solutions and buffers

*Yeast extract peptone dextrose medium (YPD)*: 10 g/L yeast extract, 20 g/L glucose, 20 g/L bacteriological peptone. For YPD agar plates, 20 g/L bacteriological agar was included. The medium was sterilised by heat using a Prestige medical 2100 classical bench top autoclave and stored at room temperature.

*Yeast selective complete medium (YS)*: 6.7 g/L yeast nutrient broth, 20 g/L glucose, 1.4 g/L amino acid drop out mix. For selective agar plates, 20 g/L bacteriological agar was included.

*Single-Stranded DNA*: Herring Sperm single-stranded DNA (2 mg/mL, Sigma) was heated at 98° for 5 min and chilled on ice before use. Stored at -20°C.

*Litium acetate solutions*: LiAC solutions (2 M, 1 M and 0.1 M; Sigma) were prepared and sterilised using 0.22 µm pore size filters. Stored at room temperature.

*Polyethylene glycol (PEG-3350)*: a solution containing 50% polyethylene glycol (PEG-3350, Sigma) was prepared and sterilised using 0.22 µm pore size filters. Stored at room temperature.

*Reducing SDS loading buffer (4×)*: 2.4 mL 1M Tris byffer, 4 mL 100% glycerol, 0.5 mL β-mercaptoethanol, 0.8 g SDS, 4 mg Bromophenolblue, and 3.1 mL water, stored at -20°C.

### 2.4.2 Culture of *Saccharomyces cerevisiae* strain YRG2

The yeast strain *Saccharomyces cerevisiae* YRG2 (Stratagene, stored at -80° in YPD containing 50% glycerol) was used for yeast two-hybrid analyses. To obtain single colonies the yeast was streaked on YPD agar plates and grown in incubator at 30°C. Once the colonies were obtained (2-4 days) the plates were stored at 4°C for several weeks. Liquid cultures were inoculated with large single colonies and grown in universal tubes containing YPD and incubated at 30°C, 200 rpm overnight.

### 2.4.3 Transformation of YRG2 cells

Plasmids pBD-GAL4 Cam and pAD-GAL4-2.1 (Stratagene) and their derivatives were used for all yeast two-hybrid analyses. *S. cerevisiae* YRG2 cells were made competent through heat shock treatment and used on the same day they were made. One single colony was picked from the YPD agar plate and used to inoculate 10 mL YPD culture and incubated overnight in a shaker incubator at 30°C, 200 rpm. This culture was then used to inoculate 50 mL of YPD medium to obtain OD<sub>600</sub> 0.2. This culture was then incubated in a shaker incubator at 30°C, 200 rpm until the OD<sub>600</sub> reached 0.8 (3-5 hours). The culture was then centrifuged to pellet the yeast at 4,000 rpm for 5 minutes at room temperature. The supernatant was discarded and the pellet re-suspended in 25 mL of sterile water followed by another centrifuge step. The pelleted yeast was then re-suspended in 1 mL of 0.1 M LiAc and transferred to a 1.5 mL microfuge tube. Cells were then centrifuged at 13,000 rpm for 30 sec in a bench micro-centrifuge. The supernatant was discarded and the cells were re-suspended to a final volume of 500 µL in 0.1 M LiAc and aliquoted in 50 µL lots in microtubes. One 50 mL culture provided sufficient cells for ten transformations. The cells were pelleted and the supernatant discarded. The 50% PEG-3350 solution was added to the yeast cells followed by the mix containing the other transformation components as described in Table 2.9.

Volume	Reagent/Component
240 $\mu$ L	50% w/v PEG-3350
36 $\mu$ L	1.0M LiAc
25 $\mu$ L	Single-stranded DNA (2.0 mg/ml)
50 $\mu$ L	500 ng of combined plasmid DNA diluted in H <sub>2</sub> O

**Table 2.9 Mix for yeast transformation**

Following the addition of all the components, the pellet was re-suspended by vortexing. The yeast transformation mix was then incubated at 30°C, 200 rpm for 30 min followed by 20 min incubation at 42°C in a water bath. The samples were then centrifuged at 4,000 g for 30sec and the yeast re-suspended in 500  $\mu$ L of sterile water. 100  $\mu$ L of this suspension was plated onto yeast selective plates and incubated at 30°C for 3-5 days until colonies formed.

#### 2.4.4 Yeast two-hybrid $\beta$ -galactosidase assay

For each transformation, three independent yeast colonies were analysed. Yeast colonies were streaked into triangles on yeast selective plates. After 48-72 hours half of the triangle of yeast was used to inoculate 2 mL of yeast selective medium. After incubation at 30°C, 200 rpm for 24 h, one ml of this culture was used to obtain the OD<sub>600</sub> as measure of the culture density. When the culture was dense, with OD<sub>600</sub> >1, the samples was diluted 1:10 with yeast selective medium. From the remaining culture 25  $\mu$ L was added to 25  $\mu$ L of  $\beta$ -Glo reagent (Promega, Cat. E469A) in a white 96 well plate and incubated for 5 minutes shaking and then 25 minutes at room temperature in the dark.  $\beta$ -galactosidase luminescence was measured using a GloMax 96 microplate luminometer (Promega). The  $\beta$ -galactosidase activities values were then normalised using the OD<sub>600</sub> values.

#### **2.4.5 Western blot analysis of yeast cell lysates**

To confirm the expression of the GAL4 AD-CNOT7 and the GAL4 BD-HA-CNOT1<sub>MIF4G</sub>, GAL4 BD-HA-CNOT6L, and GAL4 BD-HA-BTG1 fusion proteins, yeast lysates were produced for western blot analysis. Individual colonies were streaked into triangles and then cultured in 2 mL yeast selective medium. One millilitre of the yeast culture was transferred into 1.5 mL microfuge tube and centrifuged at 3000 g for 2 minutes. The supernatant was discarded and the pellet was re-suspended in 100  $\mu$ L of 2.0 M LiAc. The suspension was centrifuged again and the pellet was re-suspended in 100  $\mu$ L of water. After the addition of 100  $\mu$ L 0.4M NaOH, the cells were vortexed and incubated for 10 minutes at room temperature. The cells pelleted again and re-suspended in 50  $\mu$ L of 1 $\times$  reducing SDS loading buffer. The samples were then boiled at 98°C for 5 minutes before being analysed by SDS-PAGE or stored at -80°C.

## 2.5 Cell culture

### 2.5.1 Reagents, stock solutions and buffers for use in tissue culture

*Dulbecco's Phosphate buffered saline (PBS)*: 137 mM NaCl, 2.7 mM KCl, 4.3 mM NaHPO<sub>4</sub>, 1.47 mM KH<sub>2</sub>PO<sub>4</sub> (1× PBS without Ca<sup>2+</sup>, Mg<sup>2+</sup>, Sigma-Aldrich)

*Dulbecco's Modified Eagle media (DMEM) complete*: DMEM (Sigma-Aldrich), 10% foetal bovine serum (FBS) (Thermo-Fisher Scientific), 100 units/mL penicillin and 100 units/mL streptomycin (PAA), L-glutamine 200 mM (Lonza)

*Trypsin/EDTA solutions*: 10× concentrated solutions of phosphate-buffered saline containing 0.5% trypsin and 0.2% EDTA were purchased from (PAA) stored at -20°C. 1× solutions obtained by the addition of PBS were stored at 4°C.

### 2.5.2 Routine maintenance of cell lines

Cell culture was performed in a Class II biological safety cabinet under aseptic conditions. Adherent cells were routinely maintained in Dulbecco's modified Eagles medium (DMEM) supplemented with 10% foetal bovine serum (FBS), 2 mM L-glutamine, 100 units/ml penicillin and 100 units/ml streptomycin, in a humidified incubator at 37°C and 5% CO<sub>2</sub>. When cells were 70%-80% confluent, they were washed with pre-warmed phosphate buffered saline (PBS) and incubated in 1-2 ml of trypsin/EDTA for 1 to 5 minutes depending on cell type. Trypsin was deactivated by addition of 8-9 ml of complete DMEM and the cells re-suspended. The suspension was then centrifuged at 1100 rpm for 5 minutes, the supernatant removed and the cells were re-suspended in complete DMEM for dilution and plating.

Cell line	Origin	Number of passage	Supplier
MCF7-NKI	Breast cancer cell	12	DSMZ
	Human embryonic kidney		
HEK293	293	12	DSMZ
HeLa	Cervical cancer cell	12	DSMZ

Cell line	Origin	Number of passage	Supplier
MCF7-NKI	Breast cancer cell	12	DSMZ
	Human embryonic kidney		
HEK293	293	12	DSMZ
HeLa	Cervical cancer cell	12	DSMZ

**Table 2.10 Mammalian cell lines used.** All cell lines were supplied from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ)

### 2.5.3 Freezing cells for long term storage

To store cells into liquid nitrogen, cells were detached from a 75 cm<sup>2</sup> flask using trypsin/EDTA, transferred in a 15 ml Falcon tube and centrifuged at 1100 rpm for 5 min. After centrifugation, the pellet was re-suspended in 2 ml of complete DMEM containing 10% of dimethylsulfoxide (DMSO) and 1 ml aliquots of the cell suspension were transferred into cryovials. The cryovials were then placed at -80°C in a Mr. Frosty container (NALGENE Labware), which allows slow freezing of -1°C per minute. After several days, cryovials were placed in liquid nitrogen for long-term storage.

### 2.5.4 Retrieving cells from liquid nitrogen storage

To start culture cells from liquid nitrogen storage, they were firstly quickly thawed by incubation at 37°C in the water bath, and then added to 9 ml of complete DMEM medium pre-warmed at 37°C. In order to remove the DMSO from the cell suspension, cells were pelleted at 1100 rpm for 5 minutes. The suspension was then re-suspended in 10 ml of complete DMEM medium pre-warmed at 37°C. The cell suspension was then transferred to a 75 cm<sup>2</sup> flask and placed in the humidified incubator (5% CO<sub>2</sub> and 37°C). After 24 hours, the medium was changed, and the cell passaged between 48 and 96h after retrieval from liquid nitrogen.

## 2.6 Immunoprecipitation analyses

### 2.6.1 Reagents, stock solutions and buffers for use in immunoprecipitations

*1 M Dithiothreitol (DTT)*: dissolved in H<sub>2</sub>O, stored at -20°C.

*Agarose beads*: protein A-agarose beads and protein G-agarose beads were purchased from Roche and stored at 4°C.

*Cell lysis buffer containing 0.2% NP-40*: 50 mM 1 M Tris-HCl pH 8.0, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.5 mM EDTA, 5% glycerol, 0.2% NP-40, 1 mM DTT. After the addition of one Roche complete protease inhibitor cocktail tablet to 50 mL, the lysis buffer was stored at -20°C.

*Phosphate buffered saline*: 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.47 mM KH<sub>2</sub>PO<sub>4</sub> (1× PBS without Ca<sup>2+</sup>, Mg<sup>2+</sup>; PAA).

*Upper Buffer (4x)*: 0.5 M Tris Base, 0.4% SDS, pH 6.8.

*Transfer buffer (10x)*: 0.25 M Tris Base, 1.92 M glycine, stored at room temperature.

*Reducing SDS loading buffer (4x)*: 2.4 mL 1M Tris buffer, 4 mL 100% glycerol, 0.5 mL β-mercaptoethanol, 0.8 g SDS, 4 mg Bromophenolblue, and 3.1 mL water, stored at -20°C.

### **2.6.2 Preparation of protein lysates for immunoprecipitation**

40-80×10<sup>4</sup> HEK293T cells were seeded in 6cm diameter culture dishes. After 24 hours when the cells were 50-70% confluent, the DMEM medium was changed and cells were transfected with the appropriate plasmid using JetPEI transfection reagent. A total of 5 µg total plasmid (2.5 µg of each vector) was added to 10 µl of JetPEI following the manufacturer instruction. The DNA/JetPEI mixture was incubated for 30 minutes before being added to the cells drop wise. After 48 hours from transfection cells harvested by pipetting the medium directly onto the cells. Cell suspension was centrifuged at 4,000 g (4°C) for 5 minutes and the pellet re-suspended in 1ml of pre-chilled PBS (4°C) to wash the cells twice. Pellets were then re-suspended in 500 µl of cold cell lysis buffer containing 0.2% NP-40. Cellular lysates were then freeze/thawed twice (-80°C) and centrifuged at 16,000 g for 5min (4°C). Supernatants were then stored at -80°C for further use.

### **2.6.3 Preparation of antibody-agarose beads for immunoprecipitation**

To prepare agarose beads for immunoprecipitations, 10 µL of protein A and 10 µL of protein G agarose beads (Roche) were mixed together in a 1.5 mL microfuge tube and centrifuged at 300 g for 30 sec. The supernatant was discarded and the beads were washed twice with 500 µL of PBS before the appropriate antibody was added. One microgram of antibody was used per 10 µL of beads and the total final volume was 500 µL in PBS. Antibodies used for immunoprecipitations are listed in Table 2.11. The beads-antibody mixture was then incubated at room temperature on a rotator shaker for 1 hour. Following centrifugation at 3000 g for 30 sec the supernatant was carefully removed with a gel loading tip. The agarose beads were then washed three times with 500 µL PBS followed by one wash with 500 µL of cell lysis buffer containing 0.2% NP-40. To the antibody coated beads, 450 µL of protein lysates was added. The mixture

was then incubated overnight on a rotator shaker at 4°C. The following day the mixture was pelleted at 300 g for 30 sec and the supernatant discarded. Beads were then washed three times with 500µL of lysis buffer containing 0.2% NP-40, prior to add 10 µl of PBD and 10 µL of SDS loading buffer. The samples were boiled for 5 minutes and subjected to SDS-PAGE for western blot or stored at -80°C until necessary.

Antibody	Dilution	Clone	Supplier
Anti-Flag	1 mg per 10 mL of agarose beads	M2, monoclonal	Sigma-Aldrich
Anti-GFP	1 mg per 10 mL of agarose beads	FL, polyclonal	Santa Cruz Biotechnology

**Table 2.11 List of antibody used for immunoprecipitations**

## 2.7 siRNA knockdown

### 2.7.1 Reagents, stock solutions and buffers for use in protein analysis

*TE Buffer*: 10 mM Tris-HCl pH 8.0, 1 mM EDTA.

*DMEM complete*: DMEM (Sigma-Aldrich), 10% fetal bovine serum (FBS) (Thermo-Fisher Scientific), 100 units/mL penicillin and 100 units/mL streptomycin (PAA), glutamine (200 mM; Lonza).

*Serum free medium*: DMEM (Sigma-Aldrich, D6546), 100 units/mL penicillin and 100 units/mL streptomycin (PAA, P11-010), glutamine (200 mM; Lonza, BE17-60SE).

*Phosphate buffered saline*: 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.47 mM KH<sub>2</sub>PO<sub>4</sub> (1× PBS without Ca<sup>2+</sup>, Mg<sup>2+</sup>; PAA).

*siRNA*: siRNAs were purchased from Ambion.

### 2.7.2 siRNA transfection

Working in class II biological safety cabinet, MCF7 cells were seeded in 6 well plates at  $15 \times 10^4$  cells per well. For each well, 2.2 pmoles of siRNA (31 ng) duplexes were diluted in 200  $\mu$ l of serum free medium vortexing gently. The siRNA used are reported in Table 2.12. Ten microliters of Interferin (Polyplus, Cat. 409-10) were then added to the siRNA duplex, homogenised immediately for 10 sec, and incubated for 10 minutes at room temperature. During the incubation time, the growth medium was replaced with 2 ml of fresh complete medium per well. The transfection mix was then added onto the cells and distributed by swirling.

Transfection efficiency was estimated by observing cells by light microscopy 48-96 hours after transfection of AllStars Cell Death Control siRNA (Qiagen, Cat. 1027299). After 48 hours from transfection cells were washed twice with PBS, and harvested in 1 ml of PBS by scraping. Cell suspension was centrifuged 4000 g at 4°C for 5 minutes and the pellet re-suspended in 1 ml of pre-chilled PBS (4°C) to wash the cells. Pellets were then re-suspended in 500  $\mu$ l of cold cell lysis buffer containing 0.2% NP-40. Cellular lysates were then freeze/thawed twice (-80°C) and centrifuged at 16,000 g for 5 min at 4°C. Supernatants were then stored at -80°C for further use or analysed immediately by western blot.

siRNA	Supplier
siRNA Pan2	Ambion, 4390824, ID: s19252
siRNA Pan3	Ambion, 4392420, ID: s48723
Silencer Negative control No 1 siRNA	Ambion, AM4611
AllStars Cell Death Control siRNA	Qiagen, 1027299

**Table 2.12 siRNA used for knockdown.**

## 2.8 RNA tethering assay

### 2.8.1 Reagents, stock solutions and buffers for use in protein analysis

*Dulbecco's Modified Eagle media (DMEM) complete*: DMEM (Sigma-Aldrich), 10% foetal bovine serum (FBS)(Thermo-Fisher Scientific), 100units/mL penicillin and 100 units/mL streptomycin (PAA), L-glutamine 200 mM (Lonza).

*Phosphate buffered saline*: 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.47 mM KH<sub>2</sub>PO<sub>4</sub> (1× PBS without Ca<sup>2+</sup>, Mg<sup>2+</sup>; PAA).

*Hepes-buffered saline (HBS)*: 280 mM NaCl, 50 mM HEPES pH 7.0, 10 mM KCl, 12 mM dextrose, 1.5 mM Na<sub>2</sub>PO<sub>4</sub>.

*Radioimmunoprecipitation assay buffer (RIPA)*: 150 mM NaCl, 10% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris, pH 8.0.

### 2.8.2 Cell seeding and transfection for RNA tethering assay

Working in class II biological safety cabinet, HEK293T cells were seeded into 12-well culture plates at 8-10×10<sup>4</sup> cells per wells. After 24 hours, cells at 70% confluency were co-transfected using CaCl<sub>2</sub>. For each transfection, 250 ng of the reporter plasmid pRL-5boxB and 250 ng of pCIλN derivative plasmids. DNA-Calcium phosphate mixture was set up as reported in Table 2.13 and incubated for 30 minutes before being added drop-wise to the cells. After 16 hours the medium was removed replaced with fresh medium after two washes with PBS. Twenty four hours after transfection the cells were harvested for the luciferase assay.

Reagent	Amount
DNA	250 ng of pRL5boxB and 250 ng of pCI $\lambda$ N plasmids
2M CaCl <sub>2</sub>	6 $\mu$ L
dH <sub>2</sub> O	Up to 50 $\mu$ L
2 $\times$ HBS	50 $\mu$ L

**Table 2.13 Transfection setup for RNA tethering assay**

### 2.8.3 Luciferase reporter assay

Twenty four hours after transfection of the reporter/tethering plasmid, cells were washed with ice-cold PBS and 100  $\mu$ L of Luciferase cell lysis buffer (NEB) were added directly on the cells. The plate was then placed at -80°C for 15 minutes until completely frozen. The samples were then thawed and harvested by pipetting to help the lysis process. The suspension was transferred into 1 mL microfuge tube and centrifuged 16,000 g for 5min at 4°C. The samples were stored in -80°C freezer or used immediately to perform the luciferase assay. Bradford assay was performed to determine protein concentration of the samples.

The luciferase assay was performed using the BioLuxGluc Luciferase Assay kit (NEB). Reagents for the assay were prepared according to the manufacturer's instruction. Ten microliters of cell lysates were transferred on white opaque flat bottom 96 well plate. Fifty microliters of BioLuxGluc reagent were added for each well and luminescence was read using the Promega GloMax 96 microplate luminometer (delay 5.0 sec, integration time 2.0 sec) Each transfection was carried out in triplicate.

### 2.8.4 Cellular lysates for western blot

Working in a class II biological safety cabinet, HEK293T cells were seeded into 6-well culture plates at 16-20 $\times$ 10<sup>4</sup> cells per wells. Transfection with the calcium phosphate protocol was performed as described previously doubling the amount of the reagents. After 48 hours from transfection, cells were washed twice with cold PBS and harvested by directly pipetting 1ml of PBS onto the cells. Cellular suspensions were then transferred to 1.5 ml microfuge tubes and centrifuged at 3,000 g for 5 min at 4°C. The

supernatant was discarded and the pellet re-suspended with 100  $\mu$ l RIPA buffer. Samples were then incubated 1 hour in a disk tube rotator at 4°C. Cell suspension was then pelleted at 13,000 g for 30 sec at 4°C and the supernatant was analysed by SDS-PAGE and western blot.

## 2.9 Protein analysis

### 2.9.1 Reagents, stock solutions and buffers for use in protein analysis

*Upper buffer (4×)*: 0.5 M Tris base, 0.4% SDS, pH 6.8; stored at room temperature.

*Lower buffer (4×)*: 1.5 M Tris base, 0.4% SDS, pH 8.8; stored at room temperature.

*Running buffer (10×)*: 0.25 M Tris Base, 1.0% SDS, 1.92 M glycine; stored at room temperature  
*Transfer buffer (10×)*: 0.25 M Tris Base, 1.92 M glycine; stored at room temperature.

*Reducing SDS loading buffer (4×)*: 2.4 mL 1M Tris buffer, 4 mL 100% glycerol, 0.5 mL  $\beta$ -mercaptoethanol, 0.8 g SDS, 4 mg Bromophenolblue, and 3.1 mL water, stored at -20°C.

*Transfer buffer (10×)*: 0.25 M Tris Base, 1.92 M glycine. Stored at room temperature.

*Tris buffer saline supplemented with 0.05% Tween-20 (TBST)*: 50 mM TrisHCl pH 7.8, 150 mM NaCl, 0.1% Tween-20. Stored at room temperature.

*10% APS*: 10% ammonium persulphate (APS) in H<sub>2</sub>O.

### **2.9.2 Bradford assay to determine protein concentration**

A Bradford assay was performed to determine the concentration of proteins. The Sigma protein reagent (Sigma-Aldrich, B6916-500ML) was diluted 1:5 in H<sub>2</sub>O before use. A protein standard curve was produced using BSA dissolved in water. Standard 0.2 mg/mL, 0.4 mg/mL, 0.6 mg/mL, 0.8 mg/mL, and 1.0 mg/mL and 2 µL of the protein sample buffer was added. Two microliters of protein sample were added to the 998 µL of diluted Bradford reagent. Samples were transferred to a 1 ml cuvette and the absorbance read at 600 nm. The standard curve was obtained by linear regression analysis using Microsoft Excel and was used to calculate the protein concentration of the samples.

### **2.9.3 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)**

Proteins were analysed using SDS-PAGE. Resolving gels were prepared in the appropriate percentage (Table 2.14) and cast in gel cassettes (Invitrogen). The separating gel was cast first and isopropanol was added on the top and allowed to set for 30 minutes. After removing the isopropanol, the stacking gel was cast and the appropriate comb was inserted in the gel and left to set for further 30 minutes. Before use the comb was removed and the wells were washed with 1× running buffer. The Invitrogen X-Cell SureLock Mini Cell system was used for SDS-PAGE analysis. Protein samples were denatured for 5 min in 1× SDS sample buffer prior to loading. Gels were run in 1× running buffer at 180 V for 60-90 min.

	Resolving gel			Stacking gel
	10%	12%	14%	4%
Protein MW (kDa)	30-200	20-150	10-80	
40% acrylamide:				
Bisacrylamide (29:1)	2.0 mL	2.4 mL	2.8 mL	300 $\mu$ L
4 $\times$ lower buffer	2.0 mL	2.0 mL	2.0 mL	-
4 $\times$ upper buffer	-	-	-	750 $\mu$ L
10% APS	80 $\mu$ L	80 $\mu$ L	80 $\mu$ L	1950 $\mu$ L
TEMED	8 $\mu$ L	8 $\mu$ L	8 $\mu$ L	60 $\mu$ L
H <sub>2</sub> O	4.0 mL	3.6 mL	3.2 mL	6 mL

**Table 2.14 Preparation of SDS-PAGE gels****2.9.4 Western blot analysis**

Proteins separated by SDS-PAGE, were transferred to nitrocellulose Whatman Protan 0.45  $\mu$ m membrane using the Invitrogen X-Cell SureLock Mini Cells system for 1 hour at 25 V in 1 $\times$  transfer buffer. The membrane was then blocked using 5% w/v skimmed milk powder in 1 $\times$  TBST for 1 hour on a platform rotator at room temperature. The membrane was then transferred to a 50 ml Falcon tube and incubated overnight at 4 $^{\circ}$ C with TBST containing 5% w/v skimmed milk powder the appropriate primary antibody (Table 2.15 Membranes were then washed three times with TBST and were transferred to a new 50 ml falcon tube containing 5 ml 5% milk in TBST with the appropriate secondary antibody solution (Table 2.16) and incubated 1 hour at room temperature on a platform shaker. The membrane was then washed three times with TBST and placed between two plastic sheets. The Pierce ECL Western Blotting substrate detection kit (Thermo-Fisher Scientific, Cat. 32209) was added directly on the membrane. After one minute of incubation, the signals were detected using a Fujifilm LAS-400 digital imaging system. Exposure time varied depending on the protein detected. Images were analysed using ImageJ.

<b>Primary Antibodies</b>	<b>Dilution</b>	<b>Clone</b>	<b>Supplier</b>
Pan2/USP52	1:1000	C-20, polyclonal	Santa Cruz Biotechnology
Pan3	1:500	A-9, monoclonal	Santa Cruz Biotechnology
Anti-Flag	1:1000	M2, monoclonal	Sigma-Aldrich
Anti-GFP	1:1000	FL, polyclonal	Santa Cruz Biotechnology
Anti-HA	1:1000	3F10, monoclonal	Roche
Anti-V5 probe	1:1000	E10, monoclonal	Santa Cruz Biotechnology
Anti-GAL4-TA	1:1000	C-10, monoclonal	Santa Cruz Biotechnology
Anti-Tubulin	1:1000	C-20, polyclonal	Santa Cruz Biotechnology

**Table 2.15 Primary antibodies used for western blotting.** Antibodies were diluted in TBST containing 5% w/v skimmed milk powder.

<b>Secondary Antibodies</b>	<b>Dilution</b>	<b>Supplier</b>
Chicken anti-mouse HRP	1:5000	Santa Cruz Biotechnology
Goat anti-rat HRP	1:1000	Santa Cruz Biotechnology
Rabbit anti-goat HRP	1:1000	Santa Cruz Biotechnology
Goat anti-rabbit HRP	1:1000	Cell signalling

**Table 2.16 Secondary antibodies used for western blotting.** Antibodies were diluted in TBST containing 5% w/v skimmed milk powder .

## **2.10 Stable isotope labelling in cell culture (SILAC)-rapid immunoprecipitations mass spectrometry of endogenous protein (RIME)**

### **2.10.1 Reagents, stock solutions and buffers for use in SILAC-RIME analysis**

*"Heavy"-labelled medium:* R/K-deficient SILAC DMEM (PAA), 10% dialysed fetal bovine serum (Sigma-Aldrich), 800  $\mu$ M L-Lysine  $^{13}\text{C}_6^{15}\text{N}_2$ -Hydrochloride (Sigma-Aldrich), 482  $\mu$ M L-Arginine  $^{13}\text{C}_6^{15}\text{N}_4$  hydrochloride (Sigma-Aldrich), L-glutamine (200 mM; Lonza).

*"Light"-labelled medium:* R/K-deficient SILAC DMEM (PAA, E15-086), 10% dialysed fetal bovine serum (Sigma-Aldrich), 800  $\mu$ M L-Lysine  $^{12}\text{C}_6^{14}\text{N}_2$ -Hydrochloride (Sigma-Aldrich), 482  $\mu$ M L-Arginine  $^{12}\text{C}_6^{14}\text{N}_4$  hydrochloride (Sigma-Aldrich), L-glutamine (200 mM; Lonza).

*Lysis buffer:* 5 mM Tris-HCl pH 8.0, 85 mM NaCl, 0.5% NP-40, protease inhibitor.

*RIPA buffer:* 10mM Tris-HCl pH 8, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% SDS, 0.1% sodium deoxycholic acid, 1% Triton X-100, and protease inhibitor.

*Sonication buffer:* 50 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.1% SDS, 0.5% sodium deoxycholic acid, and protease inhibitors.

*Trypsin solution:* 10  $\mu$ g/mL trypsin in 100 mM  $\text{NH}_4\text{HCO}_3$

### **2.10.2 Preparation of magnetic beads**

Protein A magnetic beads suspension (100  $\mu$ L) was placed on a magnetic stand and the supernatant was removed by aspiration. One millilitre of PBS supplemented with 5 mg/mL BSA was added to wash and block the beads and then the tube was put back on the magnetic stand. This step was repeated further three times. The beads were then re-suspended in 500  $\mu$ L of PBS/BSA and added to cellular lysates.

### **2.10.3 Preparation of cell lysates**

MCF7 cells ( $1 \times 10^7$ ) were grown in "heavy"-labelled complete media or "light"-labelled complete media.

When confluent, the media were removed and replaced with media containing 1% formaldehyde (Tebu-bio). After 8 min, glycine was added to a final concentration of 0.2 M. The cells were then washed with 10 mL of cold PBS and harvested with a cell scraper in 500  $\mu$ L of PBS. The cell suspensions were then centrifuged at 2,000 g for 3min at 4°C, the supernatant was removed and the cells were re-suspended in 500  $\mu$ L of PBS. After repeating the centrifugation step, the supernatant was removed and the cell pellet frozen at -80°C until further use. The frozen pellet was then re-suspended in 5 mL of lysis buffer and incubated one hour on ice followed by 50 strokes in a tight pestle Dounce homogenizer. Nuclei were separated from the cytoplasmic extract by centrifugation and washed with 25 mL of lysis buffer. Washed nuclei were re-suspended then in 1 mL of sonication buffer and subjected to 15 cycles of sonication in a cooled waterbath sonicator (Diagenode Bioruptor, Medium intensity, 30sec on/off cycles).

#### **2.10.4 Immunoprecipitation**

Nuclear and cytoplasmic fractions were investigated separately or combined. The extracts were centrifuged at 14,000 rpm at 4°C to clear from aggregates and were then incubated overnight with the appropriate antibodies while tumbling at 4°C. For control immunoprecipitations pre-immune rabbit IgG was used. "Heavy" SILAC labelled extracts were used for immunoprecipitations using a mix of 10 µg anti-CNOT6 and 5 µg anti-CNOT6L, or 10 µg anti-CNOT7 and 10µg anti-CNOT8 custom antibodies (Eurogentec). The extracts were then incubated with 100 µL of protein A coated magnetic beads equilibrated with sonication buffer for 4 hours at 4°C tumbling. RIPA buffer was used to rinse the beads and control and specific immunoprecipitations were combined followed by five washes with 0.5 mL RIPA buffer followed by two rinses with 1.0 mL 50 mM ammonium bicarbonate. The supernatant was removed and the beads subjected to on-based tryptic digestion adding 10 µL of trypsin solution directly onto the beads. Beads were re-suspended by vortexing and incubated at 37°C overnight. The tubes were placed on a magnet to allow the beads to collect at the tube wall; the supernatant was removed and directly placed into formic acid to give a final concentration of 5%. Digested peptide mixtures were diluted 1:10 with loading buffer (0.1% formic acid/2% acetonitrile/water), and 2–5 µL aliquot of the diluted peptide mixture was analysed by nano-LC-MS/MS.

The SILAC-RIME experiments were carried out by Dr Klaas Mulder, Cambridge Research Institute, Cancer research UK and the proteomics facility, Cambridge Research Institute, Cancer research UK.

## **2.11 Bioinformatics**

### **2.11.1 Protein-protein interaction network generation**

Protein-protein interaction data, deriving from previous SILAC-RIME experiments, were processed by using Microsoft Excel and imported in Cytoscape as "Network from table". Attributes tables were also imported in order to have a correlation between the node and the Heavy/Light ratio. Once imported, networks were visualized by "Spring Embedded" layout. To analyse network different plugins were used, which were downloaded from the Cytoscape App store (<http://apps.cytoscape.org/>). By using the VizMapper option, bg\_name was set as Node Label showing canonical names of proteins. CNOT6-CNOT6L and CNOT7-CNOT8 protein-protein interaction networks were merged using the "merge networks" option in the Tool.

### **2.11.2 Expansion of intra-network protein-protein interactions**

The BisoGenet Cytoscape plugin (v.1.41) (Martin et al., 2010) was used to search for molecular interactions from the SysBiomics database of BisoGenet and to display the results as interaction networks. The SysBiomics database integrates data from interaction several databases including DIP, BIOGRID, HPRD, BIND, MINT and INTACT. The associated information of the protein networks includes annotations from the NCBI, UniProt, KEGG, and Gene Ontology databases. The BisoGenet plugin was launched with the option "Expand current network" and identifiers were set as "Protein identifiers only", data settings as "Protein Protein Interaction", the output as "Proteins" and the method used was "by adding edges connecting input nodes".

### **2.11.3 Clustering analysis**

Clustering analysis was carried out by two Cytoscape plug-ins. The ClusterOne plugin (v.1.0) (Zhang et al., 2014) was used to find potential overlapping protein complexes in a protein interaction network. It was used to group proteins into functional clusters by using the parameters: minimum size 3, node penalty 2, merging method "single-pass". The clusters found were sorted by number of nodes, density, quality and P-value.

As an alternative method, MCODE was used to find clusters defined as highly interconnected regions (Bader et al., 2003). After the MCODE run, the original network has a set of nodes called MCODE\_SCORE that can be used to colour the nodes in the network by their neighbourhood density using Cytoscape's Visual Styles. The MCODE plug-in was used with default parameters.

### **2.11.4 Generation of Residues interaction networks (RIN)**

Residue interaction networks were obtained by using the RINalyzer 2.0 package (Doncheva et al., 2011). Residue interaction network (RIN) format data was obtained using the RINertator (v0.5) module together with the Reduce (v3.14) and Probe (v2.12) tools starting from Protein databank (PDB) data of the NOT-module crystal structure (PDB accession number 4C0D). RIN data was then visualised and explored using UCSF Chimera and Cytoscape 3.1.1.

### **2.11.5 Analysis of protein variants using UCSF Chimera**

UCSF Chimera (Pettersen et al., 2004) was used to inspect how mutated residues are accommodated within the CNOT7 protein structure (PDB accession number 4GMJ). Residues within 5 angstroms from the investigated amino acid were visualised as atoms. Residues were mutated by the option Tool> Structure editing > rotamers. The Dunbrack rotamer library was selected and only rotamers with a probability >5% were examined. Clashes were found by the option Surface/Binding Analysis >

Find Clashes/ Contacts. The "Find atoms with VDW overlap" was set 0.6 angstroms and "Subtract from overlap for potentially H-bonding pairs" was set 0.6. Found clashes were resolved by Tool > Structure Editing > Minimise. Clashing residues were selected and minimisation of the structure was performed by selecting the appropriate model to minimise, using Steepest descent steps 100, Steepest descent step size 0.02 angstroms, Conjugate gradient steps 10, Conjugate gradient step (angstrom) 0.02, Update interval 10, and Fixed atoms "unselected". Hydrogen atoms were added while considering H-bonds. Charges were assigned with default parameters.

## 2.12 Phylogenetic analysis

### 2.12.1 Retrieval of homologous proteins

Homologues sequences of the examined proteins were retrieved by using the Domain Enhanced Lookup Time Accelerated–Basic Local Alignment Sequence Tool (DELTA-BLAST) (Boratyn et al., 2012). The reference protein (refseq\_protein) database of was used as the query dataset, and humans proteins as subjects.

Parameters were set as below:

- *evaluate* indicating the expect value (E) for saving the hits was set to 10;
- *max\_target\_seqs* defining the number of aligned sequences to keep was set to 500;
- *wordsize* defining the word size of initial match was set to 3;
- *matrix* indicating the scoring matrix name was set BLOSUM62;
- *gapopen* indicating the cost to open a gap was set to 11;
- *gapextend* indicating the cost to extend a gap was set to 1;
- *comp\_based\_stats* indicating the composition-based statistic set to 2 (default).

Only proteins with a query length coverage >75% were considered as homologues of the subject protein. Low quality or incomplete proteins were discarded for the further analyses.

### 2.12.2 Alignment of retrieved proteins

Retrieved homologues amino acid sequences were aligned using MAFFT (v7.221)(Kato et al., 2002; Kato and Standley, 2013).

The multialignment was performed using the parameters and the E-INS-i accuracy oriented method:

- *--ep* indicating the offset value for group-to-group alignment was set to 0 in order to allow large gaps and preserving the domains alignment;
- *--genafpair* specifying the local algorithm with the generalised affine gap cost (Altschul, 1998) to compute the pairwise alignment;

- *--maxiterate* indicating the number of cycle of iterative refinement to perform was set to 1000;

The used command line was:

```
MAFFT-einsi --ep 0 --genafpair --maxiterate 1000 input > output
```

Input and output files were in FASTA format.

The amino acidic multialignment was then converted into nucleotide alignment using the PAL2NAL tool (Suyama et al., 2006) for further analyses.

### **2.12.3 Phylogenetic reconstruction with Neighbour-joining method**

Phylogeny reconstruction of the homologues sequences was performed using MEGA v.6.06 (Tamura et al., 2013). Neighbour-joining method was used as statistical method, using the Maximum Composite Likelihood substitution model including transitions and transversions. Inferred trees were tested by bootstrap method (1000 replicates).

### **2.12.4 Selection of the best substitution model**

Statistical selection of the best-fit models of nucleotide substitution was carried out by using JModelTest (Posada, 2008). Likelihood calculation was performed on seven substitution schemes (56 substitution models) with rate variation parameters +I, +G, and nCat=4. The base tree for likelihood calculations was optimised using maximum likelihood and the base tree search was the best between Nearest-Neighbour Interchange (NNI) and the Subtree-Pruning-Regrafting (SPR) heuristic searches. Once the likelihoods were calculated, the substitution model was chosen according to the Akaike Information Criterion (AIC).

### **2.12.5 Phylogenetic reconstruction by Maximum-Likelihood method**

Maximum-likelihood phylogenesis was estimated by using PhyML software (Guindon et al., 2010;). The input multialigned sequence used was the JModeltest output in phylip format.

Parameters used were:

- *-i* specifying the input file;
- *-d* indicating the data type was set nt;

- *-n* specifying the number of dataset was set 1;
- *--run\_id* attributing a name to the analysis;
- *-b* specifying the approximate Likelihood-Ratio Test was set as SH-like branch supports alone (-4)
- *--run\_id*
- *-m* specifying the nucleotide substitution model, was set as resulted from jModelTest AIC calculations;
- *-f* was set *m* to estimate equilibrium frequencies using maximum likelihood;
- *-v* indicating the proportion of invariable sites was set to *e* in order to estimate this parameter by maximising the likelihood of the phylogeny;
- *-c* indicating the number of substitution rate categories was set to 4 as default;
- *-a* specifying the shape of gamma distribution if required was set to be estimated (*e*);
- *-o* indicates the output file;
- *-s* defining the topology search operation set to BEST to chose the best one between Nearest Neighbour Interchange (NNI) and Subtree Pruning and Regrafting (SPR) search.

The general command line used was:

```
./phymml -l <inputfile> -d nt -n 1 -b -4 --run_id "name analysis" -m "model code" -f m -v e -c 4 -a e -no_memory_check -o <output file> -s BEST.
```

### 2.12.6 Phylogenetic analysis by Maximum-Likelihood (PAML)

Variability analyses were performed using the tool *codeml* in the Phylogenetic Analysis by Maximum-Likelihood (PAML 4) package (Yang, 2007).

*Codeml* required the multialigned proteins in *phylip* format and the Maximum Likelihood phylogenetic tree in *newick* format as inputs. The program was run using the *.ctl* file edited appropriately as shown in Figure 2.1. The four models of natural selection required different parameters as summarised in Table 2.17. Comparisons between the null and selection models were performed by Likelihood Ratio Tests and  $\chi^2$  calculations to select the model with the best fit. In order to run branch-site models, the phylogenetic tree was edited specifying the examined lineage with "#". Sites under positive selective pressures were selected with the Bayes Empirical Bayes (BEB) analyses (Yang et al., 2005).

Parameters	Site models		Branch-site models	
	m1	m2a	A <sub>0</sub>	A
<i>model</i>	0	0	2	2
<i>Nsite</i>	0	2	2	2
<i>fix_omega</i>	0	0	1	0
<i>omega</i>	0.4	0.4	0.4	1

**Table 2.17 Parameters of the models used for Phylogenetic Analysis by Maximum-Likelihood (PAML).**

```

seqfile = sequence data filename
treefile = tree structure file name
outfile = main result file name

noisy = 3 * 0,1,2,3,9: how much rubbish on the screen
verbose = 0 * 0: concise; 1: detailed, 2: too much
runmode = 0 * 0: user tree; 1: semi-automatic; 2: automatic
          * 3: StepwiseAddition; (4,5):PerturbationNNI; -2: pairwise

seqtype = 1 * 1:codons; 2:AAs; 3:codons-->AAs
CodonFreq = 3 * 0:1/61 each, 1:F1X4, 2:F3X4, 3:codon table

*
  ndata = 1
  clock = 0 * 0:no clock, 1:clock; 2:local clock; 3:CombinedAnalysis
  aaDist = 0 * 0:equal, +:geometric; -:linear, 1-6:G1974,Miyata,c,p,v,a
  aaRatefile = dat/jones.dat * only used for aa seqs with model=empirical(_F)
                * dayhoff.dat, jones.dat, wag.dat, mtmam.dat, or your own

  model = 0
          * models for codons:
            * 0:one, 1:b, 2:2 or more dN/dS ratios for branches
          * models for AAs or codon-translated AAs:
            * 0:poisson, 1:proportional, 2:Empirical, 3:Empirical+F
            * 6:FromCodon, 7:AAClasses, 8:REVaa_0, 9:REVaa(nr=189)

  NSsites = 0 * 0:one w;1:neutral;2:selection; 3:discrete;4:freqs;
              * 5:gamma;6:2gamma;7:beta;8:beta&w;9:beta&gamma;
              * 10:beta&gamma+1; 11:beta&normal>1; 12:0&2normal>1;
              * 13:3normal>0

  icode = 0 * 0:universal code; 1:mammalian mt; 2-10:see below
  Mgene = 0
          * codon: 0:rates, 1:separate; 2:diff pi, 3:diff kapa, 4:all diff
          * AA: 0:rates, 1:separate

  fix_kappa = 0 * 1: kappa fixed, 0: kappa to be estimated
  kappa = 2 * initial or fixed kappa

```

**Figure 2.1 Phylogenetic Analysis by Maximum-Likelihood (PAML) control file.** PAML analysis was carried out using the parameters as indicated in the control (.ctl) file.

### 2.13 Statistical data analyses

Statistical analyses were performed using GraphPad Prism v.6.0g for Mac OS X, GraphPad Software, La Jolla California USA, [www.graphpad.com](http://www.graphpad.com).

Correlation analyses were performed on the datasets deriving from SILAC-RIME experiments. Pearson correlation coefficients ( $r$ ) and two-tailed  $P$  values were computed to infer statistical significance.

One-way ANOVA followed by Dunnett's multiple comparisons tests were performed to compute statistical significance of the yeast two-hybrid  $\beta$ -galactosidase assays data and the RNA tethering assays data.

Phylogenetic trees reconstructed by Neighbour-joining method were tested by bootstrap tests (1000 replicates) using MEGA v.6.06 (Tamura et al., 2013). Phylogenetic trees reconstructed by Maximum-Likelihood method were tested by SH-like bootstrap test (1000 replicates) using PhyML (Guindon et al., 2010).

Likelihood Ratio Test and  $\chi^2$  calculations were used to select the selection model with the best fit using PAML4 (Yang, 2007).

# **Chapter 3**

## **The interactome of the nuclease module of the Ccr4-Not complex**

## Chapter 3. The interactome of the nuclease module of the Ccr4-Not complex

### 3.1 Introduction

Degradation of mRNA plays a central role in regulation of gene expression and in quality control of mRNA biogenesis.

The shortening of poly(A) tail (deadenylation) in most cases is the first step of cytoplasmic mRNA degradation (Yamashita et al., 2005; Zheng et al., 2008). In the traditional view, mature mRNAs have long poly(A) tails ranging from 70 to 80 nucleotides (nt) in *S. cerevisiae*, and from 200 to 250 nt in mammalian cells (Chen and Shyu, 2011). However, more recent studies based on direct sequencing techniques suggest that the poly(A) tail is noticeably shorter. Both using TAIL-SEQ and PAL-SEQ methods, which were developed in the Kim and Bartel labs, respectively, it was found that the medium tail length in mammalian cells is just under 100 nt long (Chang et al., 2014; Subtelny et al., 2014). In mammals, deadenylase enzymes are required to shorten the poly(A) tails, triggering decapping and exposing the mRNA to the Xrn1 or exosome ribonucleases (Couttet et al., 1997; Meyer et al., 2004). Mammalian deadenylation is a process with biphasic kinetics (Yamashita et al., 2005; Bartlam and Yamamoto, 2010). According to one model, the first phase is a slow deadenylation led by the Pan2-Pan3 complex that shortens the poly (A) tail to approximately 110 nt without degradation of the mRNA body. The second phase involves the Ccr4-Caf1 complex that shortens the poly(A) tail to 10 nucleotides, in this phase degradation of the mRNA body was observed (Yamashita et al., 2005). The Ccr4-Caf1 and Pan2-Pan3 complex can form a large complex on mRNA in the cytoplasm that coordinates the two phases (Zheng et al., 2008). The nuclease subunits of the Ccr4-NOT complex underwent to a duplication event in vertebrates; CNOT6 and CNOT6L are orthologues of the Ccr4 subunit and CNOT7 and CNOT8 are orthologues of the Caf1 subunit. In addition to the catalytic subunits, the Ccr4-Not and Pan2-Pan3 complexes contains non-catalytic subunits whose role in mRNA decay is not completely understood.

Deadenylation is finely regulated on global level and on mRNA specific level. As soon the mRNA is imported from the nucleus in the cytoplasm, the poly(A) tail undergoes to a basal shortening.

The deadenylation of specific mRNAs can be triggered by destabilising elements with a protein-mediated recruitment of the deadenylase complex (Goldstrohm and Wickens, 2008).

AU-rich elements (AREs) are the most common destabilising elements found in the mRNA (Chen et al., 1995; Barreau et al., 2005). AREs are sequence regions rich in adenosines and uridines found in the 3'UTR of the mRNAs (Wu and Brewer, 2012; Chen et al., 1995). The protein Tristetraprolin (TTP) associates with the AREs and promotes the mRNA degradation directly binding the central scaffold CNOT1 of the Ccr4-Not complex (Lykke-Andersen and Wagner, 2005; Fabian et al., 2013).

Another mRNA decay pathway that requires interacting factors for the recruitment of the deadenylase complexes is the microRNA-mediated decay. Most miRNAs repress gene activity by base-pairing to target mRNAs triggering degradation (Bushati and Cohen, 2007; Chen and Shyu, 2011). The miRNA-mediated decay requires Argonaute (AGO) proteins, the GW182 that recruits the Ccr4-Not, and the Dcp1-Dcp2 complexes (Behm-Ansmant et al., 2006).

The study of the interactions of the Pan2-Pan3 and Ccr4-NOT complexes led to the characterisation of the deadenylation role in different regulatory pathways. The work described in this chapter aimed to identify new partners that interact with Ccr4 and Caf1, as well as with Pan2 and Pan3 to determine new pathways that trigger mRNA degradation. Stable Isotope Labelling with Amino acids in Cell culture (SILAC)-Rapid Immunoprecipitation and Mass spectrometry of Endogenous proteins (RIME) together with bioinformatics tools for protein-protein network generation and exploration (Cytoscape) were used to discover the Ccr4 and Caf1 interactome. Pan2-Pan3 preliminary experiments to discover its interactome were performed. This study identified unique and shared interaction partners of the deadenylase subunits. Moreover, new interaction partners for the Ccr4 and Caf1 subunit were identified.

### **3.2 Stable Isotope Labelling with Amino acids in Cell culture-Rapid Immunoprecipitation and Mass spectrometry of Endogenous proteins (SILAC-RIME) of the Ccr4-Not nuclease module**

Stable Isotope Labelling with Amino acids in Cell culture (SILAC)-Rapid Immunoprecipitation and Mass spectrometry of Endogenous proteins (RIME) was used to discover the interactome of the Ccr4 and Caf1 subunits of the Ccr4-Not complex. The approach is based on the *in vivo* labelling of endogenous proteins by incorporation of amino acids with substituted stable isotopic nuclei. Two populations of cells are grown in culture media containing amino acids with the “light” or “heavy” isotopes. The labelled amino acids are incorporated in the newly synthesized proteins. This technique, coupled to affinity purification and mass spectrometry allowed to determine the interactome of several proteins (Fang et al., 2008; Trinkle-Mulcahy et al., 2008; Zhong et al., 2011; Foster et al., 2013) (Figure 3.1).

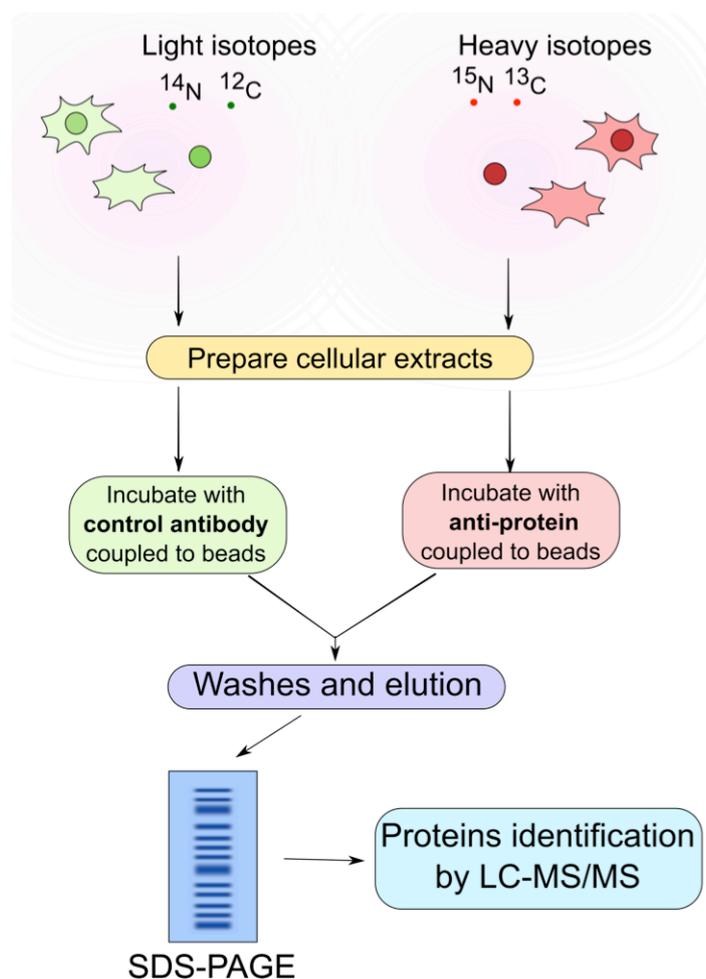
This approach was used to discover the interactome of the Caf1 (CNOT7 and CNOT8) and Ccr4 (CNOT6 and CNOT6L) deadenylase subunits and was performed in collaboration with Dr. Klaas Mulder (Faculty of Science, Radboud Institute for Molecular Life Science, Radboud University, Netherlands).

MCF7 cells were labelled by growth in culture medium containing  $^{13}\text{C}_6$   $^{15}\text{N}_2$ -L-lysine and  $^{13}\text{C}_6$   $^{15}\text{N}_4$ -L-arginine and are referred hereafter as heavy SILAC labelled cells. In parallel, cells were grown in culture media containing  $^{12}\text{C}_6$   $^{14}\text{N}_2$ -L-lysine and  $^{12}\text{C}_6$   $^{14}\text{N}_4$ -L-arginine, and are hereafter referred as light SILAC labelled cells. Following the labelling and crosslinking, cellular extracts were obtained and immunoprecipitations with magnetic beads were performed. Light SILAC cell extracts were used for control immunoprecipitations with purified pre-immune rabbit IgG. Heavy SILAC labelled cell extracts were used for immunoprecipitations using a mix of CNOT6 and CNOT6L antibodies, or a mix of CNOT7 and CNOT8 antibodies. Cytoplasmic and nuclear extracts were initially analysed separately. Following RIME immunoprecipitations, the eluted

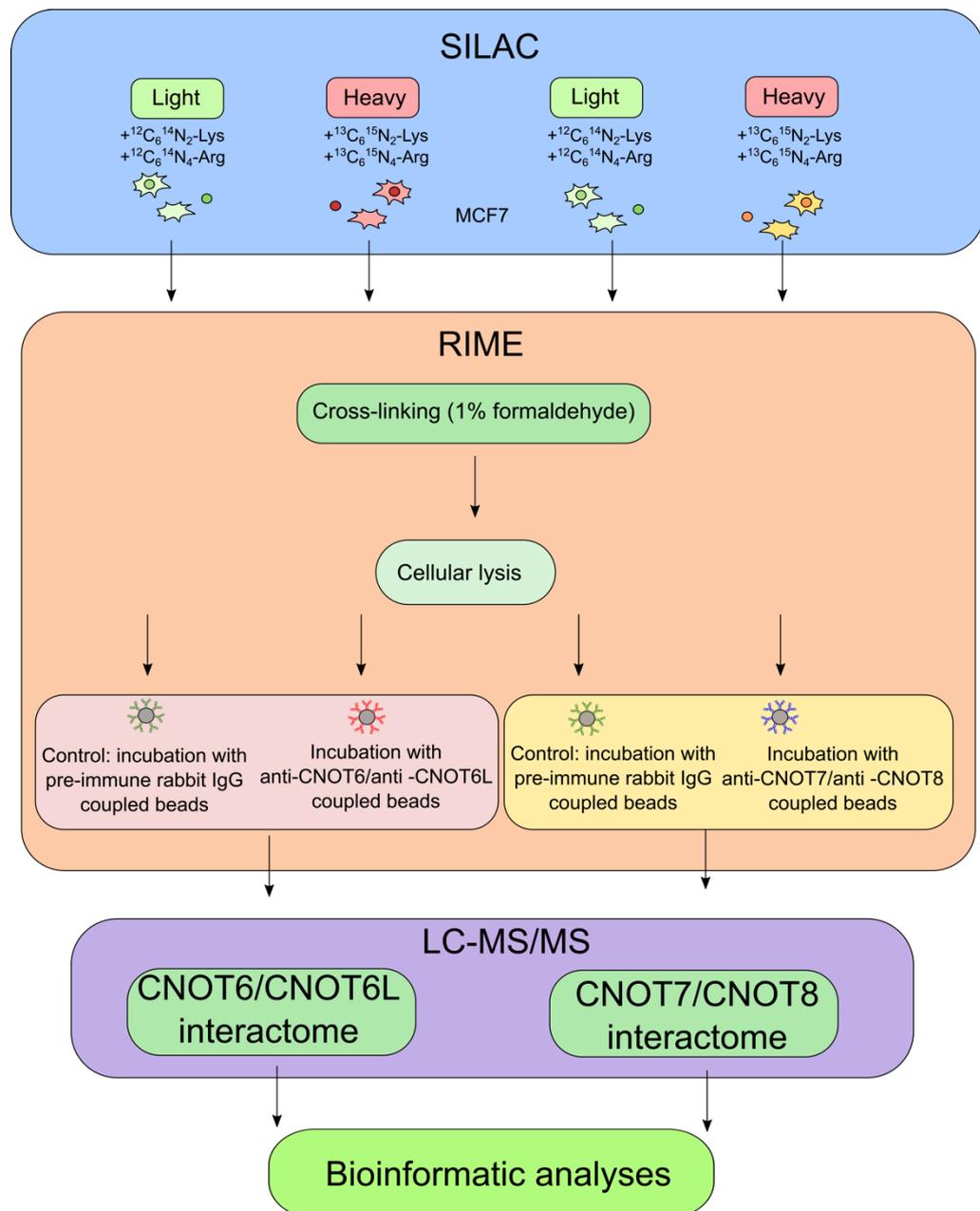
proteins were analysed by LC-MS/MS mass spectrometry for the identification of unique peptides.

A schematic overview of the approach is represented in Figure 3.2.

The advantage of this approach relies on its high confidence of detecting not only direct but also low affinity and indirect protein-protein interactions preserving the cellular localisation.



**Figure 3.1 Stable isotope labelling with amino acids in cell culture (SILAC).** SILAC is based on the in vivo labelling of endogenous proteins by incorporation of amino acids with substituted stable isotopic nuclei. Two populations of cells are grown in the presence of amino acids with the “light” isotope or the “heavy” form of the amino acids. The labelled amino acids are incorporated in the newly synthesized proteins. Proteins are purified by affinity purification by incubating cellular extracts with control antibody or with specific anti-protein antibody. Eluted proteins are identified by LC-MS/MS.



**Figure 3.2 SILAC-RIME approach to obtain Ccr4 and Caf1 interactome.**

MCF7 cells were grown in medium containing  $^{12}\text{C}_6^{14}\text{N}_2$ -lysine and  $^{12}\text{C}_6^{14}\text{N}_4$ -arginine for light labelling and  $^{13}\text{C}_6^{15}\text{N}_2$ -lysine and  $^{13}\text{C}_6^{15}\text{N}_4$ -arginine for heavy labelling (SILAC). The RIME consisted in cross-linking, cellular lysis and immunoprecipitations with magnetic beads coated with CNOT6 and CNOT6L antibodies or with CNOT7 and CNOT8 antibodies.

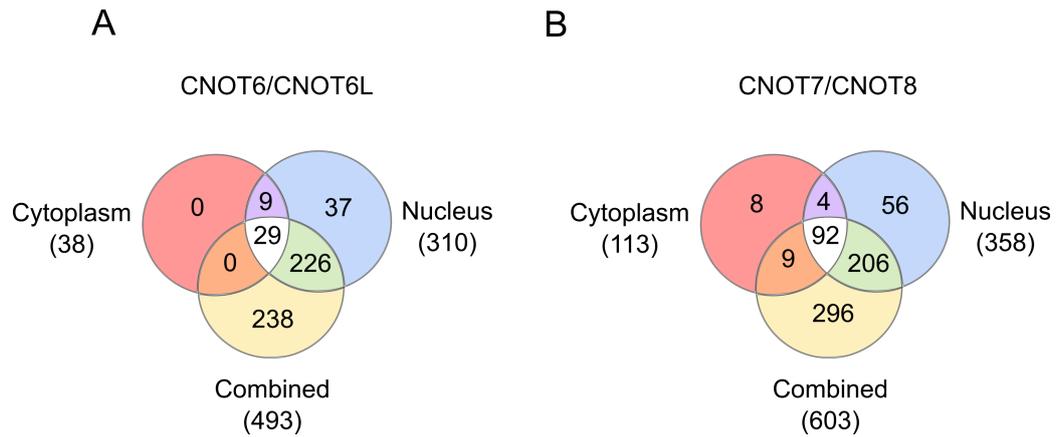
Immunoprecipitation with pre-immune rabbit IgG coated beads were used as control. Following RIME, the eluted proteins were analysed by LC-MS/MS for the identification of unique peptides and results were analysed by computational tools.

### 3.3 Visualisation of SILAC-RIME data

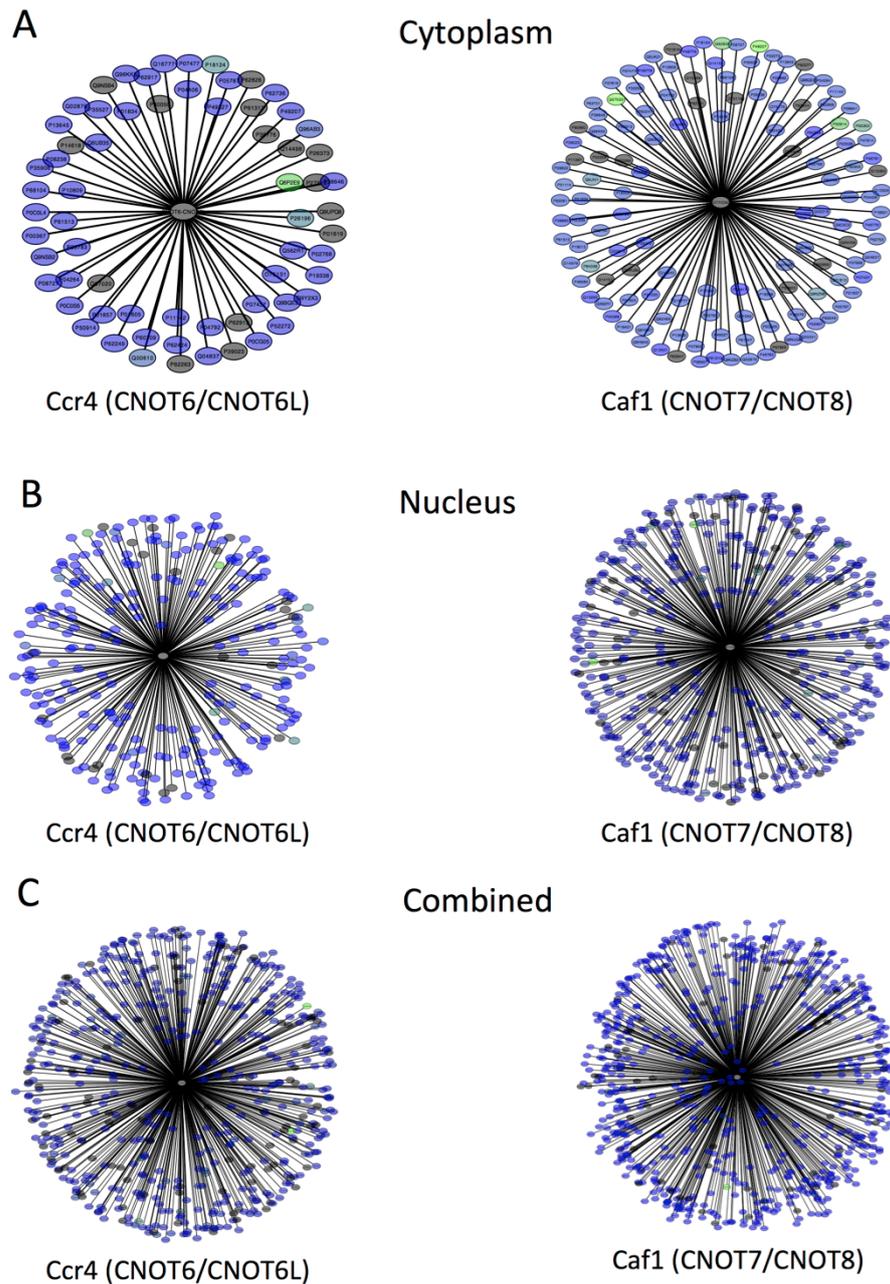
Two SILAC-RIME experiments were performed. In the first experiment, the nuclear and cytoplasmic fractions were analysed separately. In the second experiment, the nuclear and cytoplasmic extracts were combined before immunoprecipitation analysis. Using CNOT6-CNOT6L antibodies, 38 interacting proteins were detected in the cytoplasmic fraction and 310 interacting proteins in the nuclear fraction. Using a mixture of CNOT7 and CNOT8 antibodies, 113 interacting proteins were detected in the cytoplasmic fraction and 358 interacting proteins in the nuclear fraction. When the nuclear and cytoplasmic extracts were combined, 493 interacting proteins were detected upon immunoprecipitation using CNOT6 and CNOT6L antibodies, and 603 interacting proteins using a mixture of CNOT7 and CNOT8 antibodies (Figure 3.3).

The SILAC-RIME data, analysed using Proteome Discoverer v.1.3 that only uses unique peptides for ratio analysis, is reported in Supporting Table 3.1. To obtain a visual overview of the immunoprecipitated proteins, the Cytoscape package was used. Nodes representing proteins were coloured according to the heavy/light chain ratio resulting from the SILAC analysis using continuous mapping from blue indicating lower values to green indicating higher values as shown in Figure 3.4. Black nodes indicated proteins with no heavy/light ratio resulting from SILAC-RIME analyses and were discarded from further analysis.

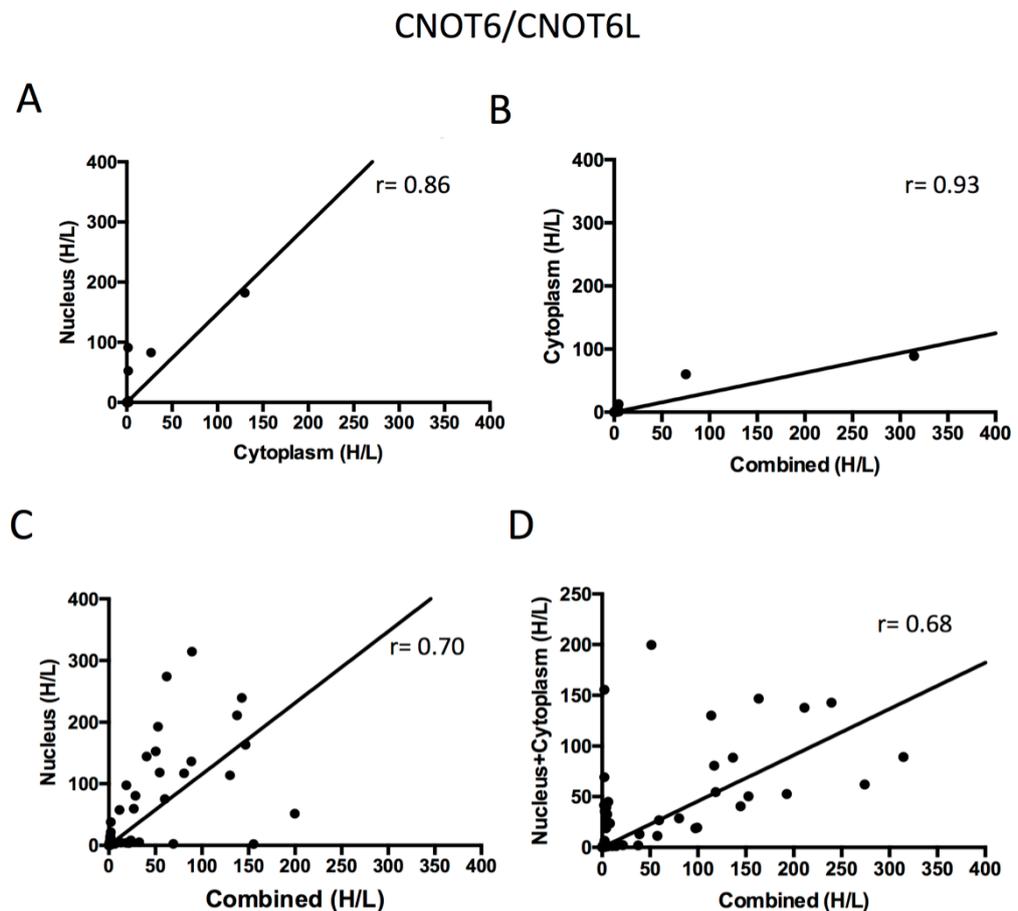
To examine the relatedness between the data sets, correlation analyses between the nuclear, cytoplasmic and combined fractions were performed. Correlation in the dataset with a P value  $<0.0001$  was found for the CNOT6/CNOT6L data between the nucleus and cytoplasm, between the cytoplasm and the combined fraction, between the nucleus and the combined fractions, and between the union of the cytoplasm and nucleus fractions interrogated separately and the combined fraction (Figure 3.5 A, B, C, and D respectively). The CNOT7/CNOT8 data was found to be correlated with a P value  $<0.0001$  (Figure 3.6 A, B, C, and D). Based on these findings, further analyses were performed on the combined fraction dataset representative of both the nuclear and cytoplasmic fractions.



**Figure 3.3 Venn diagrams indicating overlap between SILAC-RIME data sets.** The table summarises the number of proteins co-immunoprecipitated with CNOT6-CNOT6L and CNOT7-CNOT8 and detected in the cytoplasmic, nuclear, and combined fractions. Proteins found in more than one fraction are grouped in overlaps of the circles in Venn diagrams.

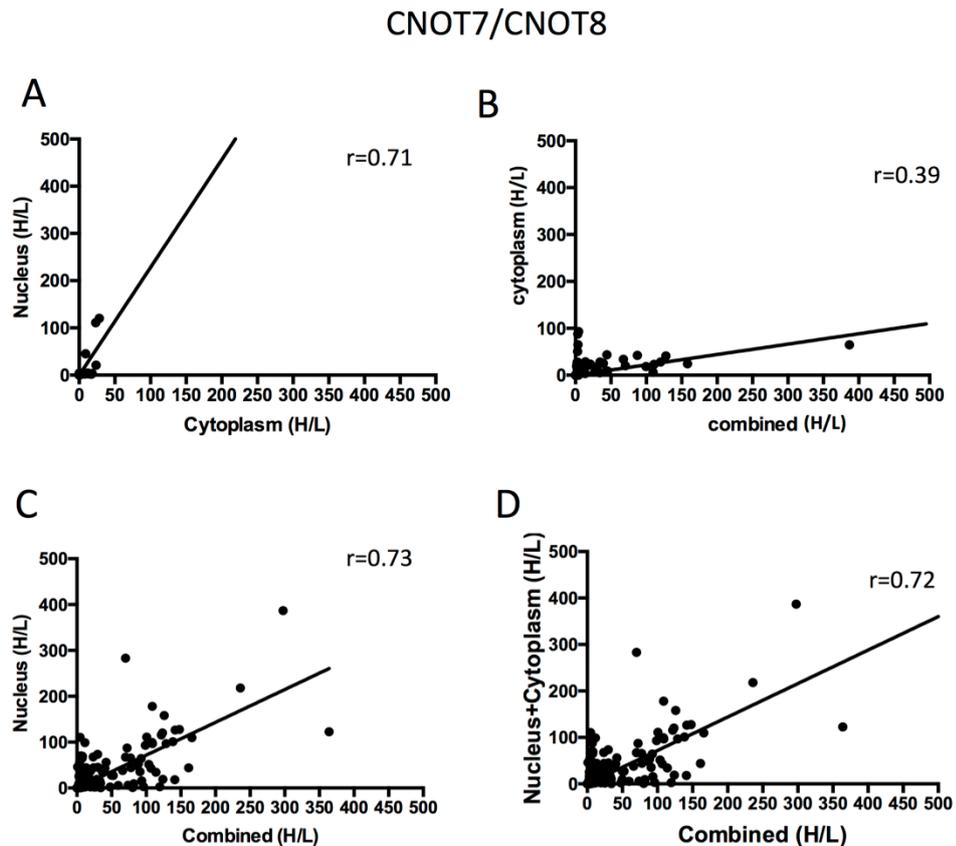


**Figure 3.4 SILAC-RIME results visualised by Cytoscape.** Cytoscape was used to visualise SILAC-RIME results of the Ccr4 (CNO6/CNOTT6L) and Caf1(CNOT7/CNOT8) co-immunoprecipitated proteins in **(A)** the cytoplasmic fraction **(B)**, in the nuclear fraction, and **(C)** in the combined fraction. Proteins were coloured as a measure of their heavy/light peptide ratio with continuous mapping from blue indicating lower values to green indicating bigger values. Nodes with no heavy/light ratio resulting from the analyses were coloured in black and were discarded from subsequent analyses.



**Figure 3.5 Correlation between CNOT6/CNOT6L**

**co-immunoprecipitated proteins from the nuclear, cytoplasmic and combined fractions.** The CNOT6/CNOT6L interactome was investigated in nuclear, cytoplasmic and combined extracts. Pearson correlation coefficients with P values  $<0.0001$  were found between Heavy/Light ratios (H/L) of the CNOT6/CNOT6L co-immunoprecipitated proteins in **(A)** the nuclear fraction and the cytoplasmic fraction; **(B)** the cytoplasmic fraction and the combined fraction; **(C)** the nucleus fraction and the combined fraction; and **(D)** between the union of the nuclear and cytoplasm fractions investigated separately and the combined fraction.



**Figure 3.6 Correlation between CNOT7/CNOT8 co-immunoprecipitated proteins from the nuclear, cytoplasmic and combined fractions.** The CNOT7/CNOT8 interactome was investigated in nuclear, cytoplasmic and combined extracts. Pearson correlation coefficient with P values  $<0.0001$  were found between Heavy/Light ratios (H/L) of the CNOT7/CNOT8 co-immunoprecipitated proteins in **(A)** the nuclear fraction and the cytoplasmic fraction; **(B)** the cytoplasmic fraction and the combined fraction; **(C)** the nucleus fraction and the combined fraction;; and **(D)** between the union of the nuclear and cytoplasm fractions investigated separately and the combined fraction.

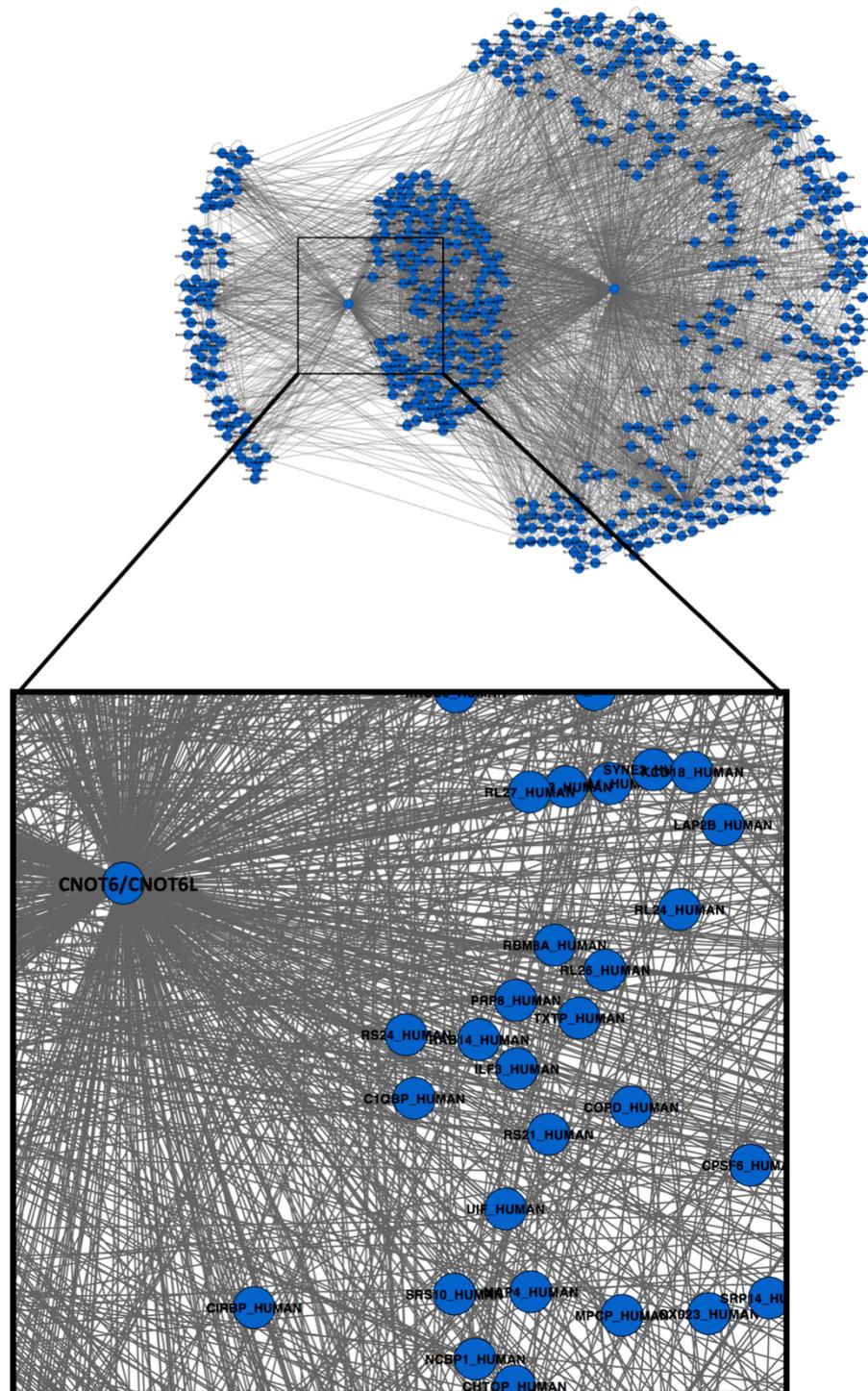
### **3.4 Combined protein-protein interaction network of the nuclease module**

Immunoprecipitates derived from total cell lysates contained the largest number of unique proteins and were significantly overlapping with the proteins immunoprecipitated using both the nuclear and cytoplasmic fractions. Thus, the attention was focussed on the networks obtained using total cell lysates.

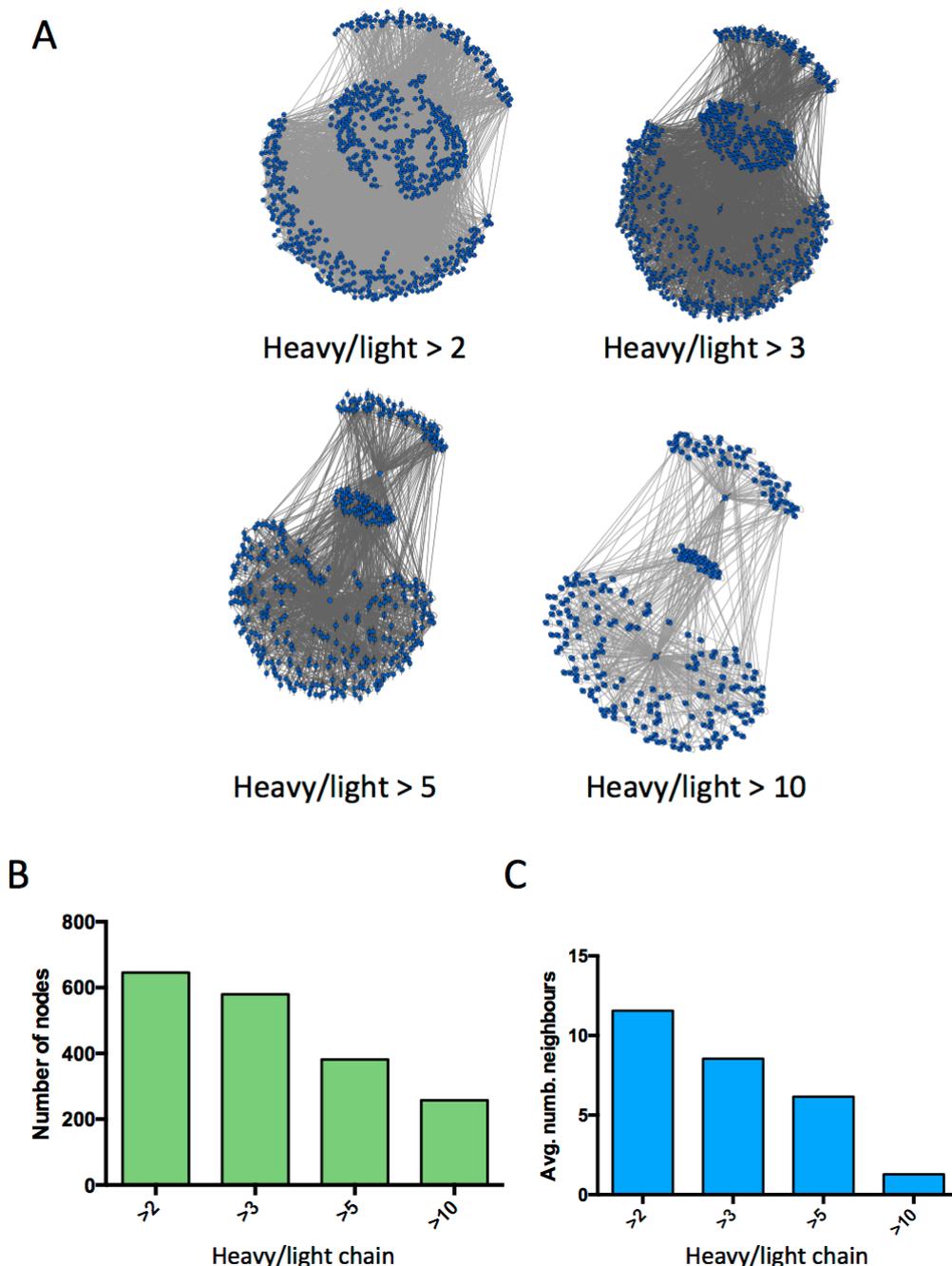
In order to obtain insight into all proteins interacting with the nuclease module, the Caf1 (CNOT7/CNOT8) and Ccr4 (CNOT6/CNOT6L) protein interaction networks were merged and visualised using Cytoscape (Shannon et al., 2003). In the network, proteins were represented as nodes and protein-protein interactions as edges between nodes (Figure 3.7). The generated network was further explored using Cytoscape plug-ins. The Bisogenet plug-in was used to integrate data from known protein-protein interaction databases (Martin et al., 2010).

Four different networks were generated by applying different cut-off values based on the relative abundance in the immunoprecipitates obtained using specific and control antibodies as evident from the heavy/light peptide ratios of the SILAC data (Figure 3.8 A). As expected, increasing the cut-off value using the heavy/light peptide ratios, the number of nodes was reduced thereby simplifying the network topology, but increasing information loss. Using heavy/light ratios  $>2$  as the cut-off threshold, the resulting network was composed of 645 nodes; selecting heavy/light ratios  $>3$  as the cut-off threshold, 579 nodes constituted the network; when heavy/light ratios  $>5$  were selected as the cut-off value, only 381 nodes were present; setting an heavy/light ratio cut-off at 10, the number of nodes constituting the network was 257 (Figure 3.8 B). Along with the number of nodes, the average number of connected nodes termed neighbours -hence the number of interactions per protein- also decreased when the cut-off value based on the heavy/light peptide ratios was increased (Figure 3.8 C). The resulting networks were manually inspected and evaluated in order to select the most informative one. The network generated with proteins with a heavy/light peptide ratio  $>3.0$  was considered as the

most biologically informative and was chosen to perform subsequent analyses.



**Figure 3.7 Protein-protein interactions network.** Proteins in the generated networks are represented as nodes and protein-protein interactions are represented as grey edges.

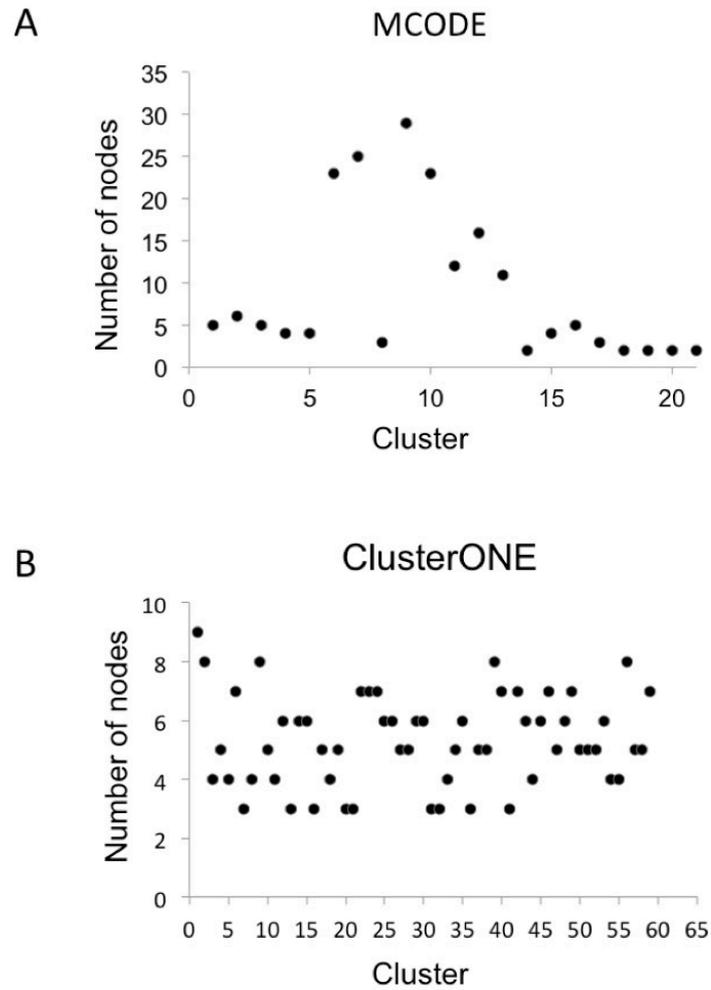


**Figure 3.8 Nuclease module interactome networks. (A)** Four different networks were generated applying different cut-off values based on the heavy/light peptide ratios of the SILAC-RIME data. **(B)** Increasing the heavy/light peptide ratio values, many nodes were lost simplifying the network topology at the expense of information loss. **(C)** Along with the number of nodes, the average number of neighbour nodes decreased when the heavy/light peptide ratio cut-off values increased.

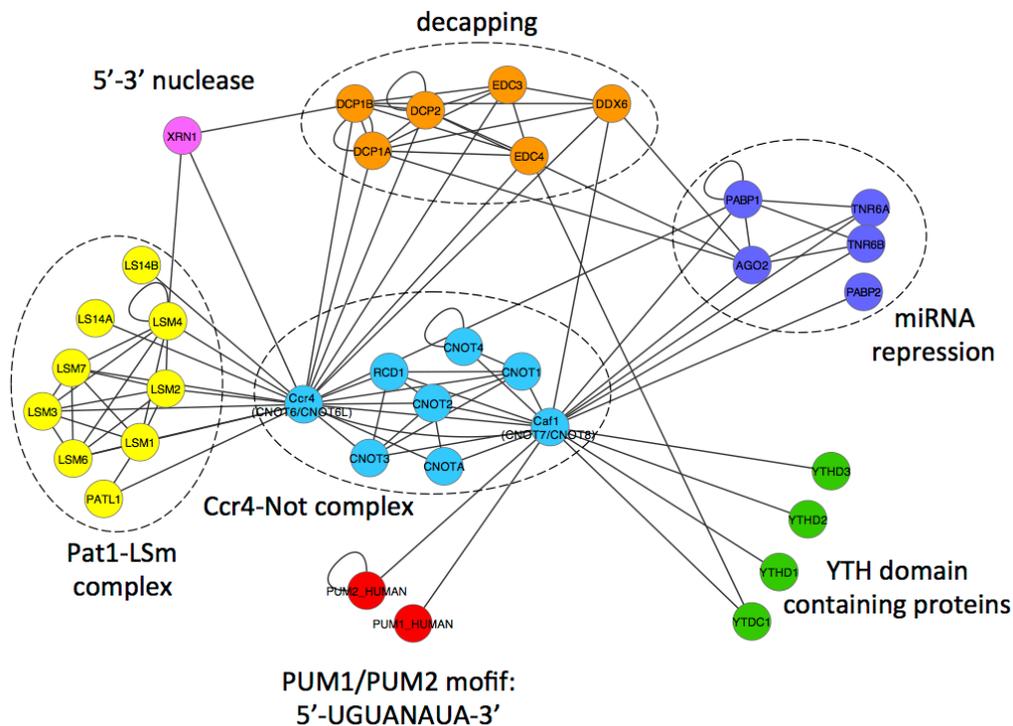
### 3.4.1 Clustering analyses

The protein-protein interaction network constituted by proteins with a heavy/light peptide ratio  $> 3.0$ , was inspected with MCODE (Bader et al., 2003) and ClusterONE (Zhang et al., 2014). The two Cytoscape plug-ins use unique methods to find protein clusters in a protein-protein interaction network. The former searches for highly interconnected regions in the network, many of which correspond to known molecular complexes. ClusterONE is a plug-in detecting potential overlapping proteins complexes in networks.

MCODE clustering approach identified 21 different clusters (Figure 3.9 A). The content of nodes grouped in the clusters varied from three to 29. ClusterONE plug-in returned 59 clusters, which nodes content varied from three nodes to nine (Figure 3.9 B). This automated search for cluster was used as starting point for the network inspection. With this approach it was possible to identify most of the known interactions with the Ccr4-Not complex and new interacting partners that indicate the involvement of the deadenylase complex in cellular pathways not yet characterised. Figure 3.10 shows how the clustering analyses allowed the identification of the Pat1-Lsm complex, proteins involved in the decapping of the mRNA, proteins involved in miRNA repression, PUM proteins, and the new interacting partners YTH domain containing protein family members proteins. Both the approaches identified the Ccr4-Not component complex, the Lsm complex, and the decapping complex components.



**Figure 3.9 Clustering analyses of the Caf1-Ccr4 interactome.** The clustering analysis was performed using two Cytoscape plug-ins. **(A)** MCODE returned twenty clusters by searching for dense areas in the protein-protein interaction network. **(B)** ClusterONE detected 59 potential overlapping protein complexes.



**Figure 3.10 Ccr4-NOT complex interactome.** The interactions of the Caf1 and Ccr4 subunits of the Ccr4-Not complex were investigated with Cytoscape. Important biological interactions were found in the interactome such as the Pat1-LSm complex components, proteins involved in decapping, miRNA repression proteins, Pumilio proteins, and YTH domain containing proteins.

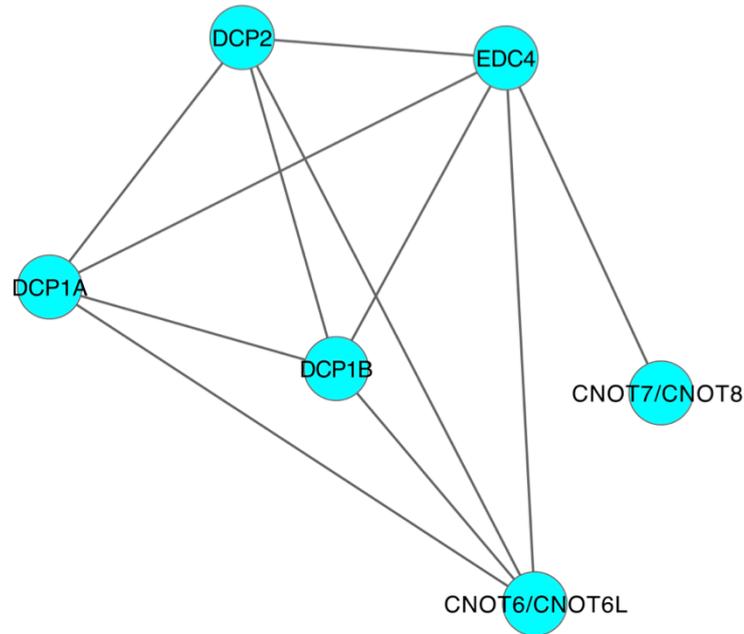
### 3.4.1.1 Molecular complexes identified by MCODE

The MCODE (Bader et al., 2003) analysis returned 21 clusters of proteins within the protein-protein interaction network and their content in terms of clustered proteins is shown in Supporting Table 3.2. Some of the clusters found are hereafter presented to support the validity of the clustering method, describing known functional clusters. The analysis detected decapping complex components to interact with the nuclease module (Figure 3.11). Ccr4 (CNOT6/CNOT6L) was found to interact with all the grouped proteins DCP1A and its paralogue DCP1B, DCP2, and EDC4 whereas Caf1 (CNOT7/CNOT8) was found to interact only with the enhancer of mRNA decapping protein EDC4 (Rehwinkel et al., 2005; Behm-Ansmant et al., 2006; Mathys et al., 2014; Nishimura et al., 2015; Ozgur et al., 2015a).

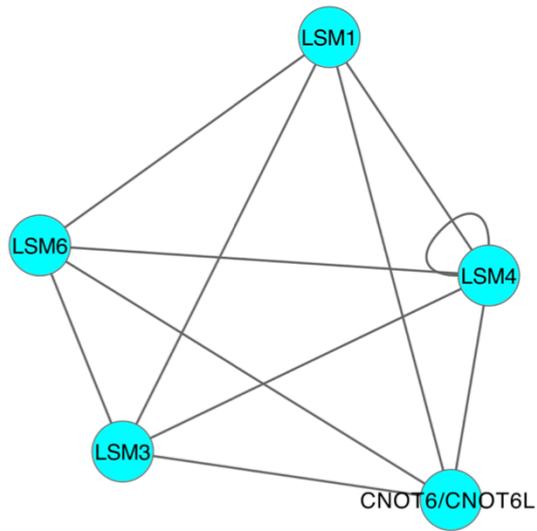
The LSm complex components LSM1, LSM3, LSM4, and LSM6 were also found to interact with the Ccr4 subunits (Figure 3.12) (Tharun et al., 2000; Yamashita et al., 2005; Ozgur et al., 2010; Totaro et al., 2011).

Furthermore, the clustering analysis identified the involvement of the Ccr4-NOT complex in the eukaryotic replication. Interactions with the mini-chromosome maintenance complex (MCM) components MCM2, MCM3, MCM5, MCM6, and MCM7 were found (Figure 3.13). Ccr4 interacted with all the five proteins of the complex, whereas Caf1 interacted with MCM7 and MCM3 subunits only. The MCM complex is an important component of the pre-replication complex and was found to be fundamental for the correct formation of the replication forks (Hyrien, 2016; Wu et al., 2014; Bell and Botchan, 2013). Recently the MCM complex interactome was studied by affinity and proximity purification coupled to mass spectrometry. The study demonstrated the interaction of the complex with proteins involved in DNA repair, in cell cycle progression, and in RNA processing (Dubois et al., 2016). Splicing factors and poly(A)-binding protein (PAPB1) were also demonstrated to interact with the MCM complex. These findings together with the SILAC-RIME results presented in this chapter may indicate an involvement of the MCM complex in the mRNA degradation. Further investigations are needed to

confirm the recruitment of the Ccr4-Not complex by MCM complex and to define the role of the latter in the mRNA processing.

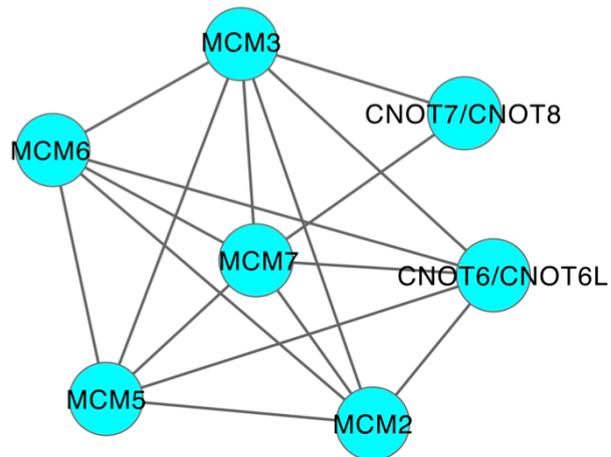


**Figure 3.11 Decapping complex identified by MCODE.** The MCODE clustering analyses detected components of the decapping complex. The CNOT6/CNOT6L subunits were found to interact with the decapping enzymes DCP1A and its paralogue DCP1B, with DCP2, and with the enhancer of mRNA decapping EDC4. The CNOT7/CNOT8 subunits were found to interact only with EDC4.



**Figure 3.12 Components of the LSm complex identified by MCODE.**

The MCODE plug-in detected LSm complex LSM1, LSM3, LSM4, and LSM6 components interacting with the Ccr4 subunits of the Ccr4-NOT complex.



**Figure 3.13 Mini-chromosome maintenance complex identified by MCODE.** MCODE plug-in identified the mini-chromosome maintenance complex components MCM2, MCM3, MCM5, MCM6 and MCM7. Although Ccr4 was found to interact with all five proteins, the Caf1 subunits were found to interact only with the MCM3 and MCM7 subunit of the MCM complex required for DNA replication and cell proliferation.

### 3.4.1.2 Molecular complexes identified by ClusterONE

ClusterONE (Zhang et al., 2014) returned 59 clusters and their content in terms of clustered proteins is shown in Supporting Table 3.3. Hereafter are presented some of the molecular clusters found in the protein-protein interaction network as support for this type of clustering method. The remaining clusters will not be discussed. The analysis identified the Ccr4-Not complex among all the proteins constituting the network as shown in Figure 3.14.

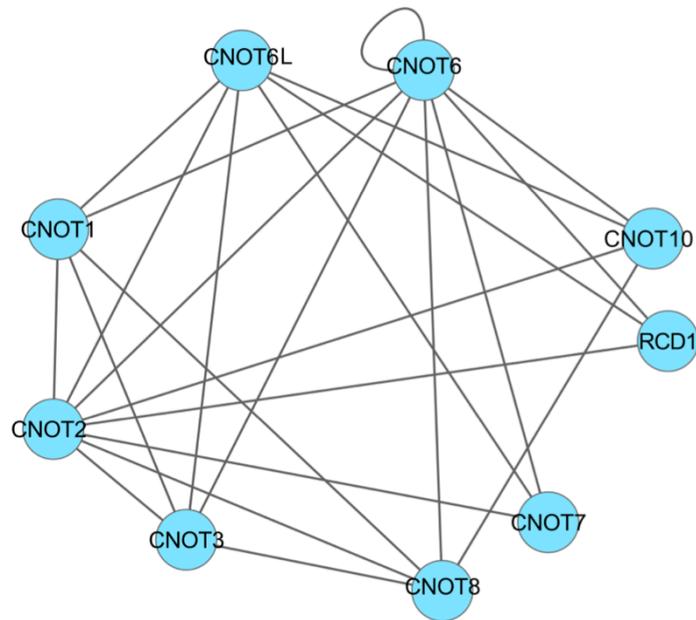
Another cluster found was constituted by the LSm complex subunits LSM1, LSM2, LSM3, LSM4, and LSM6 grouped together with the small nuclear ribonucleoprotein polypeptides (SNRP) SMD2, RUXE, and RUXF (Figure 3.15). Caf1 was found to interact with the SNRP peptides D2, E, and F but not with the LSm complex subunits. Ccr4 interacted with the LSm complex subunits and RUXE. Contacts between the SNRP peptides and LSm complex subunits were also detected (Tharun et al., 2000; Yamashita et al., 2005; Azzouz et al., 2009; Ozgur et al., 2010; Totaro et al., 2011).

Interestingly, the analysis grouped polypyrimidine binding proteins interacting with Caf1 and Ccr4 (Figure 3.16). The polypyrimidine binding protein PTBP1 was grouped with the poly(rC) binding proteins PCBP1, PBC2, and HRNPK. These proteins represents the major cellular polypyrimidine binding proteins and are associated with pre-mRNA processing and mRNA transport and metabolism (Patton et al., 1991; Zhou et al., 2010; Fukuda et al., 2009; Choi et al., 2009). In this cluster is present as well the ATP-dependent DNA helicase RUVB1, which contacts PCBP1, PBC2, and Caf1. The Ccr4 subunit interacted with all the four polypyrimidine binding proteins but not with the ATP-dependent DNA helicase.

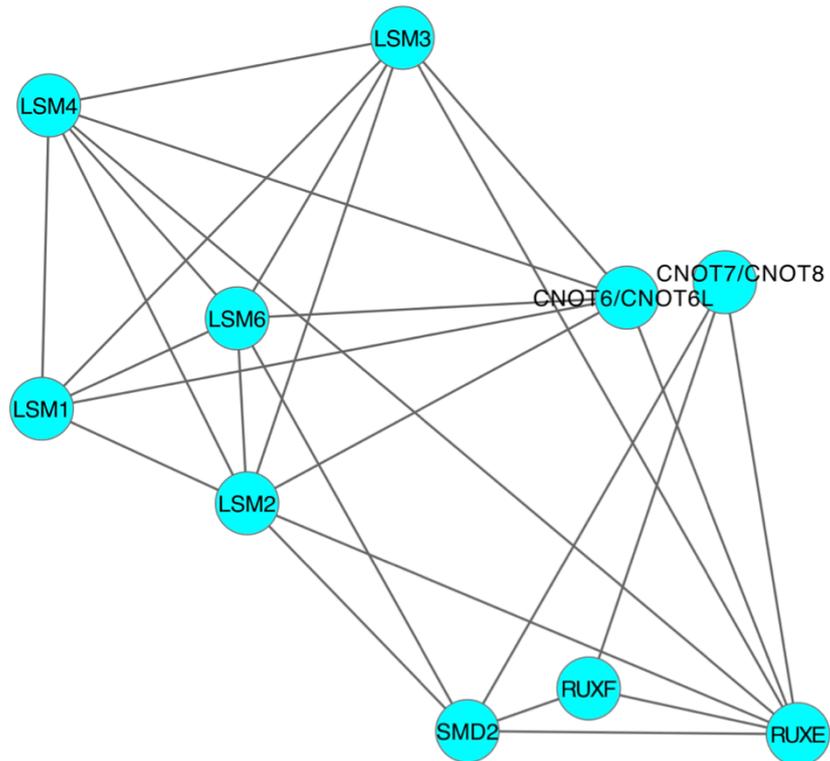
A further cluster found was constituted by the four subunits of the coatomer protein complex COPA, COPB2, COPD, and COPG interacting with the Caf1 subunit of the Ccr4-NOT complex (Figure 3.17). This multi-

subunit complex is important for intracellular transport by initiating the formation of transport vesicles and cargo sorting (Wang et al., 2016).

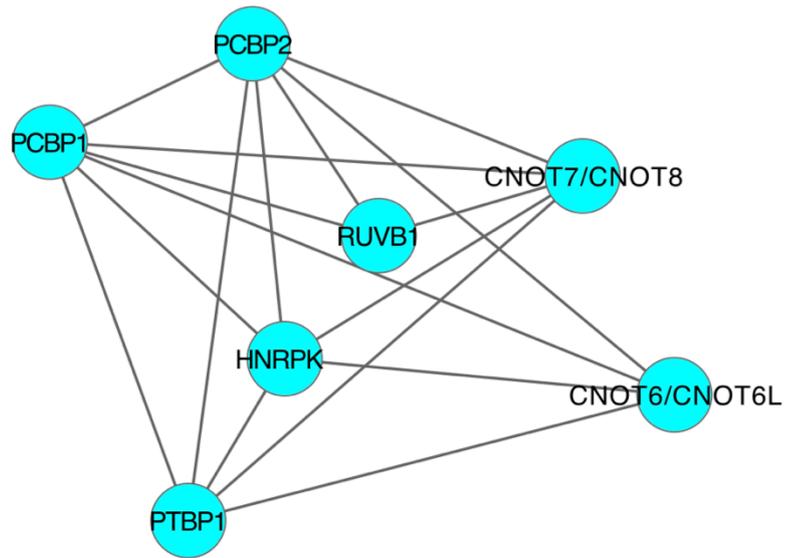
Importantly, a more heterogeneous cluster found was composed by Serine/Arginine-Rich splicing factors SRF1 and SRF2, the U2 small nuclear RNA auxiliary factor 1 (U2AF1), the polymerase (DNA) Delta interacting protein 3 (PDIP3), and the YTH domain containing protein 1 (YTHDC1). The polymerase (DNA) Delta interacting Protein 3 resulted to interact with both SRSF2 and Caf1. SRSF1 and SRF2 interacted with the YTHDC1 putative splicing factor. Ccr4 was found to interact only with the Serine/Arginine-Rich splicing factors SRF1 and SRF2, and with U2AF1 (Figure 3.18). The PDIP3 protein is a component of the TREX complex involved in mRNA export (Folco et al., 2012). Interestingly, YTHDC1 is a protein characterised by the presence of the YTH domain that specifically binds N<sup>6</sup>-methyladenosine (m<sup>6</sup>A). Recently, a key role for this protein was demonstrated in alternative splicing as well as the regulation of RNA turnover and RNA translation (Wang et al., 2014; Adhikari et al., 2016; Roundtree and He, 2016; Xiao and Guifang, 2016).



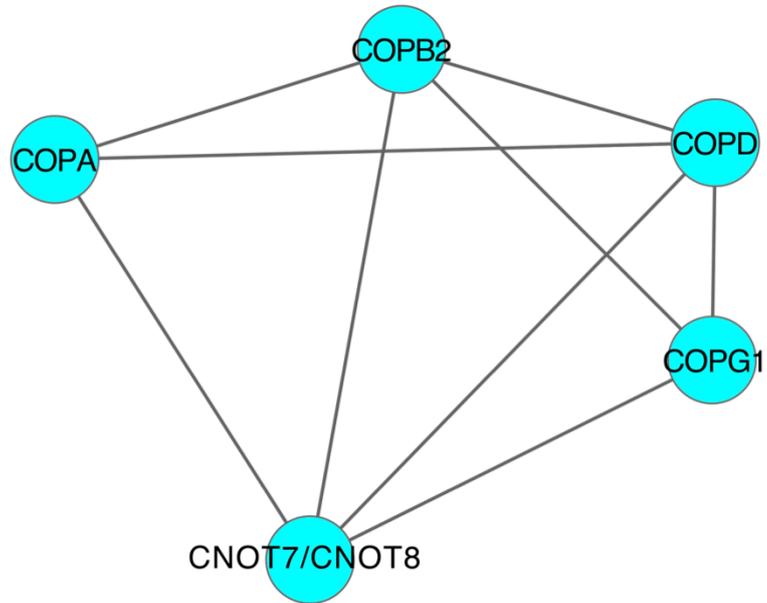
**Figure 3.14 The Ccr4-Not complex identified using ClusterONE.** The clustering analysis performed by ClusterONE identified the components of the Ccr4-Not complex CNOT1, CNOT2, CNOT3, CNOT6 and CNOT6L, CNOT7 and CNOT8, RQCD1, and CNOT10.



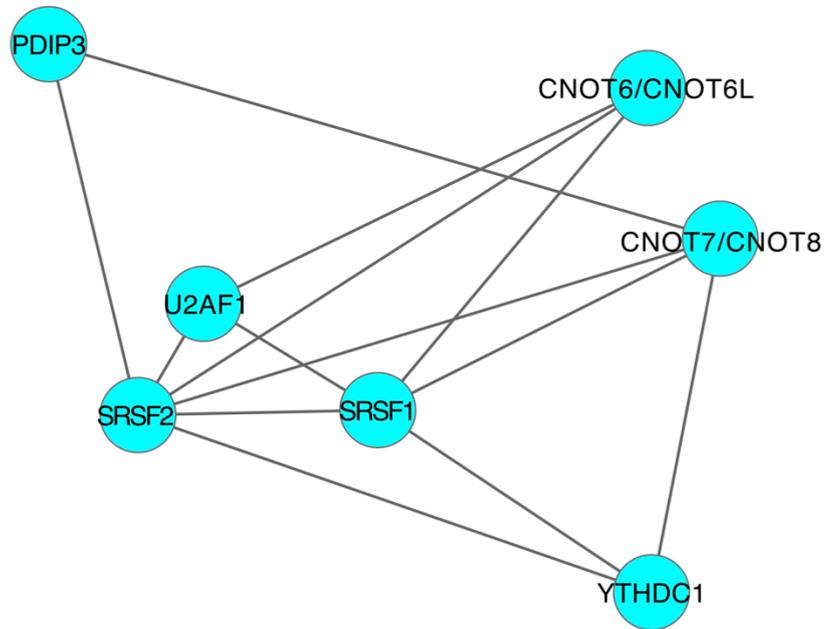
**Figure 3.15 LSm complex subunits and small nuclear ribonucleoproteins (SNRP) identified by ClusterONE.** The clustering analysis performed by ClusterONE grouped Lsm complex subunits LSM1, LSM2, LSM3, LSM4, and LSM6 together with small nuclear ribonucleoprotein polypeptides (SNRP) SMD2, RUXE, and RUXF.



**Figure 3.16 Polypyrimidine binding proteins identified by ClusterONE.** The ClusterONE analysis identified a group of pyrimidine binding proteins. The polypyrimidine binding protein PTBP1 was grouped with the poly(rC) binding proteins PCBP1, PBC2, and HRNPK. In this cluster is present as well the ATP-dependent DNA helicase RUVB1.



**Figure 3.17 Coatomer protein complex subunits identified by ClusterONE.** The ClusterOne plug-in identified four subunits of the coatomer protein complex (COPA, COPB2, COPD, and COPG1) interacting with the Caf1 proteins.



**Figure 3.18 Splicing factors identified by ClusterONE.** The ClusterOne plug-in grouped Serine/Arginine-Rich splicing factors SRF1 and SRF2 together with the U2 small Nuclear RNA Auxiliary Factor 1 (U2AF1). The polymerase (DNA) Delta interacting Protein 3 resulted to interact with both SRSF2 and CNOT7/CNOT8. SRSF1 and SRF2 interacted with the YTHDC1 putative splicing factor.

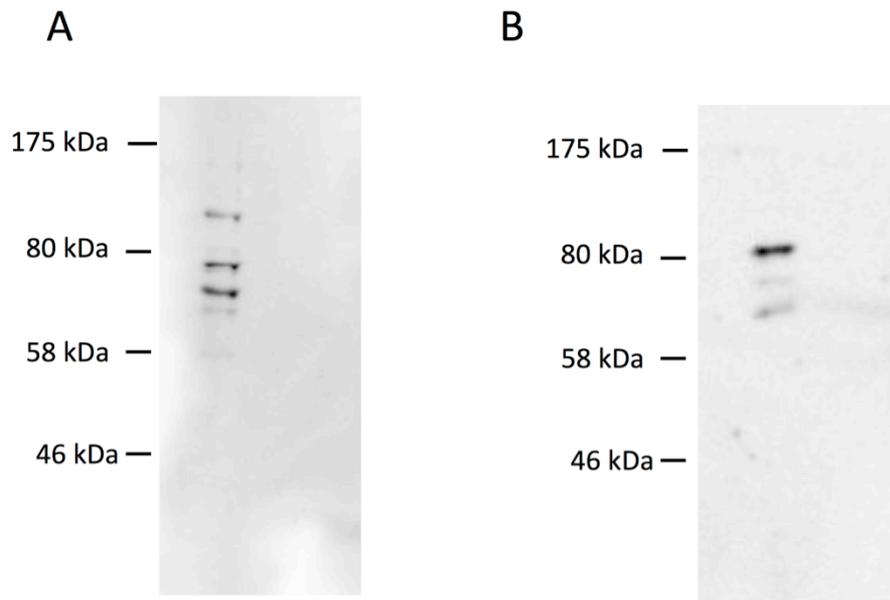
### 3.5 Evaluation of Pan2 and Pan3 antibodies

According to one biphasic kinetic model, Pan2-Pan3 complex is the first deadenylase involved in the deadenylation of the mRNA (Yamashita et al., 2005). The complex is composed of two proteins: Pan2 (127 kDa) and Pan3 (76 kDa) (Sachs and Deardorff, 1992; Boeck et al., 1996a). Pan2 is the subunit equipped with catalytic activity. Understanding the interactome of the Pan2-Pan3 deadenylase would be useful to determine and clarify the function of the Pan2-Pan3 complex and provide insight into its relationship with the Ccr4-Not complex.

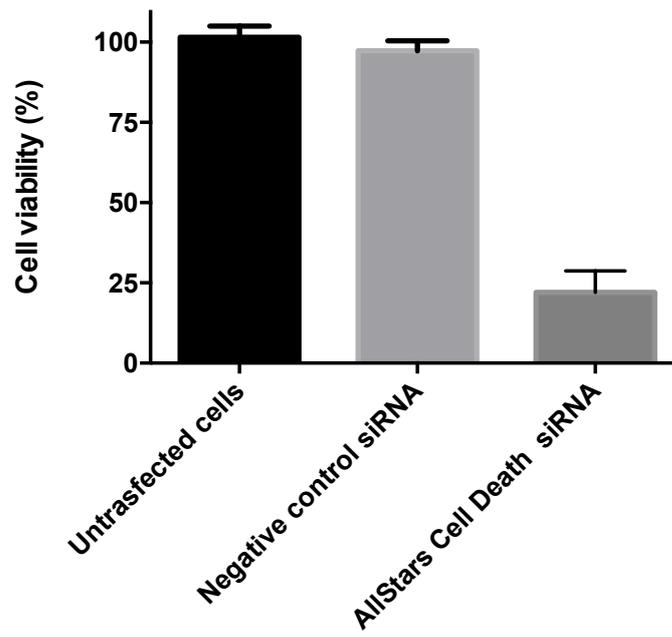
Therefore, in order to perform SILAC-RIME analysis of Pan2-Pan3 complex, commercially available antibodies targeting the two proteins were tested by western blot of MCF7 cellular lysates (Figure 3.19).

Unfortunately, both antibodies failed to specifically identify the two proteins.

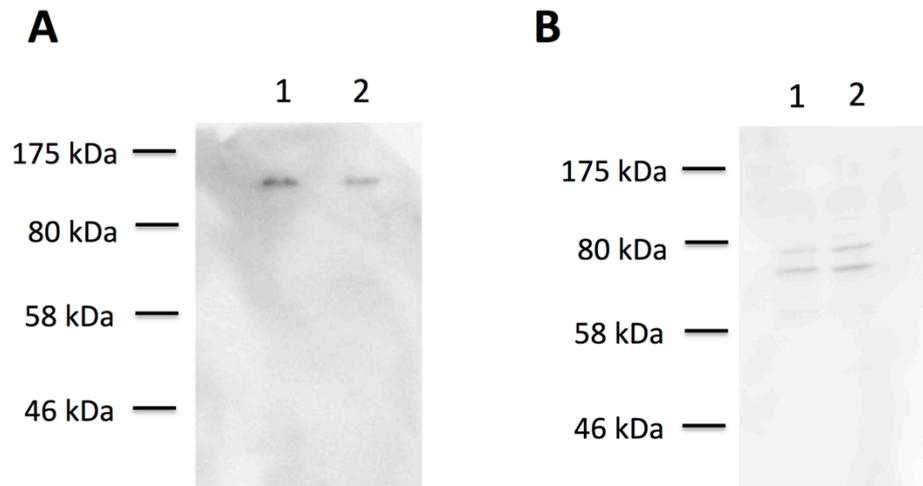
To further investigate the quality of the antibodies, siRNA knockdown of Pan2 and Pan3 was used to assess the quality of the antibodies. To ensure that the siRNA transfection was efficient, the transfection efficiency was evaluated using a cell death inducing siRNA (Allstar Cell Death) as positive control and a non-targeting siRNA as negative control. This experiment indicated that the siRNA transfection protocol was efficient (Figure 3.20). Subsequently, siRNAs were used to knockdown protein levels of Pan2 and Pan3 in MCF7 cells. Pan2 and Pan3 antibodies were used respectively for the detection of Pan2 and Pan3 in non-transfected and siRNA transfected cells by western blot as shown in Figure 3.21. These experiments did not validate the quality of the commercially available antibodies. At this point, further SILAC-RIME experiments were not further pursued due to lack of availability of high quality antibodies.



**Figure 3.19 Pan2 and Pan3 antibodies test. (A)** Pan2 (UPS52) (C-20) (Santa Cruz) antibody. A whole-cell lysate from MCF-7 breast cancer cells was separated by 10% SDS-PAGE. Proteins were transferred to a nitrocellulose membrane and subjected to immunoblot analysis. **(B)** Pan3 (A-9) (Santa) antibody. A whole-cell lysate from MCF-7 breast cancer cells was separated by 10% SDS-PAGE. Proteins were transferred to a nitrocellulose membrane and subjected to immunoblot analysis.



**Figure 3.20 Transfection efficiency of Interferin tested with AllStars Cell Death siRNA.** MCF7 human breast cancer cells were transfected using Interferin with siRNA inducing cell death (AllStar Cell Death) and negative control (non targeting siRNA).



**Figure 3.21 Western blot of cellular lysates of MCF7 cells transfected with Pan2 and Pan3 siRNAs. (A)** Pan2 (UPS52) antibody. A lysate from MCF-7 breast cancer cells was separated by 10% SDS-PAGE. Proteins were transferred to a nitrocellulose membrane and subjected to immunoblot analysis. Lane 1, lysate from mock-transfected MCF-7 cells; lane 2, lysate from MCF-7 cells transfected with siRNAs targeting Pan2. **(B)** Anti-Pan3 antibody. A lysate from MCF-7 breast cancer cells was separated by 10% SDS-PAGE. Proteins were transferred to a nitrocellulose membrane and subjected to immunoblot analysis. Lane 1, lysate from mock-transfected MCF-7 cells; lane 2, lysate from MCF-7 cells transfected with siRNAs targeting Pan3.

## 3.6 Discussion

### 3.6.1 The interactome of the Ccr4-Not deadenylase module

The deadenylase module of the Ccr4-Not complex is the responsible for its catalytic activity. SILAC-RIME analyses discovered the interactome of the Ccr4-Not complex deadenylase module. SILAC coupled affinity purification and mass spectrometry has been largely used to determine the interactome of proteins (Fang et al., 2008; Trinkle-Mulcahy et al., 2008; Zhong et al., 2011; Foster et al., 2013). This is a rapid, high sensitivity and cost effective method for direct identification of endogenous protein-protein interactions. Crosslinking allowed to preserve the endogenous protein-protein interactions and to maintain low-affinity interactions. The high specificity was obtained by the use of non-specific IgG as control taken along in parallel in all experiments (Mohammed et al., 2013). The SILAC-RIME analyses allowed the detection of proteins directly or indirectly interacting with the Caf1 and Ccr4 deadenylase subunits. The SILAC-RIME data was visualised by Cytoscape allowing a schematic representation of the interacting proteins. The relative ratios of heavy and light peptides are indicative of the enrichment of proteins in samples obtained with a specific antibody (Figure 3.4). The experiments were performed separately for Caf1 (CNOT7 and CNOT8) and Ccr4 (CNOT6 and CNOT6L) and the cytoplasmic, nuclear as well as combined nuclear and cytoplasmic fractions were analysed. The Caf1 subunit was found to have more interactions in all analysed fractions as compared to the Ccr4 subunit (Figure 3.3). The Bisogenet plugin was used in conjunction with Cytoscape to retrieve interactions from protein-protein interactions databases in order to complete the network with interactions between all its components (Martin et al., 2010). In the generated network of protein-protein interactions, proteins were represented as nodes and the interactions were represented as edges (Figure 3.7). Proteins interacting with both Caf1 and Ccr4 constituted the central dense region of the protein interaction network. The complexity of the resulting network did not allow a clear and straightforward analysis. Four different cut-off

thresholds based on the heavy/light peptide ratios (>2, >3, >5, and >10) were used to generate networks and evaluating their topology. From an accurate inspection, the network generated using a cut-off threshold of Heavy/Light peptide ratio >3 was chosen (Figure 3.8).

The protein-protein interaction network was inspected in order to find important biological pathways that may involve the recruitment of the Ccr4-Not complex. To this aim, two clustering methods were used to find protein clusters based on connectivity data (MCODE) or based on potential overlapping protein complexes (ClusterONE) (Bader et al., 2003, Zhang et al., 2014). Both MCODE and ClusterONE were used in conjunction with Cytoscape. The two approaches identified different numbers of protein clusters. MCODE identified twenty clusters with number of nodes grouped in the clusters varying from three to 29. ClusterOne identified 59 potential clusters with node content ranging from three to nine (Figure 3.9).

The clustering analysis identified clusters of known protein-protein interactions with the Ccr4-Not complex and new partners such as the mini-chromosome maintenance complex (MCM) component of the eukaryotic pre-replication complex (Figure 3.13), polypyrimidine binding proteins involved in pre-mRNA processing and mRNA transport and metabolism (Figure 3.16), and YTH domain containing protein family 1 (YTHDF1) involved in alternative splicing. The YTH domain containing protein family members specifically bind N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) (Figure 3.18).

Interestingly other proteins belonging to the YTH domain containing protein family were found in the SILAC-RIME produced interactome even though were not included in the found clusters. YTHDF1-3 proteins were found to interact with the Caf1 nuclease of the Ccr4-Not complex (Figure 3.10). YTHDF (or YTHD) proteins were found to be involved in mRNA stability (Wang et al., 2014). The SILAC-RIME data presented in this chapter suggest a recruitment of the Ccr4-NOT complex by the YTHDC1, YTHDF1, YTHDF2, and YTHDF3 in agreement with the recent findings that

demonstrated a recruitment of the Ccr4-NOT complex by YTHDF2 via CNOT1 (Du et al., 2016).

### **3.6.2 Pan2 and Pan3**

According to one model the first phase of deadenylation is carried out by the Pan2-Pan3 deadenylase complex. It is suggested that this complex is composed of the deadenylase subunit Pan2 and the non-catalytic subunits Pan3 forming a 1:2 trimeric complex (Uchida et al., 2004; Boeck et al., 1996a; Brown et al., 1996).

In order to characterise the Pan2-Pan3 interactome by SILAC-RIME, the quality and specificity of two commercially available antibodies recognising the two endogenous proteins were tested. The Pan2/UPS52 (C-20) (Santa Cruz) antibody was tested on MCF7 cell lysates. Western blots analysis detected a specific pattern of two bands but these were not of the correct size (Figure 3.19 A). Optimisation of the western blot and cellular lyses was carried out but the same result was achieved. Unlike the Pan2 antibody, the Pan3 (A-9) (Santa Cruz) antibody allowed the detection of a protein of the correct size of 76 kDa (Figure 3.19 B). Further test for antibody specificity were carried out using siRNA-mediated knockdown.

Unfortunately, in both cases it was not possible to observe a reduction of the protein expression (Figure 3.21). These preliminary experiments did not indicate that the tested antibodies were suitable for the SILAC-RIME analysis of Pan2 and Pan3.

# **Chapter 4**

## **Analysis of the NOT module of the Ccr4-Not complex**

## Chapter 4. Analysis of the NOT module of the Ccr4-Not complex

### 4.1 Introduction

The Ccr4-Not complex consists of two catalytic and six non-catalytic subunits. The largest subunit of the complex, CNOT1, acts as a central scaffold (Maillet and Collart, 2002; Winkler et al., 2006). Several modules are associated with CNOT1 in the complex (Bai et al., 1999). The CNOT11/CNOT10 module is attached to the N-terminus of CNOT1 containing anti-parallel  $\alpha$ -helices; its function it is still not clear (Basquin et al., 2012; Bawankar et al., 2013). The nuclease module is attached to the MIF4G domain of CNOT1. This sub-complex contains both catalytic subunits, of which each is encoded by two paralogous genes in vertebrata: Ccr4 is encoded by CNOT6 and CNOT6L while Caf1 is encoded by CNOT7 and CNOT8 (Winkler and Balacco, 2013).

A neighbouring DUF3819 domain interacts with CNOT9 (Bawankar et al., 2013).

On its C-terminus, CNOT1 binds CNOT2 and CNOT3 constituting the NOT module. CNOT2 and CNOT3, contain at their C-terminal a conserved domain named NOT-box, which is the region responsible for the hetero-dimerisation of the two proteins. Moreover, the C-termini are extended by two connector sequences termed CS that wrap like clamps around CNOT1. The region responsible for the interaction with CNOT1 is termed Not Anchoring Region (NAR) (Boland et al., 2013).

The NOT module components are involved in several cellular processes and their depletion may cause important changes in cell development and survival also phenotypically afflicting the organisms. Knockdown of CNOT2 leads to apoptotic cell death (Ito et al., 2011a). Several roles for CNOT3 were observed. This protein was described to promote the mitotic progression by down-regulating the expression of the spindle check-point protein MAD1 (Takahashi et al., 2012). In yeast, the CNOT3 orthologous protein Not5, binds poly(U) *in vitro* with a site at the NOT box (Bhaskar et al., 2013). Moreover, studies revealed the involvement of CNOT3 in B cell

differentiation destabilising p53 mRNA and in preventing cell death by targeting specific mRNA (Inoue et al., 2015; Suzuki et al., 2015b). CNOT3 is important for cardiac function and was shown to regulate the mRNA of metabolic enzymes in mice (Neely et al., 2010; Morita et al., 2011). Genetic studies on *retinitis pigmentosa* have identified CNOT3 as penetrance factor of t PRPF31 (Venturini et al., 2012; Rose et al., 2014). Moreover, studies identified mutations in CNOT3 involved in T-cell acute lymphoblastic leukemia onset (De Keersmaecker et al., 2013).

The aim of the results presented in this chapter was to identify amino acid residues that are important and essential for the trimeric assembly and function of the NOT module.

Somatic mutations found in different type of cancers were collected from the COSMIC database and the Ensemble genome browser and their effects on protein function were predicted using bioinformatics analysis.

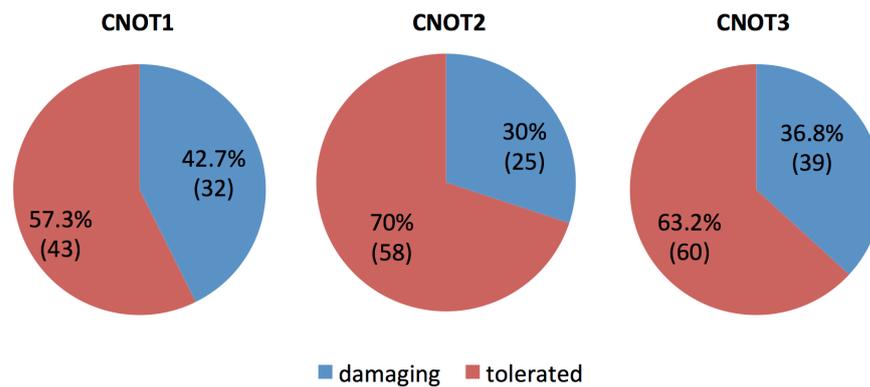
Investigation of the structure in the crystal structure published by Boland et al., 2013, was performed by a two-dimensional crystal map exploration using RINalyzer and PISA tools. These approaches narrowed the investigation to fewer residues of the CNOT1, CNOT2 and CNOT3 proteins that are likely for NOT module assembly and function.

## 4.2 Collection of mutations and prediction of their impact on the function of the NOT-module

A list of somatic mutations of the NOT module components was retrieved from COSMIC version v.67 and Ensembl. The Catalogue Of Somatic Mutations In Cancer (COSMIC) is the world largest comprehensive resource of information on somatic mutations in human cancer annotating data from the scientific literature and tumour resequencing data from the Cancer Genome Project at Sanger Institute, UK (Bamford et al., 2004; Forbes et al., 2008, 2011). The Ensemble genome browser annotates information for genomic analyses in different species with an emphasis on the human genome. The browser includes a collection of mutations found in human cancer as for the case of COSMIC (Chen et al., 2010; Rios et al., 2010; Flicek et al., 2014). In total 257 mutations were collected: 75 somatic mutations for the CNOT1 C-terminus (amino acid residues 1560-2376), 83 for CNOT2, and 99 for CNOT3.

In order to predict their impact on the function of the proteins, the mutations were analysed with Sorting Intolerant From Tolerant (SIFT) software. This tool allows prediction of amino acid substitutions that may affect protein function based on the conservation of amino acid residues in sequence alignments derived from closely related sequences, collected through PSI-BLAST (Ng and Henikoff, 2001, 2002, 2003, 2006; Kumar et al., 2009). The analysis predicted mutations having no effect on the protein function (tolerated) and mutations deleterious for the protein function (damaging). This analysis indicated that a significant proportion of the mutations annotated in COSMIC and Ensembl are predicted to be damaging to protein function (Figure 4.1). The output of a more detailed analysis is presented in Supporting Table 4.1, Supporting Table 4.2, and Supporting Table 4.3 for CNOT1, CNOT2 and CNOT3, respectively.

Tissue localisation as annotated in the databases was investigated for all the mutations retrieved and was represented as pie charts in Figure 4.2 A-C respectively for CNOT1, CNOT2 and CNOT3. The highest number of mutations of the three proteins was found in lung, endometrium, and large intestine tumor samples.

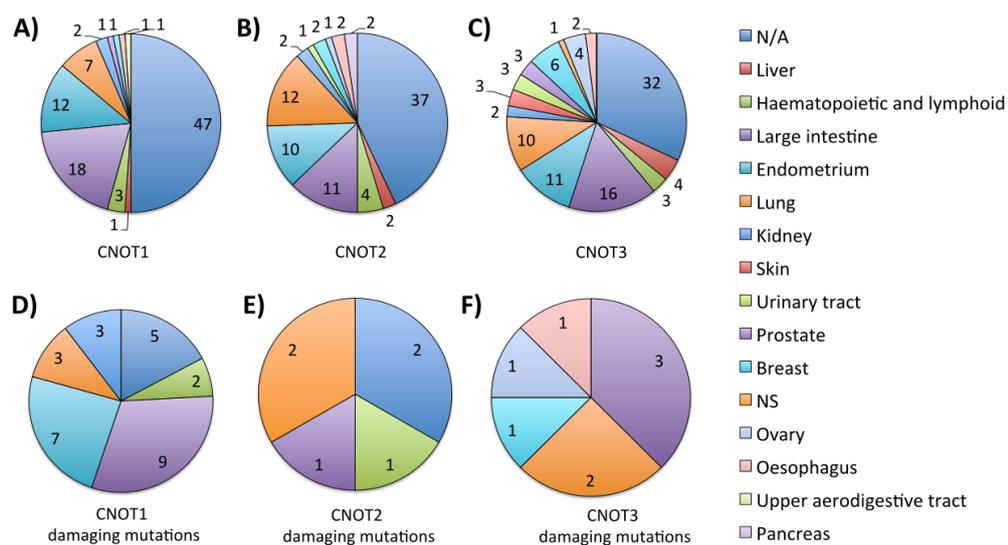


**Figure 4.1 Impact of the NOT module mutations on the proteins function.** The mutations impact on the proteins function was predicted using the Sorting Intolerant From Tolerant (SIFT) software that distinguished between damaging and tolerated mutations based on the conservation of amino acid residues in sequence alignments derived from closely related sequences. Of in total 75 CNOT1 mutations found in its C-terminus (amino acids 1560-2376), 43 were predicted to be tolerated and 32 were predicted to affect the protein function. COSMIC annotated 83 mutations found in CNOT2. Only 25 of these were predicted to have an impact on the function of the protein. The mutations found in the CNOT3 protein were 99. SIFT predicted 39 to affect the protein function.

Considering the high number of mutations found, the attention was focused on the residues that are involved in the assembly of the NOT module based on the available crystal structure of the human NOT module (PDB accession number 4C0D) (Boland et al., 2013). The fragments considered were amino acid residues 1560 to 2353 for CNOT1; residues 350 to 541 for CNOT2; and residues 607 to 748 for CNOT3.

The number of predicted damaging mutations in these regions were: 29 for CNOT1, six for CNOT2, and eight for CNOT3. The corresponding tissue localisation is summarised in Figure 4.2 D, E, and F respectively. CNOT1 mutations affecting the function were localised in large intestine (9), endometrium (seven) and lung (three), kidney (three), and haematopoietic and lymphoid (five) tumour tissues. Five mutations were annotated with unknown localisation (N/A). Distribution of the six CNOT2 mutations ranged in haematopoietic and lymphoid (one), large intestine (one), and lung (two) tumors and the remaining two had not annotated localisation. CNOT3 deleterious mutations were detected in large intestine (three), lung (two), breast (one), ovary (one), and oesophagus (one) cancers.

Table 4.1, Table 4.2, and Table 4.3 summarise the damaging mutation, the SIFT score and their tissue localisation relatively to CNOT1, CNOT2, and CNOT3. The SIFT score (0 to 1) is indicative of the probability of the mutation to be deleterious for the protein function. The smaller the value, the higher is the likelihood for the protein to affect the protein function. Mutations with a SIFT score <0.5 were considered to be damaging. This approach allowed to reduce the investigation field to 43 mutations likely to affect the function of the NOT module of the Ccr4-Not deadenylase complex.



**Figure 4.2 Tissue localisation of mutations found in the NOT module.**

Pie charts show the tissue localisation of the mutations annotated in COSMIC and Ensembl databases. Tissue localisation of all the retrieved mutations is shown for CNOT1 (A), CNOT2 (B) and CNOT3 (C). Tissue localisation of the mutations predicted by SIFT to be damaging for the proteins function is shown for CNOT1 (D), CNOT2 (E), and CNOT3 (F).

Mutation ID	Position	Residue Ref.	Residue alt.	SIFT score	Tissue Localisation
R1793G	1793	R	G	0.002	N/A
P1803A	1803	P	A	0.01	Haematopoietic and lymphoid
G1827C	1827	G	C	0.004	Large intestine
R1870H	1870	R	H	0.003	Large intestine
R1941Q	1941	R	Q	0.011	Large intestine
L1946P	1946	L	P	0.001	Large intestine
D1977G	1977	D	G	0	Haematopoietic and lymphoid
R1992Q	1992	R	Q	0	Large intestine
R2045W	2045	R	W	0	Endometrium and large intestine
R2045Q	2045	R	Q	0	Endometrium
M2063V	2063	M	V	0.021	Lung
K2073R	2073	K	R	0.023	Kidney
Y2092S	2092	Y	S	0	N/A
L2102F	2102	L	F	0.004	N/A
D2105Y	2105	D	Y	0	Haematopoietic and lymphoid
E2108A	2108	E	A	0	Large intestine
A2135V	2135	A	V	0.002	N/A
V2152G	2152	V	G	0	N/A
I2158T	2158	I	T	0.025	Liver
R2163Q	2163	R	Q	0.007	N/A
R2197H	2197	R	H	0.009	Endometrium and large intestine
P2206H	2206	P	H	0.001	Large intestine
R2209H	2209	R	H	0.045	Endometrium
N2211S	2211	N	S	0.003	Endometrium
I2250M	2250	I	M	0	Lung
Q2252H	2252	Q	H	0	Lung
R2263C	2263	R	C	0	Endometrium
R2263H	2263	R	H	0	Endometrium and large intestine
N2268H	2268	N	H	0	Large intestine
T2297M	2297	T	M	0.001	Large intestine
P2319S	2319	P	S	0	Kidney
K2331Q	2331	K	Q	0	Endometrium

**Table 4.1 CNOT1 damaging mutations.** List of 29 mutations found in CNOT1 (residues 1560-2353) predicted to be deleterious for the protein function by SIFT and their corresponding tissue localisation. The SIFT score (0 to 1) is indicative of the probability of the mutation to be deleterious.

Mutation ID	Position	Residue Ref.	Residue alt.	SIFT Score	Tissue localisation
V374E	374	V	E	0.032	Haematopoietic and lymphoid
T423M	423	T	M	0.011	N/A
K435Q	435	K	Q	0.027	Large intestine
A460T	460	A	T	0.001	N/A
D468Y	468	D	Y	0.007	Lung
R524W	524	R	W	0.001	Lung

**Table 4.2 CNOT2 damaging mutations.** List of six CNOT2 mutations relative to the NOT module crystal predicted to be deleterious for protein function by SIFT and their relative tissue localisation. The SIFT score (0 to 1) is indicative of the probability of the mutation to be deleterious.

Mutation ID	Position	Residue Ref.	Residue Alt.	SIFT Score	Tissue localisation
R633C	633	R	C	0	Lung
S667L	667	S	L	0.007	Large intestine
Q684R	684	Q	R	0.003	Large intestine
Q684H	684	Q	H	0.001	Large intestine
H709Q	709	H	Q	0.033	Ovary
R735Q	735	R	Q	0	Lung
K736N	736	K	N	0.01	Oesophagus
R750Q	750	R	Q	0.01	Breast

**Table 4.3 CNOT3 damaging mutations.** List of eight CNOT3 mutations relative to the NOT module crystal predicted to be deleterious for the protein function by SIFT and their relative tissue localisation. SIFT score (0 to 1) is indicative of the probability of the mutation to be deleterious.

### 4.3 Structure analysis

To determine important residues for the assembly of the NOT module, the NOT module crystal structure (PDB accession number 4C0D) was inspected (Boland et al., 2013). Two approaches were used.

Proteins, Interfaces, Structures and Assemblies (PISA) was used to explore the NOT module interfaces. The package provides an assessment of macromolecular interfaces, characterising interactions between residues between proteins, thermodynamic stability of the assemblies and their probable dissociation patterns (Krissinel and Henrick, 2007).

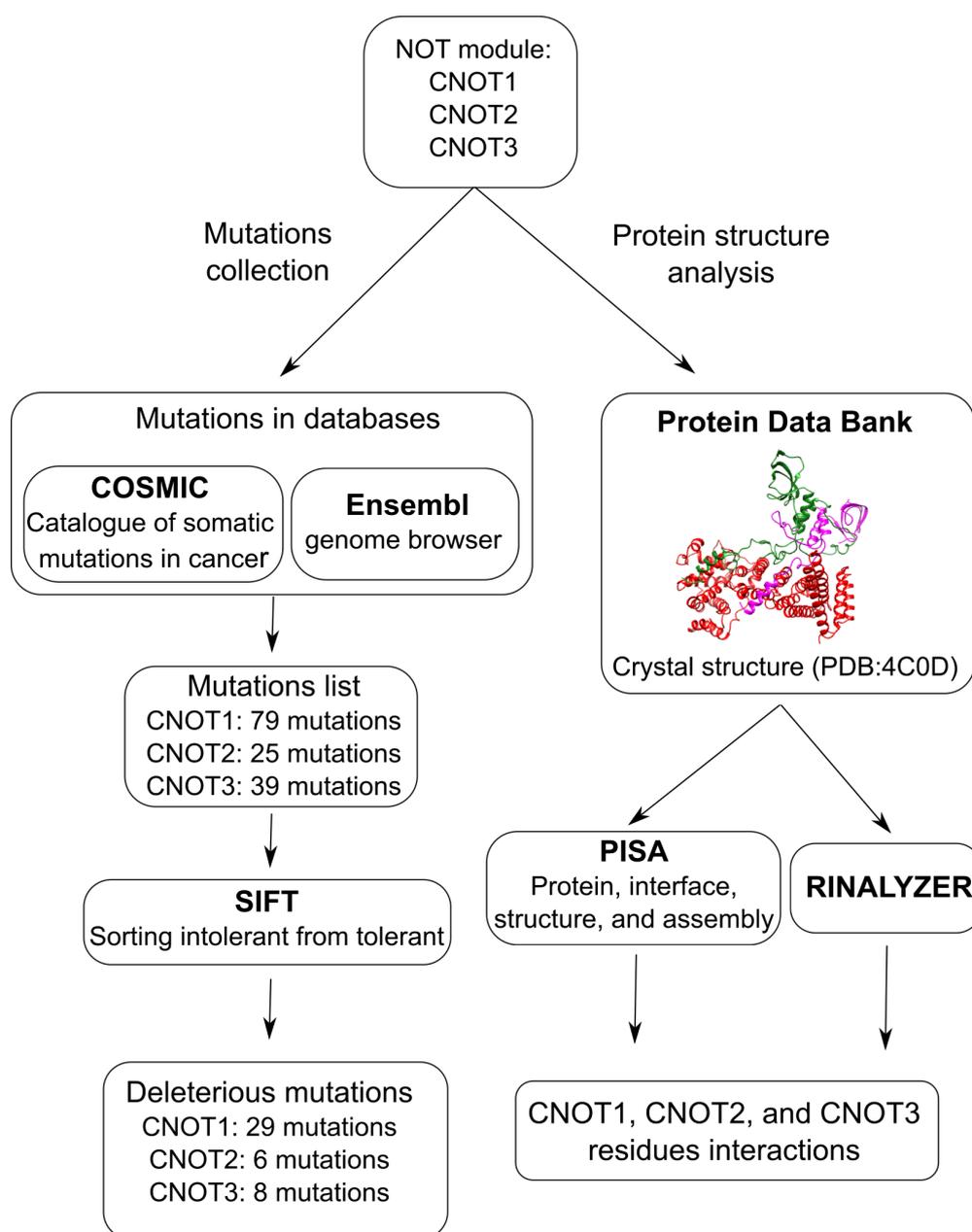
The second approach provided an insight of the contacts between all the amino acids present in the crystal by exploration of a two-dimensional map of the crystal structure obtained by RINalyzer (Doncheva et al., 2011) in combination with UCSF Chimera (Pettersen et al., 2004) and Cytoscape (Shannon et al., 2003). This analysis provided an insight on all the residues of the crystal structure and the attention was focused on the residues annotated in the COSMIC database and Ensemble and that were predicted to be deleterious for the protein function by SIFT analysis (Figure 4.3).

#### 4.3.1 Proteins, Interfaces, Structures and Assemblies (PISA)

PISA is a tool for the exploration of macromolecular interfaces (Krissinel and Henrick, 2007).

The crystal structure of the human NOT module (PDB accession number 4C0D) was used as input for the analysis. The software evaluated all the possible interfaces and their chemical and thermodynamic properties. The number of atoms (Nat) and the number of residues (Nres) of the single components of the crystal structure were calculated. Moreover, PISA provided information about the surface of the three proteins such as the number of atoms (sNat), the number of residues (sNres), and the surface area. Furthermore the solvation energy of folding (SEFD), the energy difference between the unfolded and folded state of the monomer, which is due to the solvation effect,  $\Delta G$  (kcal/mol) was calculated for the monomers. It indicates the thermal stability of the monomers, more negative values indicates more stable folded states. These results are

shown in Table 4.4. Importantly, PISA provided information about the contacts between residues of the three proteins. Hydrogen bonds and the salt bridges formed between the interfaces were identified, annotating their distance, the residues involved and the atoms involved in the interactions. CNOT1-CNOT3 interface contacts are shown in Table 4.5, CNOT1-CNOT2 interface contacts are shown in Table 4.6, and CNOT2-CNOT3 hydrogen bonds and salt bridges contacts are shown in Table 4.7 and Table 4.8 respectively. Prediction of the Gibbs free energy of dissociation of the assemblies ( $\Delta G_{\text{diss}}$ ) suggested that the most stable assembly is the heterotrimer (49.8 kcal/mol), followed by the CNOT2-CNOT3 dimer (40.8 kcal/mol), CNOT1-CNOT2 dimer (30.7 kcal/mol) and the less stable is the CNOT1-CNOT3 dimer (6.8 kcal/mol). A dissociation pattern was predicted as dissociation of the dimer CNOT2 -CNOT3, and the monomer CNOT1.



**Figure 4.3 Bioinformatic approaches to identify amino acid residues critical for the NOT module.** Identification of amino acid residues critical for the NOT module function and assembly was performed using two approaches. A list of mutations was collected from the COSMIC database and Ensembl and SIFT was used to predict tolerated and the deleterious mutations. The crystal structure of the NOT module was inspected by PISA and RINalyzer in order to identify residues important for the CNOT1, CNOT2 and CNOT3 interactions.

Interfacing monomers						
	Structure		Surface		SEFD	
	Nat	Nres	sNat	sNres	Area, Å <sup>2</sup>	ΔG (kcal/mol)
CNOT1	4096	512	2253	482	23936.0	-513.0
CNOT2	1558	191	1125	188	15257.2	-146.6
CNOT3	1210	142	847	139	11229.8	-109.0

**Table 4.4 Interfacing monomers.** PISA calculated the total number of atoms (Nat) and the number of residues (Nres) that form interactions in the crystal structure of the NOT module (PBD: 4C0D). Moreover it calculated the number of atoms (sNat), the number of residues (sNres), and the area (Å<sup>2</sup>) of the interaction surfaces of the three proteins. The solvation energy of folding (SEFD) for the corresponding monomer is shown ΔG (kcal/mol).

CNOT3		Distance (Å)	CNOT1		Contact
Residue	Atom		Residue	Atom	
TYR 613	[ OH ]	2.62	GLN1982	[ OE1 ]	Hb
HIS 624	[ ND1 ]	3.25	GLN1987	[ OE1 ]	Hb
SER 629	[ OG ]	3.06	TYR1933	[ OH ]	Hb
SER 631	[ OG ]	3.83	GLU2040	[ OE1 ]	Hb
ARG 633	[ NH2 ]	3.85	GLU2000	[ OE2 ]	Hb
TYR 613	[ OH ]	3.24	HIS1978	[ NE2 ]	Hb
GLN 614	[ O ]	3.43	TYR2074	[ OH ]	Hb
GLU 618	[ O ]	3.71	TYR2036	[ OH ]	Hb
GLU 618	[ OE1 ]	2.70	TYR2074	[ OH ]	Hb
GLU 618	[ OE2 ]	3.68	TYR2036	[ OH ]	Hb
HIS 624	[ O ]	3.01	GLN1987	[ NE2 ]	Hb
SER 629	[ O ]	3.02	ARG1992	[ NH2 ]	Hb
SER 629	[ OG ]	3.55	ARG1929	[ NH1 ]	Hb
ASP 630	[ O ]	3.27	ARG1992	[ NH2 ]	Hb
ASP 630	[ OD1 ]	2.33	TYR1933	[ OH ]	Hb
GLU 632	[ O ]	3.11	ARG1992	[ NH2 ]	Hb
ARG 633	[ NH2 ]	3.85	GLU2000	[ OE2 ]	Sb

**Table 4.5 CNOT1 and CNOT3 contacts.** PISA was used to identify interaction between the two indicated proteins, the distance between the residues involved in the interaction, the atoms of the residues involved and the type of bond formed (Hb, hydrogen bond; Sb, salt bond). The table summarise the contacts between the CNOT1 and CNOT3 residues.

CNOT2		Distance (Å)	CNOT1		Contact
Residue	Atom		Residue	Atom	
VAL 351	[ N ]	2.92	GLU2302	[ OE1]	Hb
ASP 353	[ OD2]	2.89	THR2260	[ OG1]	Hb
GLN 354	[ O ]	2.95	ARG2306	[ NH2]	Hb
PHE 355	[ O ]	2.82	GLN2303	[ NE2]	Hb
GLY 356	[ O ]	2.64	ARG2306	[ NE ]	Hb
MET 357	[ N ]	3.52	GLU2302	[ OE2]	Hb
ARG 365	[ NH2]	3.11	GLU2348	[ OE2]	Hb
ARG 365	[ NH2]	3.37	GLU2348	[ OE1]	Sb
ARG 365	[ NH2]	3.11	GLU2348	[ OE2]	Sb
HIS 375	[ NE2]	3.17	VAL2082	[ O ]	Hb
LEU 378	[ O ]	3.38	ASN2315	[ N ]	Hb
LEU 378	[ O ]	2.89	ASN2315	[ ND2]	Hb
GLY 379	[ O ]	3.08	ASN2315	[ ND2]	Hb
SER 380	[ O ]	2.65	GLN2127	[ NE2]	Hb
LEU 395	[ N ]	3.38	ASP2105	[ OD1]	Hb
LEU 395	[ N ]	3.48	ASP2105	[ OD2]	Hb
TYR 396	[ N ]	3.06	ASP2105	[ OD2]	Hb
PHE 399	[ O ]	3.88	ARG2045	[ NH1]	Hb
PRO 402	[ O ]	3.19	ARG2097	[ NH1]	Hb
TRP 403	[ NE1]	2.45	GLU2040	[ OE2]	Hb
ALA 404	[ O ]	3.16	ARG2097	[ NE ]	Hb
CYS 408	[ SG ]	3.45	ARG2045	[ NH1]	Hb
ASP 412	[ O ]	3.52	ARG2045	[ NH2]	Hb
PHE 540	[ O ]	3.60	LYS2093	[ NZ ]	Sb

**Table 4.6 CNOT1 and CNOT2 contacts.** PISA was used to identify

interaction between the two indicated proteins, the distance between the residues involved in the interaction, the atoms of the residues involved and the type of bond formed (Hb, hydrogen bond; Sb, salt bond). The table summarise the contacts between the CNOT1 and CNOT2 residues.

CNOT3		Distance (Å)	CNOT2		Contact
Residue	Atom		Residue	Atom	
TRP 622	[ NE1 ]	3.41	GLN 537	[ O ]	Hb
TRP 622	[ O ]	2.73	GLN 537	[ NE2 ]	Hb
HIS 623	[ O ]	2.74	GLN 537	[ NE2 ]	Hb
MET 625	[ O ]	3.71	ASN 534	[ ND2 ]	Hb
PRO 626	[ O ]	3.64	ASN 534	[ ND2 ]	Hb
ARG 633	[ N ]	3.53	ASP 414	[ OD2 ]	Hb
ARG 635	[ NH2 ]	3.21	GLN 411	[ O ]	Hb
ARG 635	[ NH2 ]	2.84	ASP 414	[ OD1 ]	Hb
ARG 635	[ NH1 ]	2.79	ASP 414	[ OD1 ]	Hb
TYR 637	[ OH ]	2.61	ASN 424	[ OD1 ]	Hb
ARG 640	[ NH1 ]	3.21	ASP 412	[ OD2 ]	Hb
ASN 641	[ N ]	2.64	ASP 454	[ OD2 ]	Hb
ASN 641	[ O ]	3.18	VAL 455	[ N ]	Hb
CYS 643	[ N ]	3.28	GLY 453	[ O ]	Hb
PRO 647	[ O ]	3.07	ARG 476	[ NH1 ]	Hb
TYR 648	[ OH ]	2.52	GLU 522	[ OE2 ]	Hb
TYR 648	[ O ]	3.27	ARG 476	[ NH1 ]	Hb
TYR 648	[ O ]	2.63	TYR 471	[ OH ]	Hb
HIS 649	[ ND1 ]	3.10	TYR 448	[ OH ]	Hb
HIS 649	[ NE2 ]	3.52	GLU 522	[ OE2 ]	Hb
HIS 649	[ O ]	3.09	ARG 476	[ NH2 ]	Hb
HIS 650	[ N ]	3.75	TYR 448	[ OH ]	Hb
HIS 650	[ NE2 ]	3.00	ARG 476	[ O ]	Hb
HIS 650	[ O ]	2.83	GLN 457	[ NE2 ]	Hb
ASP 658	[ OD2 ]	2.43	LYS 398	[ NZ ]	Hb
ARG 665	[ NH1 ]	2.68	LEU 505	[ O ]	Hb
SER 667	[ N ]	3.08	ASP 442	[ OD1 ]	Hb
SER 667	[ OG ]	2.76	ASP 442	[ OD1 ]	Hb
GLU 669	[ OE1 ]	3.24	GLY 440	[ N ]	Hb
THR 670	[ OG1 ]	2.61	ASP 442	[ OD1 ]	Hb
TYR 677	[ OH ]	2.64	ALA 432	[ O ]	Hb
GLU 679	[ OE1 ]	3.79	ASN 424	[ ND2 ]	Hb
GLU 679	[ OE2 ]	3.19	THR 423	[ N ]	Hb
GLU 679	[ OE2 ]	3.11	ASN 424	[ N ]	Hb
GLY 680	[ O ]	3.62	PHE 415	[ N ]	Hb
LYS 682	[ N ]	2.86	ILE 413	[ O ]	Hb
ALA 683	[ N ]	2.95	PRO 410	[ O ]	Hb
GLN 684	[ NE2 ]	3.21	TYR 421	[ O ]	Hb
TYR 729	[ O ]	3.01	ALA 432	[ N ]	Hb
GLU 730	[ OE1 ]	3.00	LYS 430	[ NZ ]	Hb

**Table 4.7 Hydrogen bonds contacts between CNOT3 and CNOT2.**

The table summarise the hydrogen bonds between CNOT2 and CNOT3 residues identified by PISA.

CNOT3		Distance (Å)	CNOT2		Contact
Residue	Atom		Residue	Atom	
ARG 635	[ NH2]	2.84	ASP 414	[ OD1]	Sb
ARG 635	[ NH1]	2.79	ASP 414	[ OD1]	Sb
ARG 635	[ NH1]	3.90	ASP 414	[ OD2]	Sb
ARG 640	[ NH1]	3.53	ASP 412	[ OD1]	Sb
ARG 640	[ NH2]	3.60	ASP 412	[ OD1]	Sb
ARG 640	[ NH1]	3.21	ASP 412	[ OD2]	Sb
HIS 649	[ NE2]	3.42	GLU 522	[ OE1]	Sb
HIS 649	[ NE2]	3.52	GLU 522	[ OE2]	Sb
ASP 658	[ OD2]	2.43	LYS 398	[ NZ ]	Sb
GLU 730	[ OE1]	3.00	LYS 430	[ NZ ]	Sb
GLU 730	[ OE2]	3.84	LYS 430	[ NZ ]	Sb

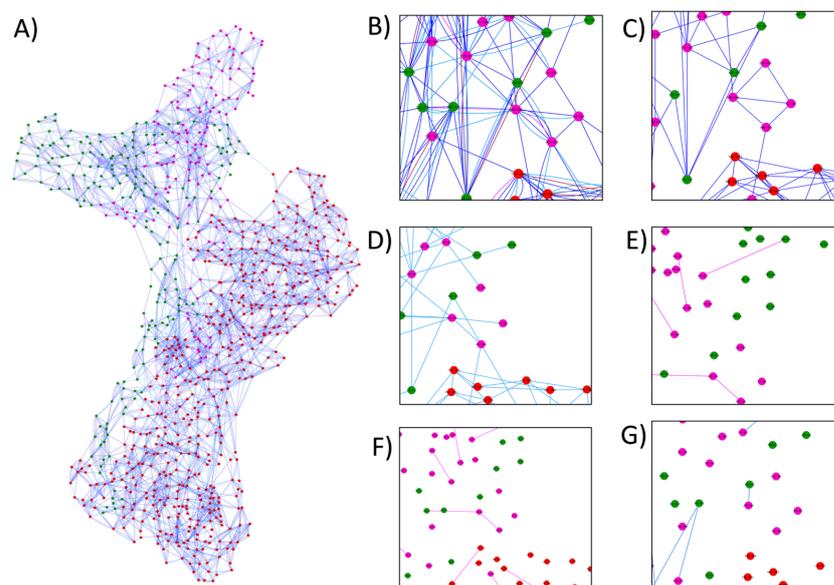
**Table 4.8 Salt bridges between CNOT3 and CNOT2.** PISA was used to identify interaction between the two indicated proteins and the distance between the residues involved in the interaction. The table summarises the salt bonds between CNOT2 and CNOT3 residues.

### 4.3.2 RINalyzer

In order to have a better understanding of all contacts between amino acids within the heterotrimer, RINalyzer was used to obtain a bi-dimensional map of the crystal.

The NOT module crystal structure (PDB accession number 4C0D) was used to generate the Residue interaction networks (RINs). The RINs were generated using the RINertator (v0.5) module together with Reduce (v3.14) and Probe (v2.12) tools (Doncheva et al., 2011). RINs data were then visualised and explored using UCSF Chimera (Pettersen et al., 2004) and Cytoscape 3.1.1 (Shannon et al., 2003).

In the RIN generated, shown in Figure 4.4 A, amino acids were represented as nodes. CNOT1 residues were coloured in red, CNOT2 residues were coloured in green, and CNOT3 residues were coloured in purple. Contacts between amino acids were shown as edges with a different colour depending on their nature (Figure 4.4 B). The residue-residue interactions were classified in contacts between main chains (Figure 4.4 C), between main chain and side chain (Figure 4.4 D), hydrogen bonds between main chains (Figure 4.4 E), hydrogen bonds between main chain and side chain (Figure 4.4 F), hydrogen bonds between side chains (Figure 4.4 G). This analysis allowed gave additional information on the interaction of the residues and their nature. Tables listing the residue contacts that were predicted to be damagingly by SIFT analysis are presented Supporting Tables 4.4, 4.5, and 4.6 for CNOT1, CNOT2 and CNOT3 respectively.



**Figure 4.4 Residue Interaction Network of the human NOT module crystal structure.** **(A)** RINalyzer was used to obtain the Residue Interaction Network (RIN) of the NOT module crystal structure (PDB: 4C0D). **(B)** In the RIN, amino acids are represented as nodes (red for CNOT1, green for CNOT2, and purple for CNOT3) and interactions are represented as edges. The interactions were classified according to their nature in **(C)** contacts between main chains of two amino acids, **(D)** contacts between main chain and side chain, **(E)** hydrogen bonds between the two main chains of two amino acids, **(F)** hydrogen bonds between the main chain and a side chain, and **(G)** hydrogen bonds between the side chains of two amino acids.

#### 4.4 Protein-protein interaction analysis

The effect of the mutations on the functionality of the NOT module can be deleterious disrupting the ability of the proteins to interact and assemble in the module. Therefore, I next aimed to investigate the functional effect of mutations using protein-protein interaction assays to complement the predictions based on computational analysis. Specifically, the aim was to detect interactions between CNOT1 and CNOT2, as well as CNOT1 and CNOT3 by yeast two hybrid analysis and by co-immunoprecipitation.

##### 4.4.1 Yeast two hybrid assay for CNOT1-CNOT2 interactions

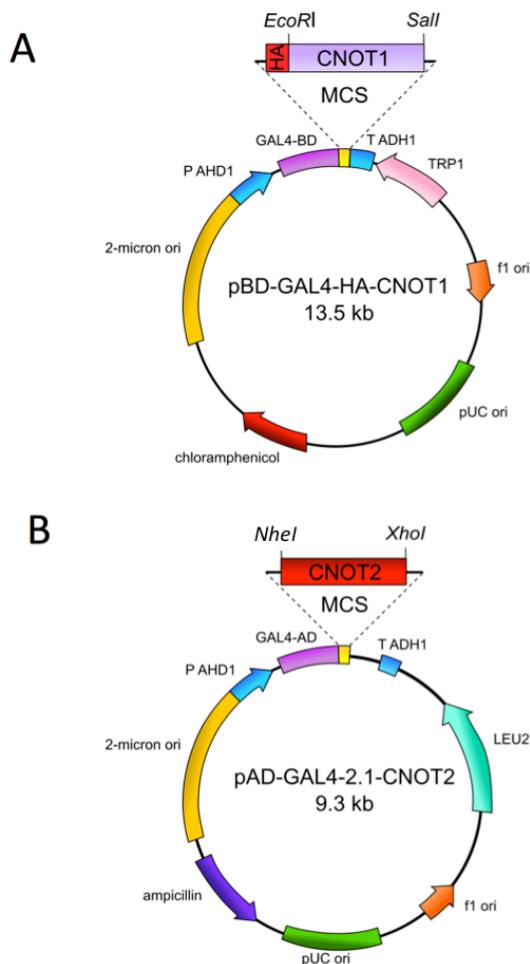
The interaction between CNOT1 and CNOT2 was investigated in *S. cerevisiae* strain YRG2 using a yeast two hybrid assay. This technique offers a sensitive, cost-effective and semi-quantitative means to test and detect interactions between two proteins. The assay relies on the capacity to separate the eukaryotic transcriptional activator GAL4 into two separate domains, the DNA binding domain (BD) that binds the upstream activating sequence (UAS) and the transcriptional activation domain (AD) that activate the transcription of the gene reporter (Ma and Ptashne, 1987). Fusion proteins BD-CNOT1 and AD-CNOT2 were produced. The interaction of the two proteins activates the transcription of the reporter gene  $\beta$ -galactosidase.

In order to investigate the interaction between CNOT1 and CNOT2 with the GAL4 two hybrids system, the pBD-GAL4-HA-CNOT1 and the pAD-GAL4-2.1-CNOT2 plasmids were generated. The CNOT1 full length cDNA was amplified to introduce the EcoRI restriction site at the 5' end, and the Sall restriction site at the 3' end of the sequence. The PCR product was then isolated by gel electrophoresis and digested with the EcoRI and Sall restriction enzymes. The pBD-GAL4-HA vector was digested with EcoRI and Sall restriction enzymes, treated with Antarctic Phosphatase in order to prevent self-ligation and ligated with the digested CNOT1.

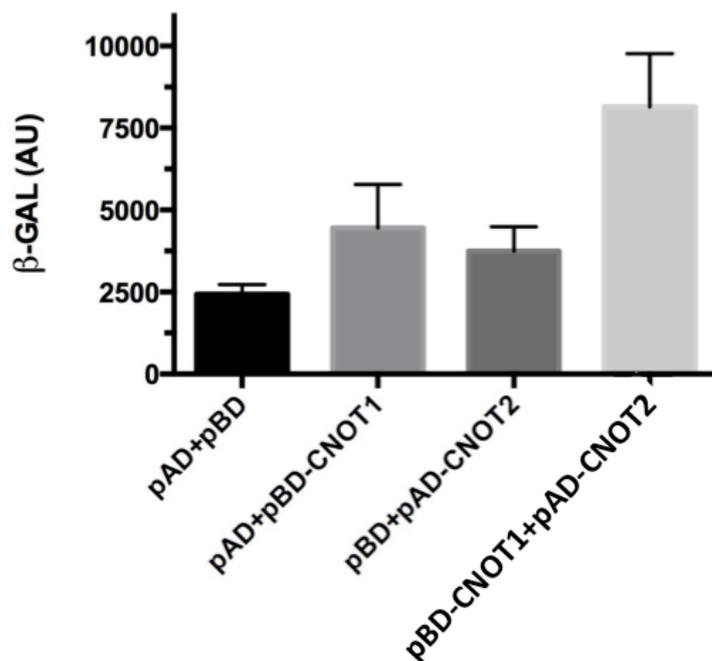
The CNOT2 cDNA was amplified by PCR to introduce the NheI restriction site at the 5' and the XhoI restriction site at the 3' of the sequence. The amplified sequence was then isolated by gel electrophoresis and digested with the NheI and XhoI restriction enzymes. The pAD-2.1 vector was

digested with NheI and XhoI and treated with Antarctic Phosphatase. Ligation between the digested CNOT2 cDNA fragment and the digested pAD-GAL4-2.1 vector was then set up. The generation of the pBD-GAL4-HA-CNOT1 and pAD-GAL4-2.1-CNOT2 was confirmed by colony PCRs, double digestions, and sequencing. Maps of the generated plasmids are presented in Figure 4.5.

*S. cerevisiae* YRG2 was transformed with the above-mentioned plasmids on selective plates. Colonies were used to inoculate culture to perform Beta-Glo Assay System (Promega) to quantitate  $\beta$ -galactosidase expression dependent on the interaction between CNOT1 and CNOT2. The Beta-Glo Assay results are shown in Figure 4.6. Empty plasmids were used as negative controls in combination with the pBD-GAL4-HA-CNOT1 and pAD-GAL4-2.1-CNOT2. Unfortunately, the interaction detected was not statistically different from the negative controls.



**Figure 4.5 Generation of pBD-GAL4-HA-CNOT1 and PAD-GAL4-2.1-CNOT2. (A)** The pBD-GAL4-HA-CNOT1 plasmid was generated by inserting CNOT1 into the EcoRI and Sall restriction sites of the multiple cloning site (MCS). The pBD-GAL4 Cam plasmid includes yeast ADH1 promoter (P ADH1), GAL4 DNA binding domain (GAL4-BD), yeast ADH1 terminator (T ADH1), yeast TRP1 ORF for selection with tryptophan (TRP), f1 origin, pUC origin, chloramphenicol resistance ORF, and 2 $\mu$  yeast origin. **(B)** The pAD-GAL4-2.1-CNOT2 was generated by inserting CNOT2 cDNA into the NheI and XhoI restriction sites of the MCS. The pAD-GAL4-2.1 plasmid has the same feature of the pBD-GAL4 plasmid except of the selection genes. The leucine (LEU) gene is used for selection in yeast, and the ampicillin is used for the selection in bacteria.



**Figure 4.6 CNOT1 and CNOT2 yeast two hybrid assay.** *S. cerevisiae* YRG2 was transformed with pBD-GAL4-HA-CNOT1 and pAD-GAL4-2.1-CNOT2. The Beta-Glo Assay System (Promega) was used to quantitate  $\beta$ -galactosidase expression. Empty plasmids were used as negative controls in combination with the pBD-GAL4-HA-CNOT1 and pAD-GAL4-2.1-CNOT2. Interactions are quantified with as  $\beta$ -galactosidase activity units (AU). Error bars indicate the standard error of the mean (n=3).

#### 4.4.2 Co-Immunoprecipitation for CNOT1-CNOT3 interaction.

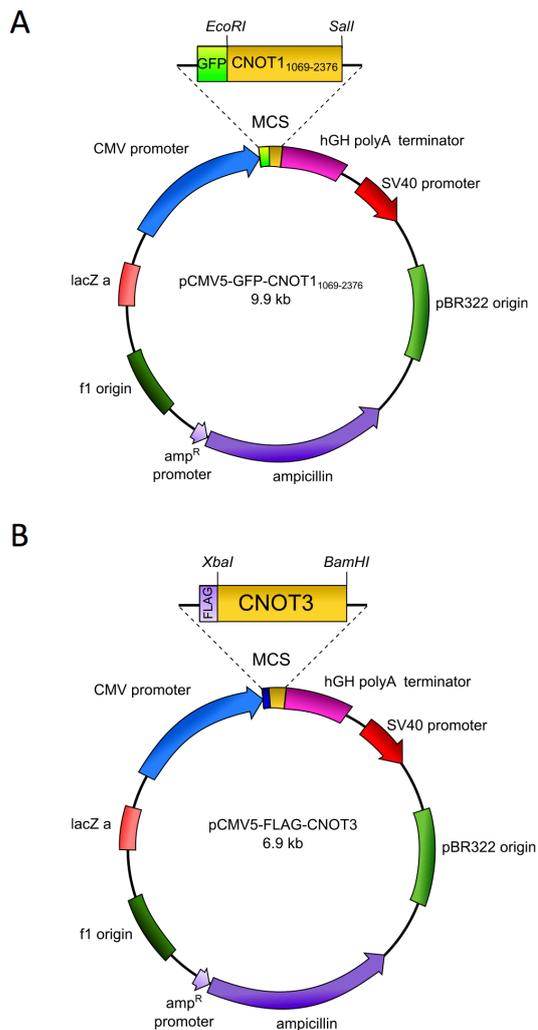
As an alternative way to detect protein-protein interactions between components of the NOT module, co-immunoprecipitation experiments were carried out.

In order to express CNOT1 and CNOT3 in HEK293 cells, plasmids pCMV5-GFP-CNOT1<sub>1079-2736</sub> and pCMV5-FLAG-CNOT3 were generated (Figure 4.7). The CNOT1 cDNA encoding the C-terminus of CNOT1 (residues 1079 to 2736) was amplified by PCR to introduce the EcoRI restriction site at the 5' end and the Sall restriction site at the 3' end. The PCR product was then digested with the EcoRI and Sall restriction enzymes. The pCMV5-GFP empty vector was digested with the same restriction enzymes and treated with Antarctic Phosphatase to prevent self-ligation. The cloning was verified by colony PCRs, double digestions, and sequencing of the resulting plasmid.

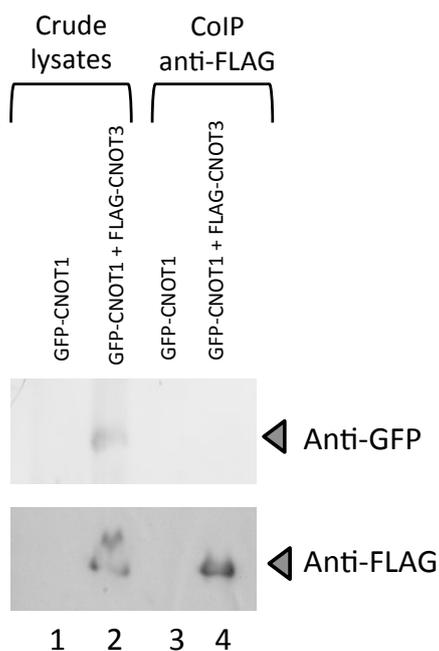
The CNOT3 cDNA was amplified by PCR to introduce the XbaI restriction site at the 5' end and the BamHI restriction site at the 3' end. The pCMV5-FLAG empty vector was digested with XbaI and BamHI and treated with Antarctic Phosphatase. Transformation was verified by colony PCR and the resulting plasmid was verified by double digestions and DNA sequencing.

The two plasmids were used to transfect HEK293 cells using JetPEI (Polyplus). Immunoprecipitations were carried out using protein A/protein G agarose beads coupled with anti-FLAG antibody.

Western blots were performed to detect the interaction between the two proteins. Cells transfected with pCMV5-GFP-CNOT1 only were used as negative controls (Figure 4.8, lanes 1 and 3). Detection of the FLAG-CNOT3 was successful in both crude lysate (lane 2) and in the immunoprecipitated sample (lane 4). Unfortunately, detection of the GFP-CNOT1 was successful only in the crude lysate of the co-transfected sample (lane 2). Optimisation of the preliminary experiment was tried but conditions for successful immunoprecipitation were not identified.



**Figure 4.7 Generation of pCMV5-GFP-CNOT1 and pCMV5-FLAG-CNOT3 plasmids. (A)** The pCMV5-GFP-CNOT1 plasmid for expression of CNOT1 (1079-2736) in HEK293 cells was generated by inserting the cDNA in the EcoRI and Sall restriction sites of the multiple cloning site (MCS). **(B)** The pCMV5-FLAG-CNOT3 was generated by inserting the CNOT3 cDNA in the XbaI and BamHI restriction sites of the MCS. The pCMV5 plasmid includes the CMV promoter, the human growth hormone polyadenylation signal (hGH polyA terminator), the SV40 promoter, the pBR322 origin, the  $\beta$ -lactamase gene for the resistance to ampicillin and its promoter ( $\text{amp}^R$  promoter), the f1 origin and lacZ alpha peptide (lacZ a).



**Figure 4.8 CNOT1 and CNOT3 co-immunoprecipitation.** GFP-CNOT1 and FLAG-CNOT3 were expressed in HEK293 cells. Immunoprecipitations were carried out using protein A/protein G agarose beads coupled with anti-FLAG antibody. Crude lysate was used to assess the expression of the CNOT1 fragment and the CNOT3 protein (lanes 1 and 2). Cells transfected with pCMV5-GFP-CNOT1 only were used as negative controls (lanes 1 and 3). Detection of the FLAG-CNOT3 was successful in both crude lysate (lane 2) and in the immunoprecipitated sample (lane 4). GFP-CNOT1 was detected in the crude lysate of the co-transfected sample (lane 2), but not in the co-immunoprecipitation sample (lane 4).

## 4.5 Discussion

The NOT-module components CNOT1, CNOT2 and CNOT3 interact via their C-terminal domains (Boland et al., 2013). In particular, CNOT2 and CNOT3 interact with each other by a conserved domain termed the NOT-box (residues 350-408 and 607-634 respectively).

The NOT-module is a fundamental component of the Ccr4-Not complex and is involved in several pivotal cellular processes. The importance of this module is sustained from its involvement in different pathologies such as heart failure, *retinitis pigmentosa*, osteoporosis, obesity and cancers (Neely et al., 2010; Morita et al., 2011; Venturini et al., 2012; De Keersmaecker et al., 2013; Rose et al., 2014; Watanabe et al., 2014). In this study we used *in silico* and *in vitro* approaches to attempt to identify residues critical for the assembly of the NOT module and its function. In order to better understand the protein-protein interactions and to detect which residues are critical for the complex function and assembly somatic mutations of the three proteins were collected from COSMIC and Ensemble databases. The investigation identified 75 mutations for CNOT1 (amino acid residues 1565-2376), 83 for the full length CNOT2 and 99 for the full length CNOT3 mostly found in lung, endometrium, and large intestine tumors samples (Figure 4.2). However in the databases the localisation of some mutations was not specified and was annotated as unknown (N/A).

SIFT analysis of the three protein segments predicted mutations that may be deleterious for protein function (Supporting Tables 4.1, 4.2 and 4.3). In the NOT-module crystal, 29 mutations of CNOT1, six of CNOT2 and eight of CNOT3 were predicted to be deleterious for the complex function. The tissue localisation is less heterogeneous due to the small number of predicted functionally damaging mutations. CNOT1 mutations affecting the function were localised in large intestine, endometrium and lung (3); but also kidney (3) and haematopoietic and lymphoid (5) tumour tissues. Five mutations have unknown localisation. Distribution of the six CNOT2 mutations range in haematopoietic and lymphoid (one), large intestine (one) and lung (two) tumors and the remaining two have not annotated

localisation. CNOT3 deleterious mutations were detected in large intestine (three), lung (two), breast (one), ovary (one) and oesophagus (one) cancers.

PISA interface and surface analysis performed on the CNOT1, CNOT2, and CNOT3 heterotrimer provided chemical information of the three possible interfaces (CNOT1-CNOT2, CNOT1-CNOT3 and CNOT2-CNOT3) such as surface residues, solvent accessible area, solvation energy, number of atoms and residues involved in the formation of the interface, hydrogen bonds, salt bonds and the possible assemblies analysis. Additional information about residue-residue interactions was obtained using RINalyzer, which provides a two-dimensional map of interacting residues. The residue interaction network is shown in Figure 4.4

By exploration of the residue interaction network obtained by RINalyzer and PISA, interactions of each residue that are likely important for protein function were annotated and reported in Supporting Tables 4.4, 4.5, and 4.6 and Table 4.5, Table 4.6, Table 4.7, and Table 4.8. Interactions were classified in main chain-main chain, main chain-side chain, side chain-side chain contacts and hydrogen bonds (Figure 4.4 B, C, D, E, F, G).

Among all CNOT1 variants predicted to be deleterious for the protein function by SIFT, no residues were found to interact with residues of both CNOT2 and CNOT3. CNOT1 Arg1992 interacted via main chain-side chain with Glu632, Ser692 and Asp630 of CNOT3. CNOT1 variant R1992Q was found in large intestine cancer. CNOT1 Arg2045 interactions resulted in contacts with several CNOT2 residues. Its main chain interacted with the side chain of Tyr369, Pro397 and Phe399 of CNOT2 and its side chain interacted with those of Ile413 and Cys408. Two variants of this residue were found in endometrium cancer, R2045Q and R2045W. Moreover, CNOT1 Tyr2092 (variant Y2092S found in unknown cancer type) interacted with CNOT2 His375. Asp2105 of CNOT1 was predicted to interact by hydrogen bonds with side chains of Asn394, Leu395 and Tyr396 of CNOT2. CNOT1 variant D2105Y was found in haematopoietic and lymphoid cancers.

Only three residues of CNOT2 were predicted to established inter-molecular contacts. Asp412 of CNOT2 was predicted to form hydrogen bonds with Arg2045 of CNOT1 and Arg640 of CNOT3. Mutations of this residue were not found in any cancer though its partner Arg2045 of CNOT1 was found mutated in endometrium cancer (variants R2045Q and R2045W). Hydrogen bonds between the main chain of Val743 of CNOT2 and the side chain of Asn2315 of CNOT1 were predicted. Thr423 of CNOT2 established side chain-side chain contacts with Phe698 and Met703 of CNOT3.

Arg633 of CNOT3 was predicted to form hydrogen bonds with Asp414 of CNOT2 and both hydrogen and salt bonds with Met1996 and Glu2000 of CNOT1. This residue was found mutated in lung cancer (R633C). Ser667 of CNOT3 established contacts with the side chain of Asp442 of CNOT2 with which formed hydrogen bonds as well. Mutation S667L of CNOT3 was found in large intestine cancer. Finally Gln684 of CNOT3 main chain was predicted to contact the side chain of Tyr421 of CNOT2. Two CNOT3 variants, Q684R and Q684H were found in large intestine cancer.

These preliminary analyses supplied a general profile of the complex interacting residues allowing to predict which mutations could affect the interactions between the heterotrimer components and which exclusively affect the function.

In order to study how mutations affect the assembly and function of the NOT module *in vitro*, preliminary experiments to detect the protein-protein interactions of the two assemblies CNOT1-CNOT2 and CNOT1-CNOT3 were performed.

Yeast two hybrids assay were performed to directly demonstrate the interactions between CNOT1 and CNOT2 (Figure 4.6). Unfortunately, it was not possible to detect the interaction of the assembly in this system. In order to prove the interaction between CNOT1 and CNOT3 co-immunoprecipitations were performed in HEK293 cells. With this analysis, FLAG-CNOT3 (100 kDa) was successfully immunoprecipitated, but no co-immunoprecipitation of GFP-CNOT1 (240 kDa) was detected (Figure 4.8) possibly due to low expression. In the future, an alternative

method to evaluate the ability of variants to interact with their partners in the NOT module may involve the reconstitution of the NOT module using recombinant (variant) proteins as described by Boland et al. (2013).

**Chapter 5**  
**Analysis of sequence variants**  
**of the CNOT7 ribonuclease**

## Chapter 5. Analysis of sequence variants of the CNOT7 ribonuclease

### 5.1 Introduction

In addition to at least six non-catalytic subunits, the Ccr4-Not complex contains two distinct catalytically active deadenylase subunits: Caf1 and Ccr4. In vertebrate cells, the complexity of the Ccr4-Not complex is further increased by the presence of paralogues of the Caf1 subunit (encoded by either *CNOT7* or *CNOT8*) and the occurrence of two Ccr4 paralogues (encoded by *CNOT6* or *CNOT6L*).

Ccr4 belongs to the exonuclease-endonuclease phosphatase (EEP) protein family catalysing phosphate ester hydrolysis requiring two  $Mg^{2+}$  ions located in the active site. Substitution of the amino acids Asn195, Glu240, Asp410, Asp489, His529 coordinating the metal ions inhibits the enzymatic activity (Dlakić, 2000; Chen et al., 2002; Wang et al., 2010). Ccr4 also contains a leucine-rich repeat (LRR) domain responsible for the interaction with the second catalytic subunit Caf1 (Malvar et al., 1992; Dupressoir et al., 2001; Basquin et al., 2012; Clark et al., 2004). Caf1 contains four conserved active site residues that coordinate divalent ions characterised by the presence of an RNase D domain (Zuo and Deutscher, 2001). The ions are coordinated to a single glutamate and three aspartate residues (DEDD) and substitution of any of these residues abolishes the enzymatic activity (Thore et al., 2003; Jonstrup et al., 2007; Horiuchi et al., 2009).

The nuclease sub-complex is anchored to the central MIF4G domain of the central scaffold of the complex CNOT1. This region is composed of  $\alpha$ -helices and binds the CNOT7/CNOT8 subunits (Bai et al., 1999; Basquin et al., 2012; Petit et al., 2012). Moreover, the interaction between CNOT6/CNOT6L and CNOT7/CNOT8 via its LRR domain is essential for the stable recruitment of CNOT6/CNOT6L to the complex (Draper et al., 1995; Dupressoir et al., 2001; Mittal et al., 2011; Basquin et al., 2012).

The nuclease module appears to contain one subunit of CNOT6/CNOT6L and CNOT7/CNOT8.

CNOT7/CNOT8. can interact with BTG/TOB proteins. This family of six proteins are the best characterised interaction partners of Caf1 in vertebrates (Mauxion et al., 2009; Winkler, 2010). BTG/TOB protein interacts with Caf1 via their highly conserve amino-terminal BTG domain (Horiuchi et al., 2009). Studies demonstrated that these proteins when overexpressed inhibit cell cycle progression, which is dependent on interactions with Caf1 (Horiuchi et al., 2009; Doidge et al., 2012; Ezzeddine et al., 2012).

A model of the sub-complex composed of CNOT1 MIF4G domain, Caf1, Ccr4 and BTG2 is shown in Figure 1.12 C and D. Although the BTG1 structure has not been resolved, BTG1 is likely to interact with Caf1 in a similar manner as BTG1 and BTG2 are highly related in their BTG domain (74% identical) (Winkler, 2010).

In the work described in this chapter, the assembly of the sub-complex composed of the CNOT1 MIF4G domain, BTG1, CNOT6L, and CNOT7 was investigated. In particular this study aims to detect CNOT7 amino acidic residues essential for its function and for the interaction with the other components of the nuclease sub-complex. The approaches used in this study are summarised in Figure 5.1. Computational methods were used to select a pool of residues that resulted mutated in tumour samples and annotated in the Catalogue of Somatic Mutations In Cancer (COSMIC) database (Bamford et al., 2004; Forbes et al., 2008, 2011) and the Ensemble genome browser (Chen et al., 2010; Rios et al., 2010; Flicek et al., 2014). The impact of each amino acid substitution on protein function was predicted using the Sorting Intolerant From Tolerant (SIFT) package (Ng and Henikoff, 2003; Kumar et al., 2009). This approach, based on conservation of amino acid residues in sequence alignments derived from closely related sequences, distinguishes between tolerated mutations and mutation deleterious for the protein function. This method allowed to reduce the number of residues likely to play a critical role in the assembly and function of the nuclease sub-complex.

Furthermore, the Disease-Susceptibility-based SAV Phenotype Prediction software (SuSPect) (Yates et al., 2014) was used to predict the phenotypic effects of the mutations and their association with disease. The approach

is based on combined sequence and structural features, including network information to identify disease-associated single amino acid variants.

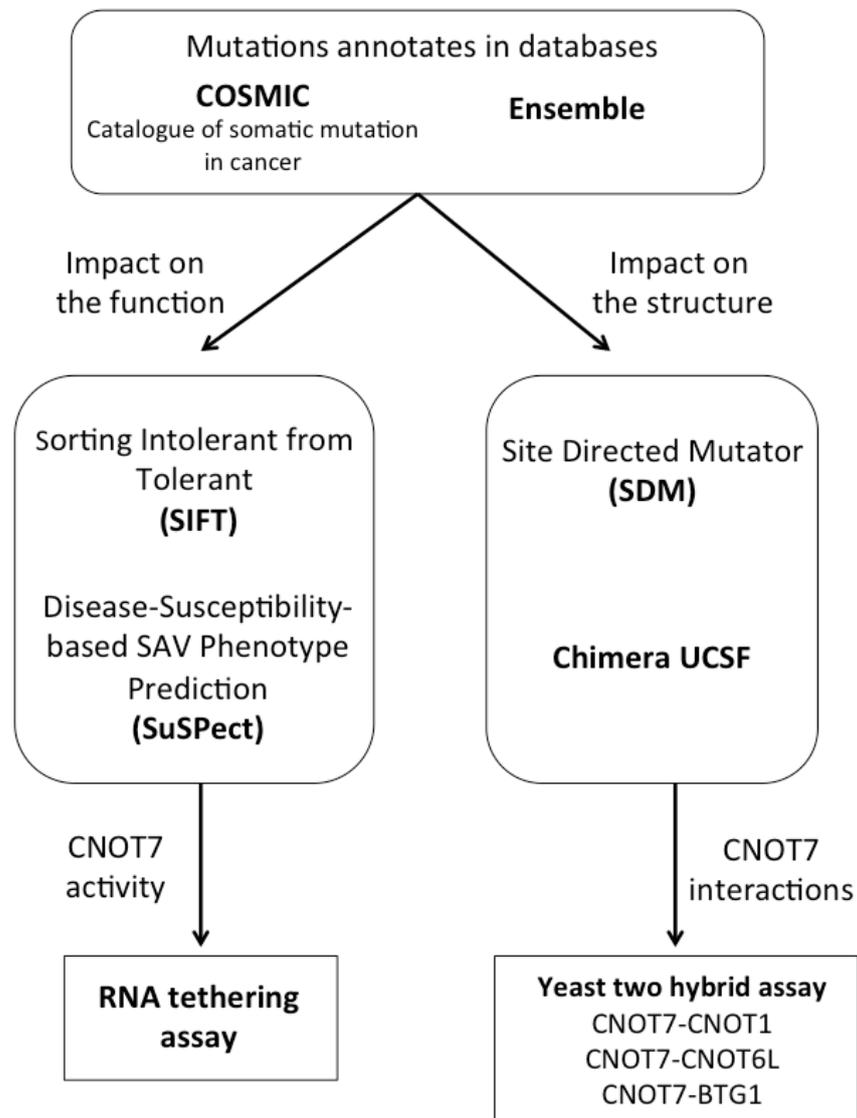
Moreover, the impact of the amino acid substitutions on the stability and structure of the CNOT7 was investigated in two ways.

The effect of single point mutations on the protein stability was predicted by Site Directed Mutator (SDM) (Worth et al., 2011). The package is based on observed substitutions occurring in homologous proteins and which are encoded in environment-specific substitution tables (ESSTs).

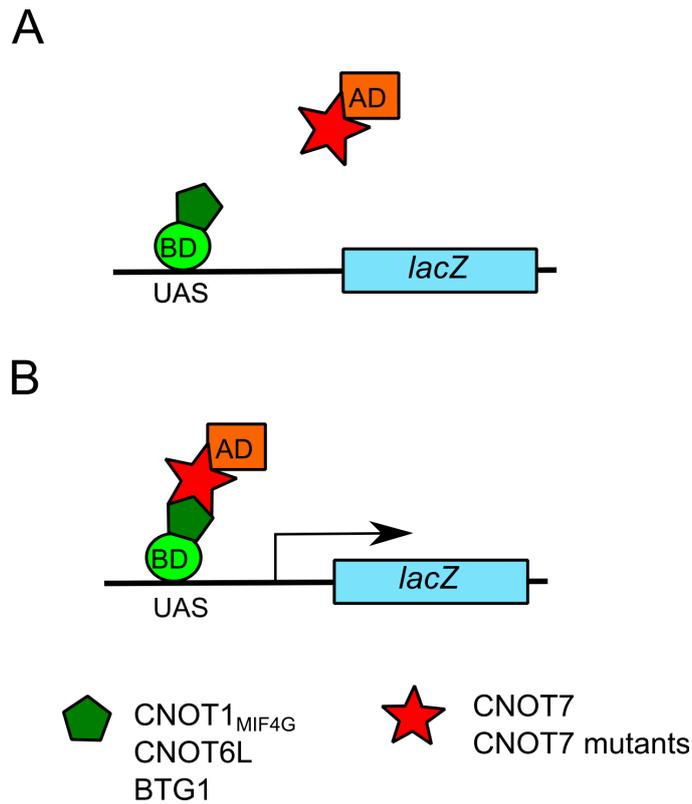
In order to investigate if the mutations can be accommodated into the CNOT7 structure model (4GMJ) without major rearrangements of the protein, UCSF Chimera was used (Pettersen et al., 2004).

These *in silico* predictions were indicative of the possible effects that amino acidic substitutions may have on the structure and function of the CNOT7. To complement these analyses, the impact of CNOT7 amino acid substitutions on the interactions with the MIF4G domain of CNOT1, CNOT6L, and BTG1 was experimentally verified using yeast two hybrid assays. This approach was used to study the interaction of the GAL4-AD-CNOT7 with GAL4-BD-CNOT1<sub>MIF4G</sub>, GAL4-BD-CNOT6L, and GAL4-BD-BTG1 fusion proteins (Figure 5.2). These analyses allowed the identification of residues critical for the interactions with the other subunits of the nuclease module.

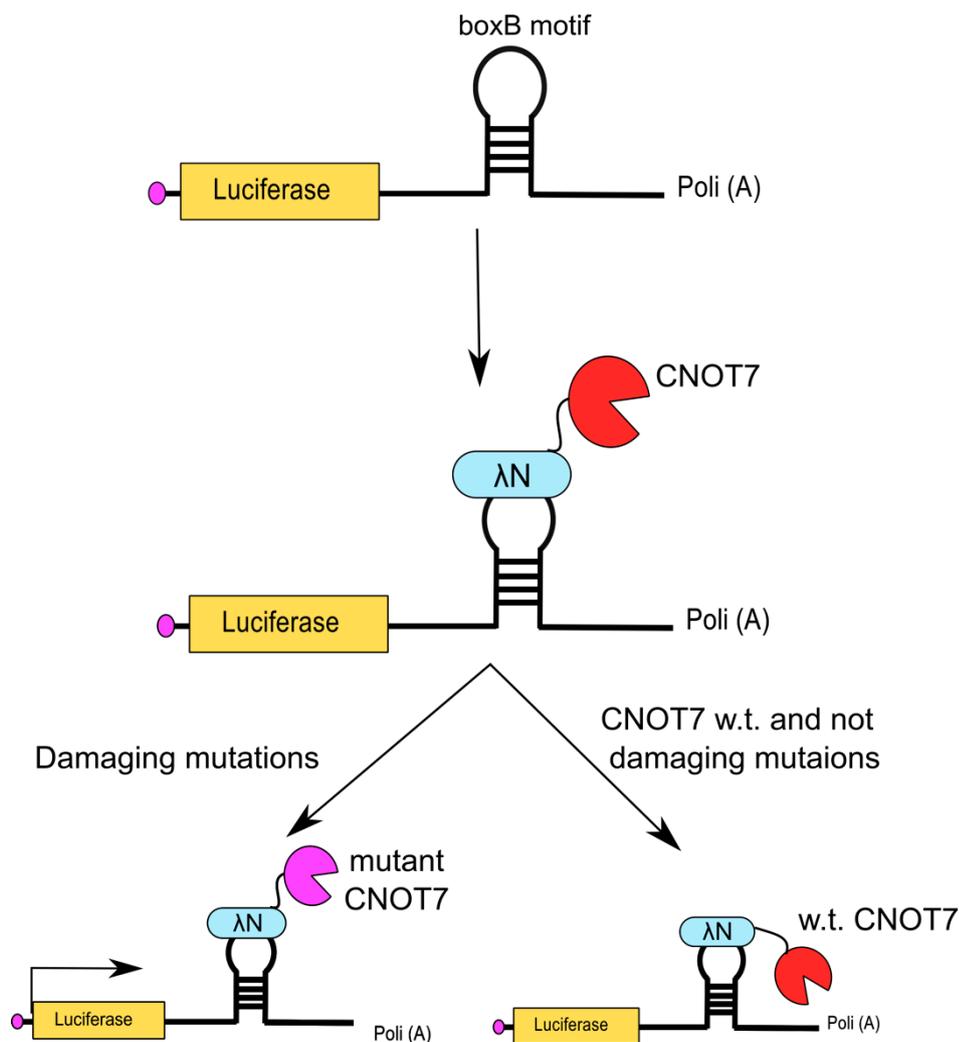
To assess the function of CNOT7 variants in mRNA degradation, RNA tethering assays were carried out. The tethering assay studies the function of RNA regulatory proteins independently of their ability to recognise and bind to RNA. The examined protein is expressed as a fusion protein with a tethering peptide ( $\lambda$ N) that recognises the 3'UTR of an mRNA reporter containing five boxB RNA motifs. Once the  $\lambda$ N fusion protein is bound to the 3'UTR, the function of the protein can be assessed by evaluating expression of the luciferase reporter gene. pCI $\lambda$ N-CNOT7 plasmid and its mutants were obtained by Gibson cloning as described in Materials and Methods. In case of wild type CNOT7, the reporter mRNA will be degraded resulting in reduced luciferase activity. Non-functional fusion proteins will not degrade the mRNA and the luciferase can be expressed (Figure 5.3).



**Figure 5.1 Overview of the approaches used for the analysis of sequence variants of CNOT7.** The scheme summarizes the approaches used in this study to characterise the impact of mutations in CNOT7 annotated in COSMIC and Ensemble. For the characterisation of the function, SIFT and SuSPect were used. The impact on the structure and the stability of the protein was predicted using Site Directed Mutator (SDM) and Chimera UCSF. Experimental approaches used to study the effect of the mutations on the protein function (RNA tethering assay) and on the interaction with other proteins of the nuclease sub-complex (yeast two-hybrid assay) are also indicated.



**Figure 5.2 Detection of CNOT7 interactions using yeast two hybrid analysis. (A)** The GAL4-BD hybrid protein (CNOT1<sub>MIF4G</sub>, CNOT6L, and BTG1) binds to the GAL4 UAS upstream to the *lacZ* reporter gene. **(B)** When AD-CNOT7 interacts with the BD hybrid proteins, the GAL4 AD and the GAL4 BD are brought in close proximity leading to expression of the reporter gene. In the absence of CNOT7 interactions, *lacZ* activity cannot be detected.



**Figure 5.3 RNA tethering assay for functional studies of CNOT7 variants.** CNOT7 was expressed as a fusion protein with a tethering peptide ( $\lambda$ N) that recognises the 3'UTR of an mRNA reporter containing five boxB motifs. Binding of wild type  $\lambda$ N-CNOT7 result in degradation of the mRNA resulting in a reduced expression of the luciferase gene. Non-functional CNOT7 variants will not degrade the mRNA and high luciferase activity can be detected.

## 5.2 Collection of mutations and prediction of their impact on the function of CNOT7

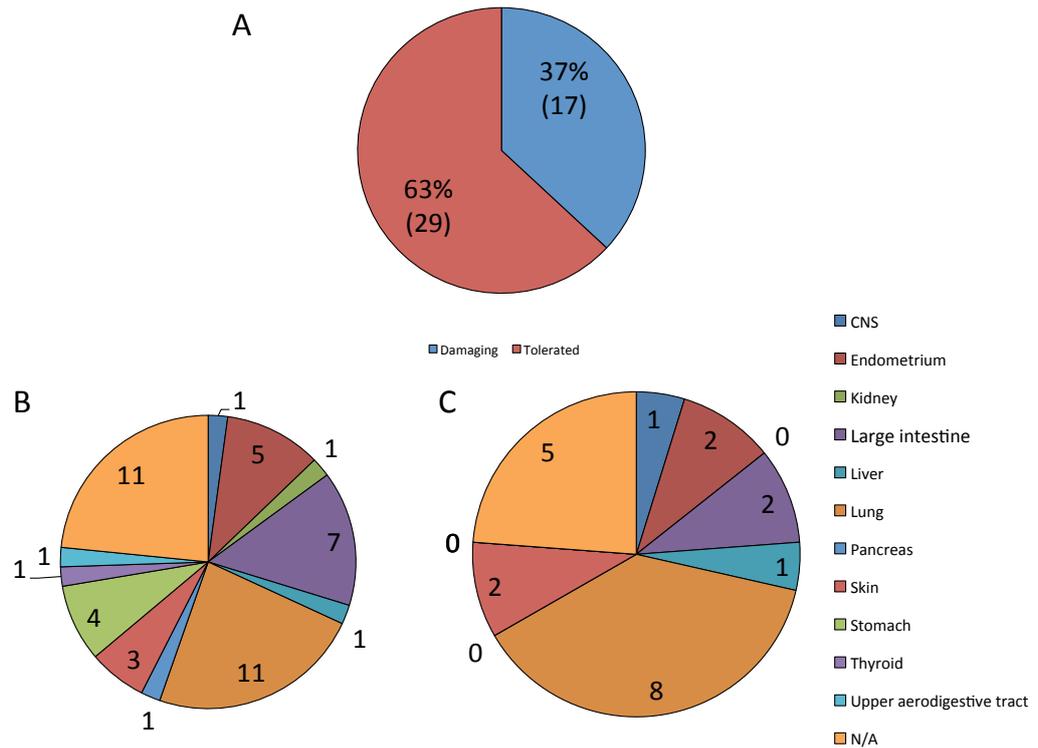
A list of somatic mutations of CNOT7 was retrieved from the Catalogue of somatic mutations in cancer (COSMIC v.72) and from Ensemble genome browser. COSMIC, as previously described, is the most comprehensive resource of information on somatic mutations in human cancer collecting data from literature and from the Cancer Genome Project at Sanger Institute (Bamford et al., 2004; Forbes et al., 2008, 2011). The Ensemble genome browser annotates genomic information of different species with emphasis on human and reporting human variants (Chen et al., 2010; Rios et al., 2010; Flicek et al., 2014). In order to predict the effects of single amino acidic variants on the function of the ribonuclease, the Sorting intolerant from tolerant (SIFT) package was used to distinguish between tolerated or damaging mutations (Supporting Table 5.1). The algorithm is based on the conservation of amino acid residues in sequence alignments derived from closely related sequences. Out of 46 somatic mutations reported in the databases, seventeen were found to be deleterious for the function (Figure 5.4 A). Tissue localisation of the 46 somatic mutations was retrieved and inspected where annotated. The tissue distribution was heterogeneous and mutations were found in the central nervous system (CNS) (one), endometrium (five), kidney (one), large intestine (seven), liver (one), lung (eleven), pancreas (one), skin (three), stomach (four), thyroid (one), and upper digestive tract (one) cancer samples. The tissue localisation of eleven mutations was not available in the databases (Figure 5.4 B).

After inspecting the tissue distribution of the mutations predicted to be functionally deleterious by SIFT, the highest number was found in lung (eight), then in endometrium (two), large intestine (two), skin (two), and central nervous system (one). Annotation of the remaining five mutations was not available in the databases (Figure 5.4 C). Mutations (n=21) selected for further analyses are reported in Table 5.1, along with their SIFT prediction scores indicative of the effects on the protein function (damaging or tolerated), and their tissue localisation.

In addition to the damaging mutations, four mutations predicted to be tolerated were inspected (M24I, A58V, E87Q, and E217Q).

In addition to SIFT, Disease-Susceptibility-based SAV Phenotype Prediction (SuSPect) was also used to predict the phenotypic effects of the mutations and their association with disease (Yates et al., 2014). This algorithm is based on combined sequence and structural features, including network information to identify disease-associated single amino acid variants. The chosen features are related to the differences between *disease-susceptible* and *disease-resistant* domains and proteins. SuSPect provides a score (0-100) indicating the mutation deleteriousness and its association with disease. A score of 50 was used as cut-off between neutral and deleterious mutations. SuSPect scores of the mutations selected on the basis of SIFT. Predictions are reported in Table 5.1 ; all other SuSPect substitution scores are reported in Supporting Table 5.2. A schematic of the mutations mapped on the linear structure of CNOT7 is shown in Figure 5.5A.

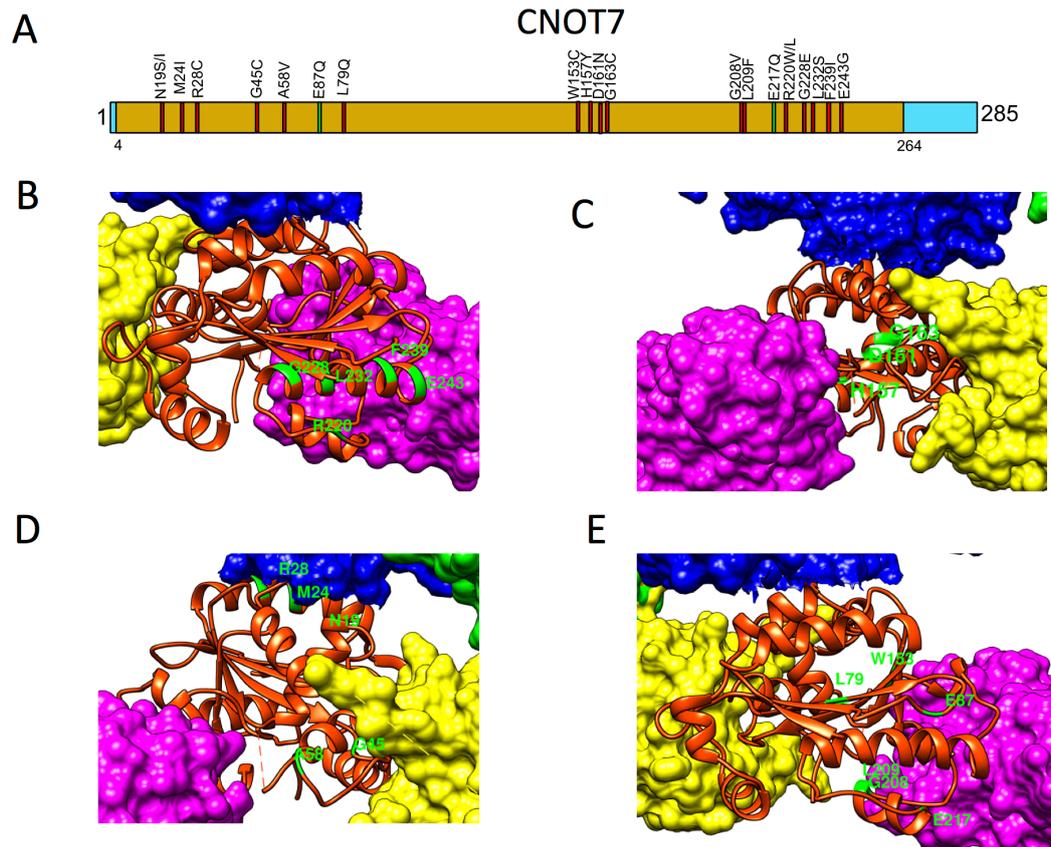
Mutations were also mapped on the crystal structure of the complex composed of CNOT7, CNOT1 MIF4G domain, CNOT6L, and BTG2 and were arranged in four groups (I to IV). Group I (R220W, R220L, G228E, L232S, F239I, and E243G) is composed of mutations mapping to the C-terminus of CNOT7 (Figure 5.5 B); group II (H157Y, D161N, G163C) comprises mutations proximal to the catalytically active site of the ribonuclease (Figure 5.5 C); group III (N19S, N19I, M24I, R28C, G45C, A58V) constitutes mutations mapping to the N-terminus (Figure 5.5 D); and group IV is composed of mutations (E87Q, L79Q, W153C, G208V, L209F, and E217Q) mapping to the region next to the catalytically active site (Figure 5.5 E).



**Figure 5.4 Overview of CNOT7 mutations annotated in COSMIC and Ensemble.** The impact of mutations annotated in COSMIC and Ensemble were investigated using the Sorting Intolerant From Tolerant (SIFT) algorithm. **(A)** The tool distinguished between 29 tolerated mutations and found 17 mutations predicted to be damaging CNOT7 function. **(B)** Tissue localisation of all mutations. **(C)** Tissue localisation of the mutations predicted to be damaging by SIFT.

CNOT7 Variant	SIFT score	Effect	SuSPect score	Disease association	Tissue localisation
N19I	0	Damaging	97	Yes	Lung cancer
N19S	0.002	Damaging	94	Yes	N/A
M24I	0.087	Tolerated	53	Yes	N/A
R28C	0.007	Damaging	66	Yes	Skin, Large intestine
G45C	0	Damaging	92	Yes	Lung cancer
A58V	0.26	Tolerated	16	No	Endometrium
L79Q	0	Damaging	90	Yes	Liver
E87Q	0.131	Tolerated	3	No	N/A
W153C	0	Damaging	87	Yes	N/A
H157Y	0	Damaging	27	No	Skin
D161N	0.02	Damaging	66	Yes	Pancreas
G163C	0.016	Damaging	36	No	N/A
G208V	0.001	Damaging	64	Yes	Lung cancer
L209F	0	Damaging	80	Yes	Lung cancer
E217Q	0.382	Tolerated	7	No	Endometrium
R220W	0	Damaging	19	No	Lung cancer
R220L	0	Damaging	22	No	Lung cancer
G228E	0	Damaging	89	Yes	Large intestine
L232S	0	Damaging	66	Yes	Lung cancer
F239I	0.022	Damaging	23	No	Lung cancer
E243G	0.019	Damaging	26	No	CNS

**Table 5.1 Selected CNOT7 variants.** List of CNOT7 variants selected for further studies. SIFT scores are reported in the table; mutations were considered damaging when the SIFT score was <0.05. The SuSPect software was used to infer the phenotypic effect (disease-association) of the mutations. SuSPect scores are reported and a cut-off score of 50 was used to distinguish between disease-associated (Yes) or non-disease-associated (No) mutations. Moreover, tissue localisation, as annotated in the databases, is reported.



**Figure 5.5 Map of selected CNOT7 variants.** The mutations selected for the further studies were mapped to the nuclease module model composed by CNOT7 (orange), CNOT1<sub>MIF4G</sub> (blue), CNOT6L (yellow), and BTG2 (pink). **(A)** Schematic of CNOT7 linear protein structure; damaging mutations are represented as red residues, tolerated mutations were coloured in green. **(B)** Mutations were arranged in four groups and mapped (green residues) on the nuclease sub-complex crystal structure. Group I mapped to the C-Terminus of CNOT7 (R220W, R220L, G228E, L232S, F239I, and E243G). **(C)** Group II are located in close proximity of the catalytic site (H157Y, D161N, and G163C). **(D)** Group III are localised in the N-terminus of CNOT7 (N19S, N19I, M24I, R28C, G45C, and A58V). **(E)** Group IV (E87Q, L79Q, W153C, G208V, L209F, and E217Q) map to the region surrounding the catalytically active residues.

### 5.3 Prediction of the impact of the mutations on the structure of CNOT7.

Prediction of the effects of the mutations on the integrity and structure of CNOT7 was performed by Site Directed Mutator (SDM) (Worth et al., 2011) and Chimera UCSF (Pettersen et al., 2004). Site Directed Mutator evaluates the effect of single point mutations on the protein stability and distinguishes disease-associated mutations from non-disease associated mutations. The package is based on observed substitutions occurring in homologous proteins and which are encoded in environment-specific substitution tables (ESSTs). It calculates a stability score, which is analogous of the free energy difference between the wild-type and the mutant (Topham et al., 1997).

The protein stability analysis by SDM was performed on all the retrieved mutations and provided information on possible changes in the secondary structure, solvent accessibility expressed in percentage, and the stability score (pseudo  $\Delta\Delta G$ ). A negative stability score indicates that the mutation is destabilising, whereas a positive score indicates a stabilising effect (Table 5.2). The analysis predicted amino acid substitutions L232S and E243G as highly destabilising and L79Q destabilising. Thus, these variants are predicted to affect the function of the protein and to be associated with disease. Variants N19S, G45C, and W153 were predicted to be destabilising but not classified as disease-associated.

Amino acid substitutions A58V, D161N, L209F, G228E, and F239I were predicted to be neutral and not disease associated. Variants N19I, R28C, E87Q, and H157Y were found to be slightly stabilizing and not disease-associated. Interestingly the H157Y mutant was predict to have a change in solvent accessibility from accessible (58.3%) to partially accessible (26.60%). G208V, E217Q, R220L, and R200W were predicted to be stabilizing and not disease-associated. The G163C mutation was the only one to be highly stabilizing and predicted to malfunction and be disease associated (Table 5.2).

Wild type		Mutant			
	2nd Structure	w.t. Solvent accessibility	Solvent accessibility	$\Delta\Delta G$	Mutation effect
L232S	Helix $\alpha$ 12	0.1%	4%	-4.56	Highly destabilizing
E243G	Helix $\alpha$ 12	48.1%	49.20%	-3.18	Highly destabilizing
L79Q	Sheet $\beta$ 3	1.2%	2%	-2.11	Destabilizing
G45C	Loop-irregular	32.4%	26%	-1.97	Destabilizing
N19S	Loop-irregular	26.5%	36.60%	-1.41	Destabilizing
W153C	Sheet $\beta$ 5	1.7%	2.80%	-1.32	Destabilizing
F239I	Helix $\alpha$ 12	10.4%	20.20%	-0.43	Neutral
A58V	Helix $\alpha$ 3	53.3%	69.70%	-0.38	Neutral
L209F	Helix $\alpha$ 11	10.3%	16.10%	-0.34	Neutral
D161N	Helix $\alpha$ 7	4.7%	6.40%	-0.17	Neutral
G228E	Helix $\alpha$ 12	0.0%	0%	0.33	Neutral
H157Y	Bend	58.3%	26.60%	0.51	Slightly stabilizing
N19I	Loop-irregular	26.5%	36.50%	0.58	Slightly stabilizing
E87Q	Loop-irregular	70.0%	68.80%	0.63	Slightly stabilizing
R28C	Helix $\alpha$ 2	48.8%	32.10%	0.67	Slightly stabilizing
E217Q	Loop-irregular	90.2%	93.40%	1.1	Slightly stabilizing
R220L	Loop-irregular	35.5%	27.20%	1.14	Slightly stabilizing
R220W	Loop-irregular	35.5%	27.70%	1.29	Slightly stabilizing
G208V	Loop-irregular	77.8%	67.10%	1.63	Slightly stabilizing
G163C	Helix $\alpha$ 7	0%	0%	2.91	Highly stabilizing

**Table 5.2 Predicted stability of CNOT7 variants.** Protein stability was investigated for all selected mutations using the Site Directed Mutator (SDM) package. The analysis provided information on changes in the secondary structure, solvent accessibility expressed in percentage, and the stability score (Pseudo  $\Delta\Delta G$ ). A negative stability score indicates that the mutation is destabilising, whereas a positive score indicates that the mutation is stabilising.

In order to investigate if the amino acid substitutions can be accommodated into the CNOT7 structure model (PDB accession number 4GMJ) without major rearrangements of the protein, UCSF Chimera was used (Pettersen et al., 2004). This package is used for molecular visualisation and provides several tools for structure exploration and editing of protein models.

The analyses identified which amino acid substitutions could be accommodated in the protein model without clashes with neighbouring residues. Only amino acid rotamers present in the rotamers library with a probability >5% were inspected (Dunbrack, 2002). Once the residues were altered, clashes with neighbouring atoms were investigated and where possible resolved. Most of the amino acid substitutions could be accommodated within the crystal structure (Supplementary Table 5.3). Six mutations though could not be accommodated in the CNOT7 structure. The clashing rotamers were annotated with the side chain torsion angles (Chi1, Chi2, and Chi3) and their probability (Table 5.3). The mutation G45C presented two cysteine rotamers that could not be accommodated without major rearrangements of the protein structure. These cysteine rotamers were present in the rotamer library with Chi1 -69.2 and probability 84%, and Chi1 -174.2 and probability 9.5% respectively. In both cases, clashes with Ser112 were found (Figure 5.6 A). The CNOT7 H157Y mutation could be accommodated with all tyrosine rotamers except for a rotamer with Chi1 -61.4 and Chi2 99.4 found in the 8.7% of the cases. This tyrosine clashed with Phe147. CNOT7 His157 could likely be substituted by a tyrosine without major atomic rearrangements of the structure (Figure 5.6 B).

The mutation D161N could be accommodated in most cases. Two asparagine rotamers with Chi1 -176.9, Chi2 64.2, and probability 15% and with Chi1 -171.6, Chi2 30.9, and probability 8% were found to clash with Met39. Thus the mutation can likely be accommodated in the structure without major structural rearrangements (Figure 5.6 C).

Furthermore, the amino acid substitution G163C was found not to interfere with the structure of the protein. In the remaining cases the

cysteine rotamer with Chi1 -177.4 and probability 33%, was found to clash with Leu174 (Figure 5.6 D).

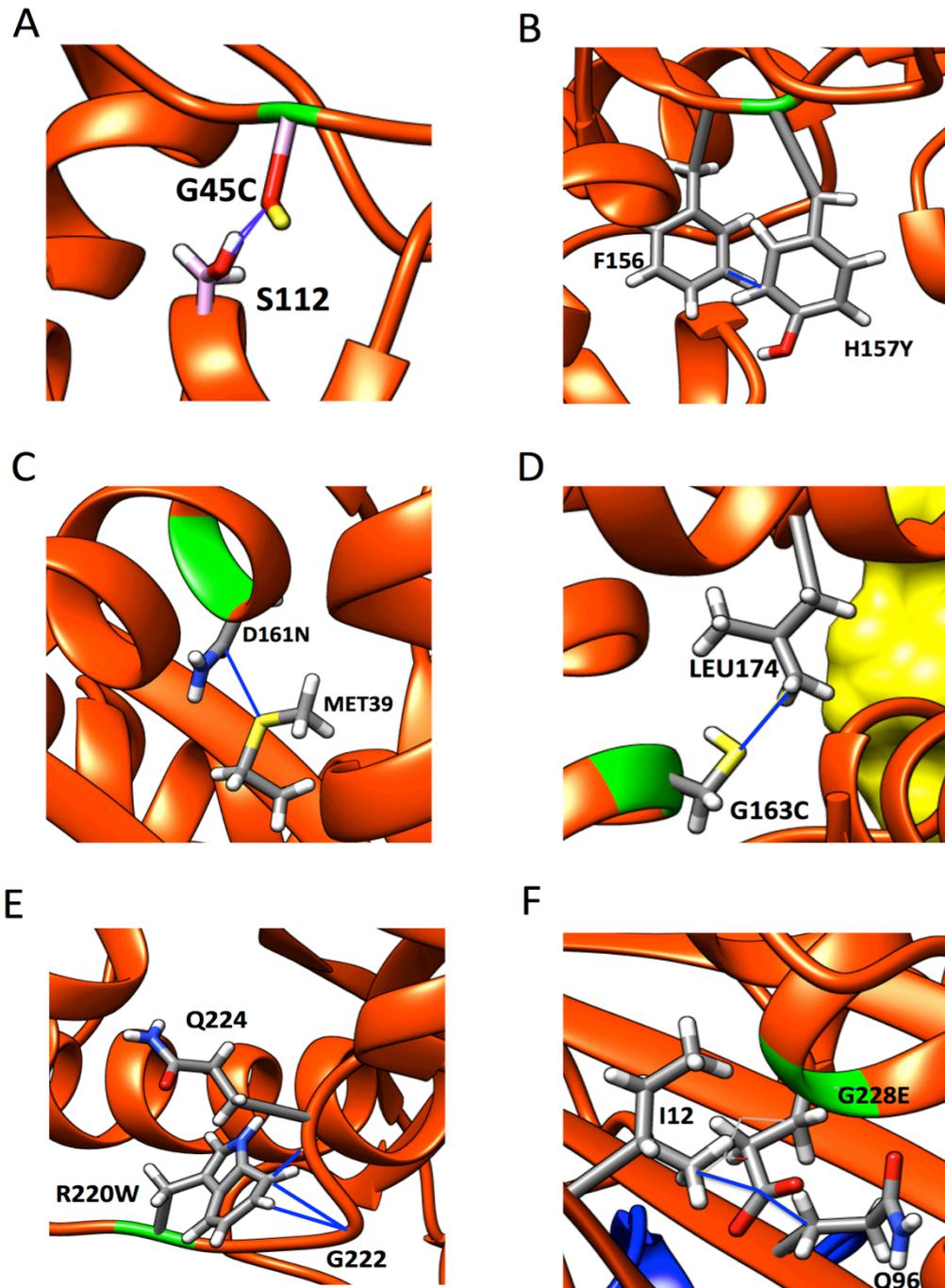
The mutation R220W was inspected. In three cases tryptophan rotamers clashed with surrounding atoms. Two tryptophan rotamers with side chain torsion Chi1 -69.1 and Chi2 102.2, and Chi1 -70.5 and Chi2 2.7 were found to clash with Gln210 respectively with 59% and 14% probability. The tryptophan rotamer with Chi1 178.4, Chi2 85.8, and probability 8.1%, was found to clash with Gln224 and Gly222. Hence in the 81.1% of the cases, Arg220 cannot be substituted with tryptophan without causing major changes in the protein structure (Figure 5.6 E).

Finally, G228E was investigated. The glutamic acid rotamers with Chi1 -177.4, Chi2 177.6, and probability 17% and the rotamer with Chi1 -68.8, -179.4, and probability 7.1% were predicted to clash with Gln96 and Ile12. In the remaining 75.9% of the cases, the glutamic acid could be accommodated in the structure without causing any structural changes (Figure 5.6 F).

None of the amino acid substitutions inspected were predicted to directly cause clashes with CNOT6L, CNOT1, or BTG1 residues.

Mutation	Side chain torsion		Probability (%)	Clashing residues
	Chi1	Chi2		
G45C	-62.9		84	SER112
	-174.2		9.5	SER112
H157Y	-61.4	99.4	8.7	PHE156,
D161N	-176.9	64.2	15	MET39
	-171.6	30.9	8	MET39
G163C	-177.4		33	LEU174
R220W	-69.1	102.2	59	GLN210
	-70.5	2.7	14	GLN210
	178.4	85.8	8.1	GLN224, GLY222
G228E	-177.4	177.6	17	GLN96, ILE12
	-68.8	-179.4	7.1	GLN96, ILE12

**Table 5.3 Accommodation of CNOT7 variants.** The effects of amino acid substitutions on CNOT7 structure was investigated. Reported are the six variants whose side chain cannot be accommodated in the reference structure of CNOT7 without causing rearrangements. For each residue the side chain torsions of the considered rotamers are indicated, along with the probability of being found in the rotamer library (Dunbrack, 2002) and the residues clashing.



**Figure 5.6 Clashes caused by amino acid substitutions in CNOT7.**

Rotamers clashes caused by amino acid substitutions were: **(A)** G45C, **(B)** H157Y, **(C)** D161N, **(D)** G163C, **(E)** R220W, **(F)** and G228E. Amino acid substitutions are highlighted in green. Clashes between atoms are indicated using blue lines.

## 5.4 The effect of CNOT7 mutations on protein-protein interactions in the nuclease sub-complex

### 5.4.1 Generation of plasmids for yeast two hybrid assays

In order to investigate protein-protein interactions of CNOT7 variants with CNOT1, CNOT6L, and BTG1, a yeast two hybrid system based on GAL4 DNA-binding domain/activation domain fusions was used. To this end, plasmids pAD-GAL4-2.1-CNOT7, pBD-GAL4-HA-CNOT1<sub>MIF4G</sub> and pBD-GAL4-HA-CNOT6L were generated. The pBD-GAL4-HA-BTG1 plasmid was provided by Miss Hibah Almasmoum. The CNOT7 cDNA was amplified to include XhoI restriction sites at the 5' and 3' ends. The PCR product was then isolated by gel electrophoresis and digested with XhoI. The pAD-GAL4 2.1 vector was digested with XhoI restriction enzyme, treated with Antarctic Phosphatase in order to prevent self-ligation and ligated with the digested CNOT7 cDNA. The selected 21 mutants were obtained by site-directed mutagenesis of the pAD-GAL4-CNOT7 plasmid (Figure 5.7 A). The cDNAs encoding HA-tagged versions of the MIF4G domain of CNOT1, CNOT6L, and BTG1 were amplified by PCR to include the EcoRI restriction site at the 5' and the Sall restriction site at the 3' of the sequences. The amplified sequences were then isolated by gel electrophoresis and digested with the EcoRI and Sall restriction enzymes. The pBD-GAL4-HA Cam vector was digested with EcoRI and Sall and treated with Antarctic Phosphatase. Ligation reactions containing the digested CNOT1<sub>MIF4G</sub> and CNOT6L cDNA sequences and the digested pBD-GAL4-HA were then set up. The generation of the pAD-GAL4-2.1-CNOT7 and its mutants, pBD-GAL4-HA-CNOT1<sub>MIF4G</sub> (Figure 5.7 B), pBD-GAL4-HA-CNOT6L (Figure 5.7 C), and pBD-GAL4-HA-BTG1 (Figure 5.7 D) was confirmed by colony PCRs, restriction enzyme digestions, and DNA sequencing. The plasmids were then used to transform *S. cerevisiae* YRG2 and perform the yeast two-hybrid assays. Unfortunately plasmids pAD-GAL4-CNOT7 L79Q and pAD-GAL4-CNOT7 L209F did not allow growth of yeast colonies and were then not used for any further analyses reducing the number of analysed variants to nineteen.

#### 5.4.2 Identification of CNOT7 residues required for the interaction with the MIF4G domain of CNOT1

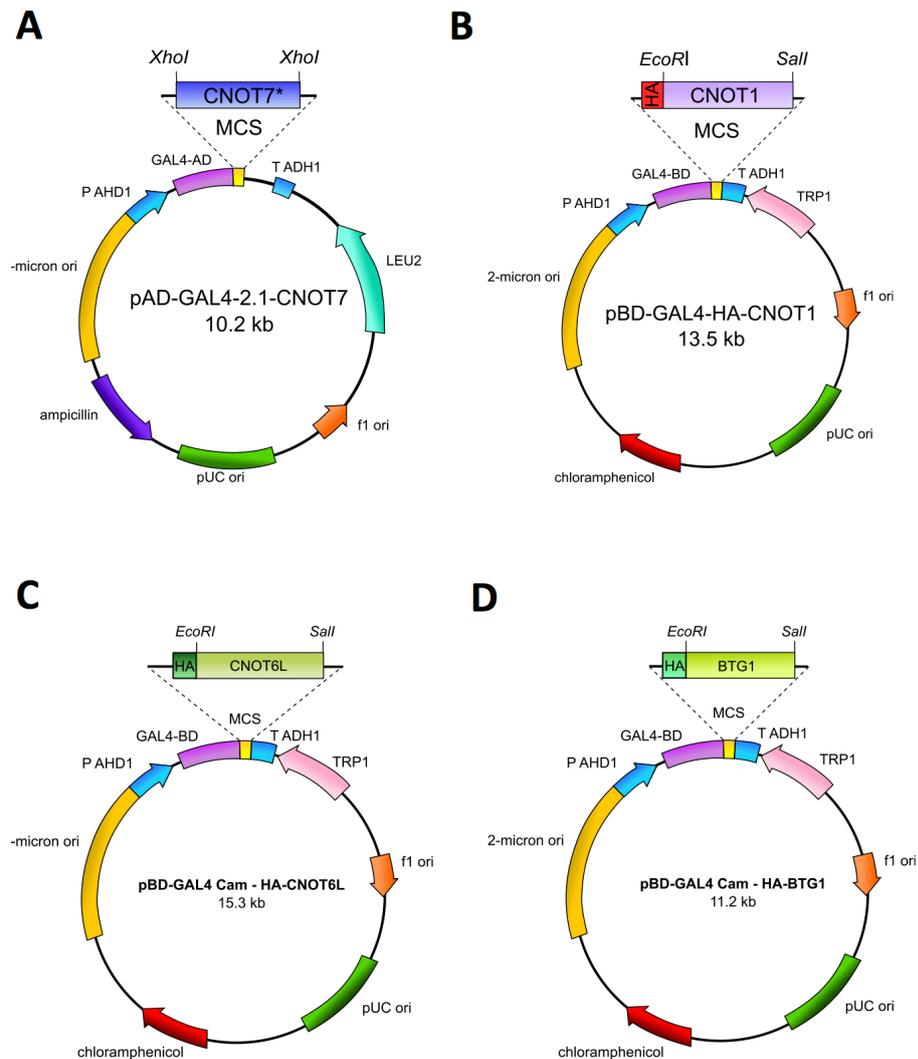
To investigate the effect of the mutations on the interaction of CNOT7 and the MIF4G domain of CNOT1, plasmids pAD-GAL4-CNOT7 and its mutants, and pBD-GAL4-HA-CNOT1<sub>MIF4G</sub> were transformed into yeast strain YRG2 and grown on media without leucine and tryptophan to select transformants. Transformation of pBD-GAL4-HA only and pAD-GAL4-CNOT7 confirmed that CNOT7 and its mutants without the CNOT1<sub>MIF4G</sub> fused to the GAL4 DNA binding domain was not able to activate  $\beta$ -galactosidase expression.

The CNOT7 mutations were analysed by grouping them as previously described. Of the six mutation analysed in the Group I, R220W, L232S, F239I, and E243G did not induce  $\beta$ -galactosidase activity and thus were unable to interact with BD-HA-CNOT1<sub>MIF4G</sub> (Figure 5.8 A). From the yeast cultures used for the  $\beta$ -galactosidase assay, cellular lysates were analysed by western blot to confirm the expression of BD-HA-CNOT1<sub>MIF4G</sub> and the CNOT7 mutants (Figure 5.8 B).

In Group II, CNOT7 variants H157Y, D141N, and G163C retained the ability to interact with CNOT1<sub>MIF4G</sub> and to induce  $\beta$ -galactosidase expression (Figure 5.9 A). Expression of BD-HA-CNOT1<sub>MIF4G</sub> and the CNOT7 mutants was confirmed by western blot of yeast lysates (Figure 5.9 B).

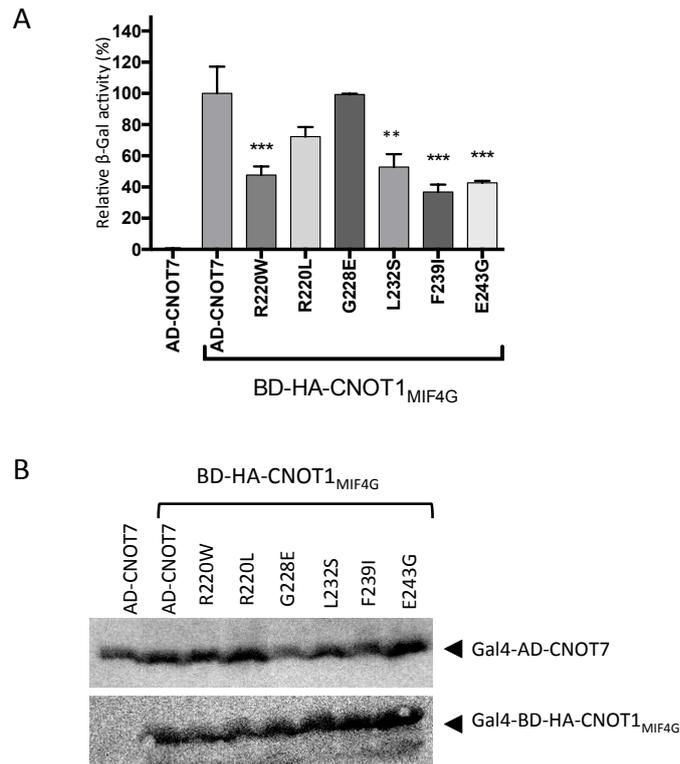
For Group III, two variant, N19S and N19I, displayed reduced interactions with BD-HA-CNOT1<sub>MIF4G</sub> (Figure 5.10 A). From yeast cultures used for the  $\beta$ -galactosidase assay, cellular lysates were analysed by western blot to confirm the expression of BD-HA-CNOT1<sub>MIF4G</sub> and the CNOT7 mutants. Interestingly, the variants N19S and N19I were found to have a reduced expression level (Figure 5.10 B).

In Group IV, two variant, W153C and G208V, displayed reduced  $\beta$ -galactosidase expression (Figure 5.11 A). From yeast cultures used for the  $\beta$ -galactosidase assay, cellular lysates were analysed by western blot to confirm the expression of BD-HA-CNOT1<sub>MIF4G</sub> and the CNOT7 mutants (Figure 5.1 B).

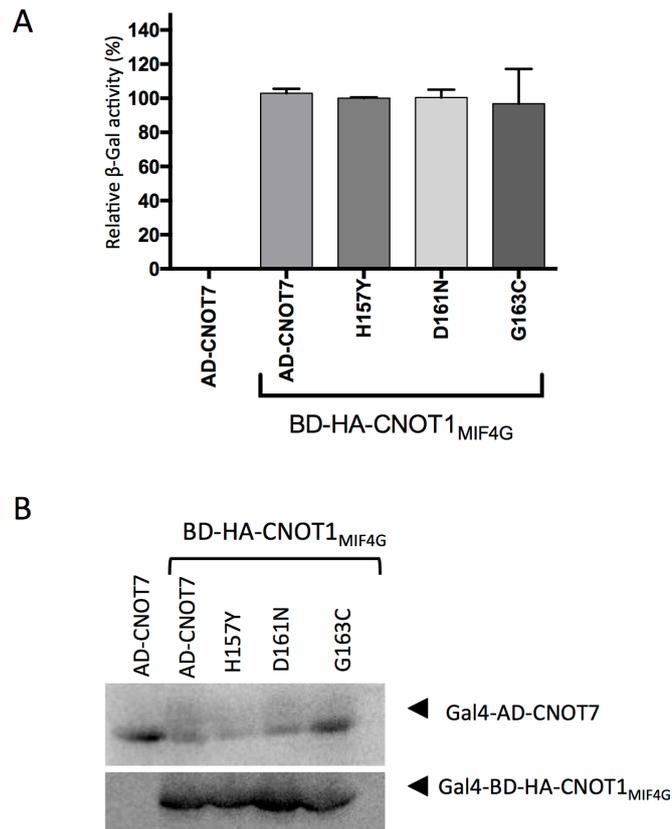


**Figure 5.7 Generation of plasmids for yeast two hybrid analysis. (A)** The pAD-GAL4-2.1-CNOT7 plasmid was generated by cloning CNOT7 cDNA using the XhoI restriction site of the multiple cloning sequence (MCS). The pAD-GAL4 2.1 plasmid includes yeast ADH1 promoter (P ADH1), GAL4 activating domain (GAL4-AD), yeast ADH1 terminator (T ADH1), yeast selection gene LEU2 ORF, f1 origin, pUC origin, ampicillin resistance ORF, and 2  $\mu$  yeast origin. **(B)** The pBD-GAL4-Cam-HA-CNOT1 plasmid was generated by inserting a cDNA encoding HA-tagged CNOT1MIF4G into the multiple cloning sequence (MCS) using the EcoRI and Sall restriction sites. Plasmid pBD-GAL4-Cam has the same features of the pAD-GAL4 plasmid except of the selection genes and the GAL4 binding domain (BD). The tryptophan gene (TRP1) is used for selection in yeast, and the chloramphenicol resistance gene is used for selection in bacteria.

**(C)** Plasmid pBD-GAL4-Cam-HA-CNOT6L was generated by inserting a cDNA encoding HA-tagged CNOT6L into the multiple cloning sequence (MCS) using the EcoRI and Sall restriction sites. **(D)** Plasmid pBD-GAL4-Cam-HA-BTG1 was generated by inserting a cDNA encoding HA-tagged BTG1 into the EcoRI and Sall restriction sites of the MCS.

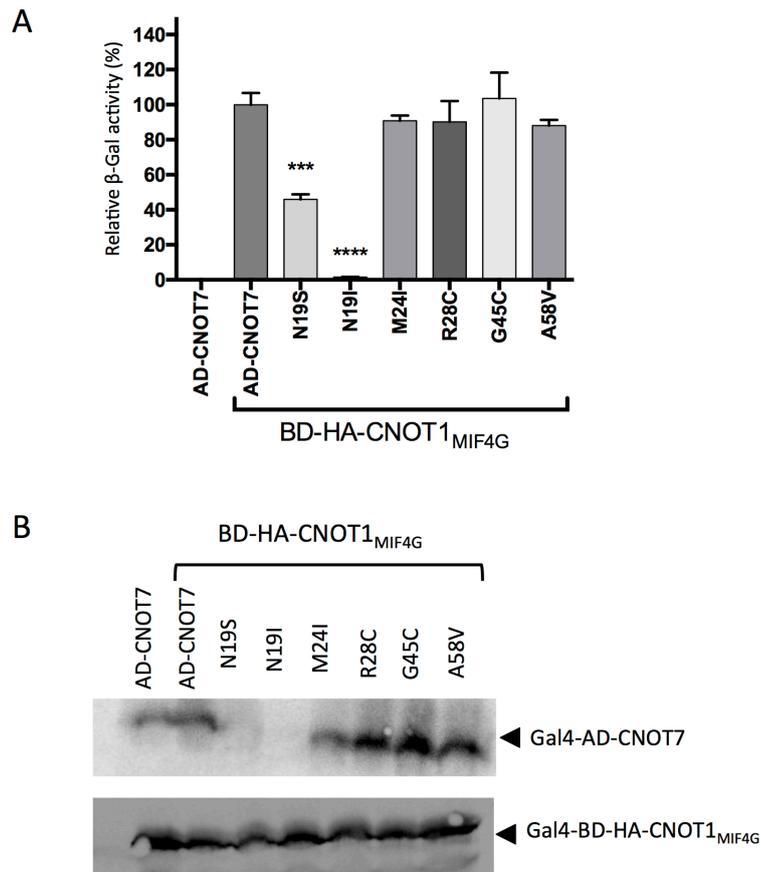


**Figure 5.8 Yeast two-hybrid analysis of the CNOT1<sub>MIF4G</sub> interaction with CNOT7 variants (Group I).** **(A)** YRG2 yeast cells were transformed with the indicated vectors; pAD-Gal4-CNOT7 and its variants, and pBD-Gal4-HA or pBD-Gal4-HA- CNOT1<sub>MIF4G</sub>.  $\beta$ -Galactosidase activity was normalised against total protein concentration. Error bars represent the standard error of the mean (n=3) and the P value is indicated (\*\*  $\leq$  0.01 and \*\*\*  $\leq$  0.001). **(B)** Western blot analysis showing expression of Gal4-BD-HA-CNOT1<sub>MIF4G</sub> and Gal4-AD-CNOT7. Anti-HA antibodies were used to detect Gal4-BD-HA-CNOT1<sub>MIF4G</sub>; anti-Gal4-AD antibodies were used to detect Gal4-AD-CNOT7.

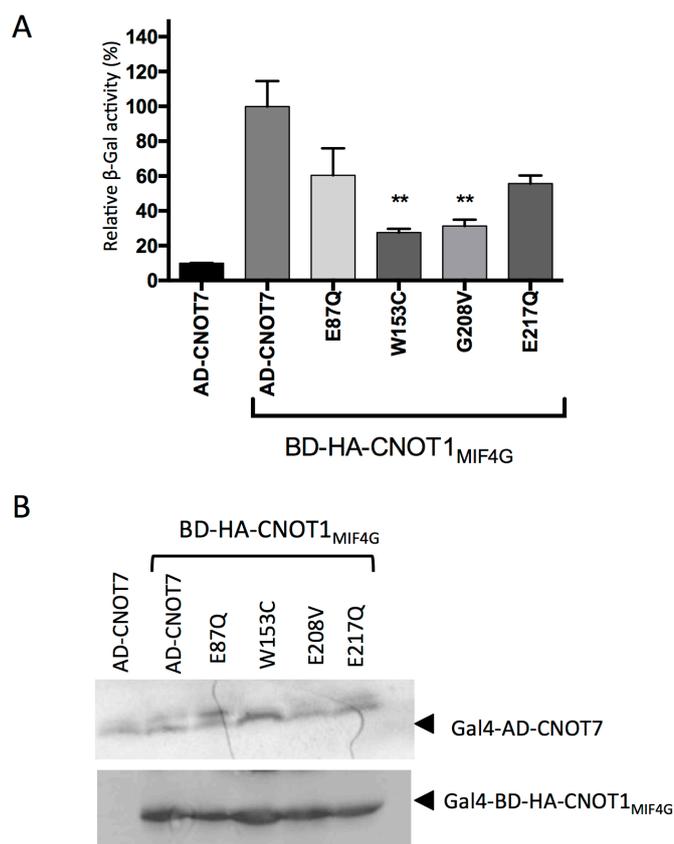


**Figure 5.9 Yeast two-hybrid analysis of the CNOT1<sub>MIF4G</sub> interaction**

**with CNOT7 variants (Group II).** (A) YRG2 yeast cells were transformed with the indicated vectors; pAD-Gal4-CNOT7 and its variants, and pBD-Gal4-HA or pBD-Gal4-HA- CNOT1<sub>MIF4G</sub>.  $\beta$ -Galactosidase activity was normalised against total protein concentration. Error bars represent the standard error of the mean (n=3). (B) Western blot analysis showing the expression of Gal4-BD-HA-CNOT1<sub>MIF4G</sub> and Gal4-AD-CNOT7. Anti-HA antibodies were used to detect Gal4-BD-HA-CNOT1<sub>MIF4G</sub>; anti-Gal4-AD antibodies were used to detect Gal4-AD-CNOT7.



**Figure 5.10 Yeast two-hybrid analysis of the CNOT1<sub>MIF4G</sub> interaction with CNOT7 variants (Group III).** (A) YRG2 yeast cells were transformed with the indicated vectors; pAD-Gal4-CNOT7 and its variants, and pBD-Gal4-HA or pBD-Gal4-HA-CNOT1<sub>MIF4G</sub>.  $\beta$ -Galactosidase activity was normalised against total protein concentration. Error bars represent the standard error of the mean (n=3) and the P value is indicated (\*\* $\leq 0.001$  and \*\*\*\* $\leq 0.0001$ ). (B) Western blot analysis showing the expression of Gal4-BD-HA-CNOT1<sub>MIF4G</sub> and Gal4-AD-CNOT7. Anti-HA antibodies were used to detect Gal4-BD-HA-CNOT1<sub>MIF4G</sub>; anti-Gal4-AD antibodies were used to detect Gal4-AD-CNOT7.



**Figure 5.11 Yeast two-hybrid analysis of the CNOT1<sub>MIF4G</sub> interaction with CNOT7 variants (Group IV).** (A) YRG2 yeast cells were transformed with the indicated vectors; pAD-Gal4-CNOT7 and its variants, and pBD-Gal4-HA or pBD-Gal4-HA-CNOT1<sub>MIF4G</sub>.  $\beta$ -Galactosidase activity was normalised against total protein concentration. Error bars represent the standard error of the mean (n=3) and the P value is indicated (\*\*  $\leq 0.01$ ). (B) Western blot analysis showing the expression of Gal4-BD-HA-CNOT1<sub>MIF4G</sub> and Gal4-AD-CNOT7. Anti-HA antibodies were used to detect Gal4-BD-HA-CNOT1<sub>MIF4G</sub>; anti-Gal4-AD antibodies were used to detect Gal4-AD-CNOT7.

### 5.4.3 Identification of CNOT7 residues required for the interaction with CNOT6L

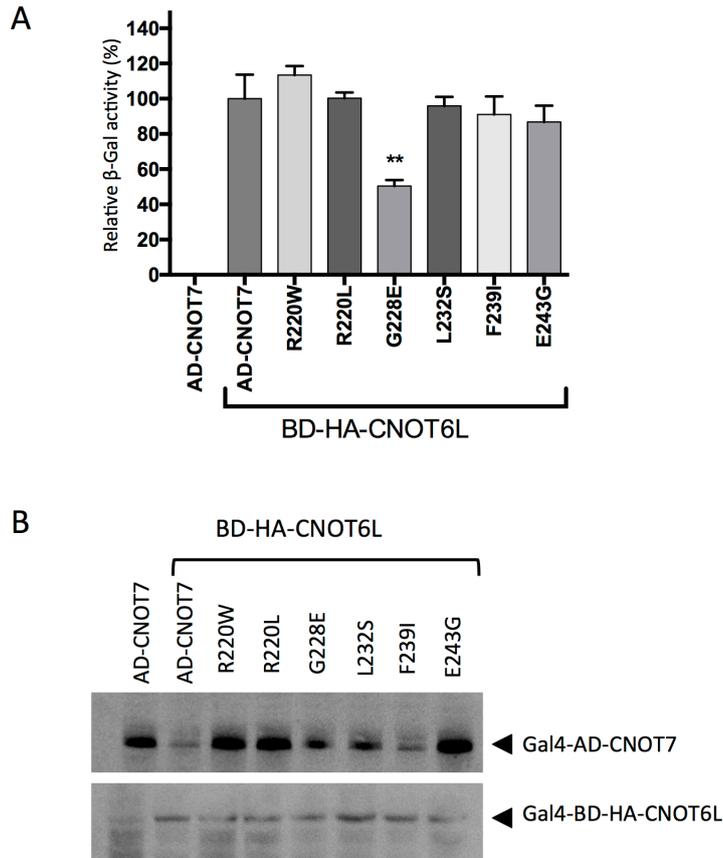
Next, the effect of the CNOT7 mutations on the interaction with the CNOT6L subunit was assessed.

The mutations in Group I induced  $\beta$ -galactosidase activity except for the G228E variant that resulted in reduced interactions with CNOT6L (Figure 5.12 A). Cellular lysates were analysed by western blot to confirm the expression of BD-HA-CNOT6L and the CNOT7 mutants (Figure 5.12 B).

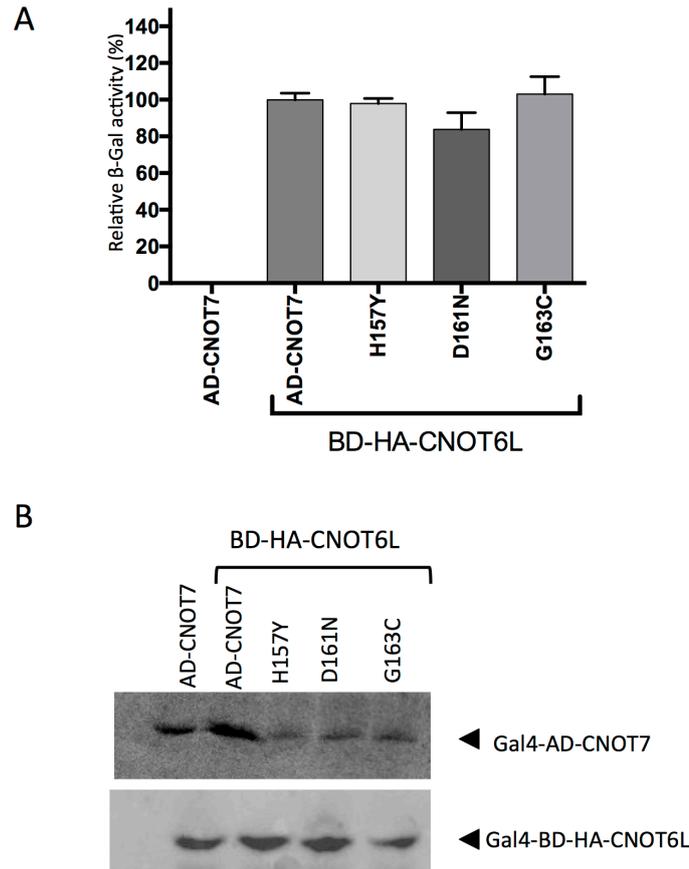
Group II, composed of H157Y, D161N, and G163C did not show any reduction in the  $\beta$ -galactosidase expression (Figure 5.13 A). Expression of BD-HA-CNOT6L and the CNOT7 mutants was confirmed by western blot of yeast lysates (Figure 5.13 B).

Of the six CNOT7 variants of Group III, N19I and G45C showed a significant reduced expression of the  $\beta$ -galactosidase and hence were unable to interact with BD-CNOT6L (Figure 5.14 A). Expression of BD-HA-CNOT6L and CNOT7 variants was confirmed by western blot of yeast cellular lysates though the variants N19S and N19I resulted to have a lower expression levels as observed before (Figure 5.14 B).

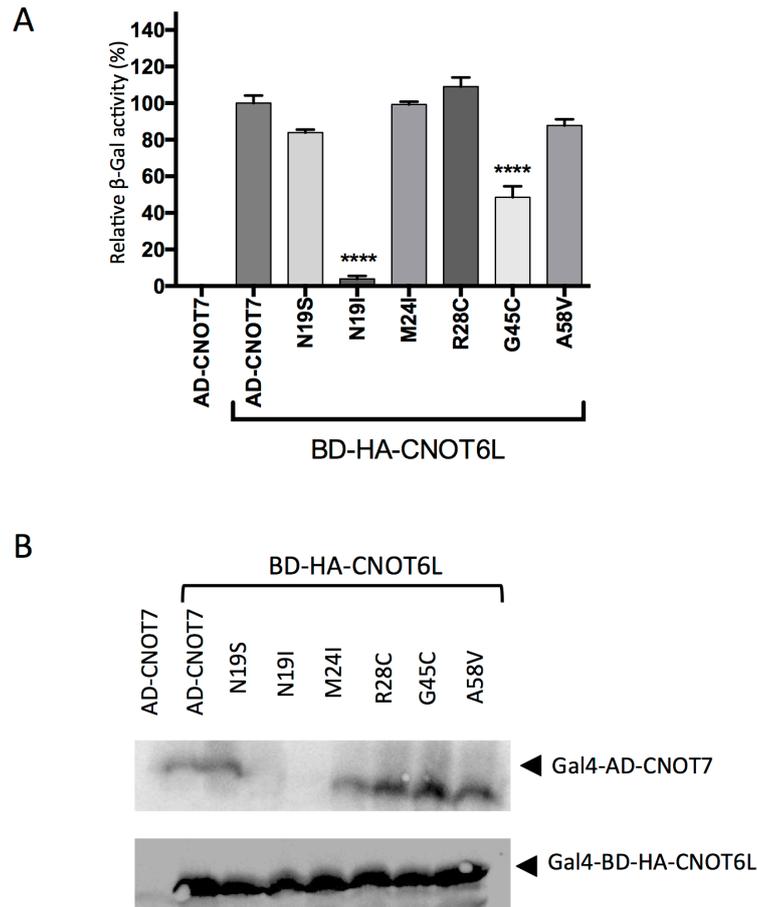
In the remaining Group IV, no variant showed decreased CNOT7-CNOT6L interactions. Only the mutation E217Q displayed a slightly increased  $\beta$ -galactosidase expression (Figure 5.15 A). Cellular lysates were analysed by western blot to confirm the expression of BD-HA-CNOT6L and the CNOT7 mutants (Figure 5.15 B).



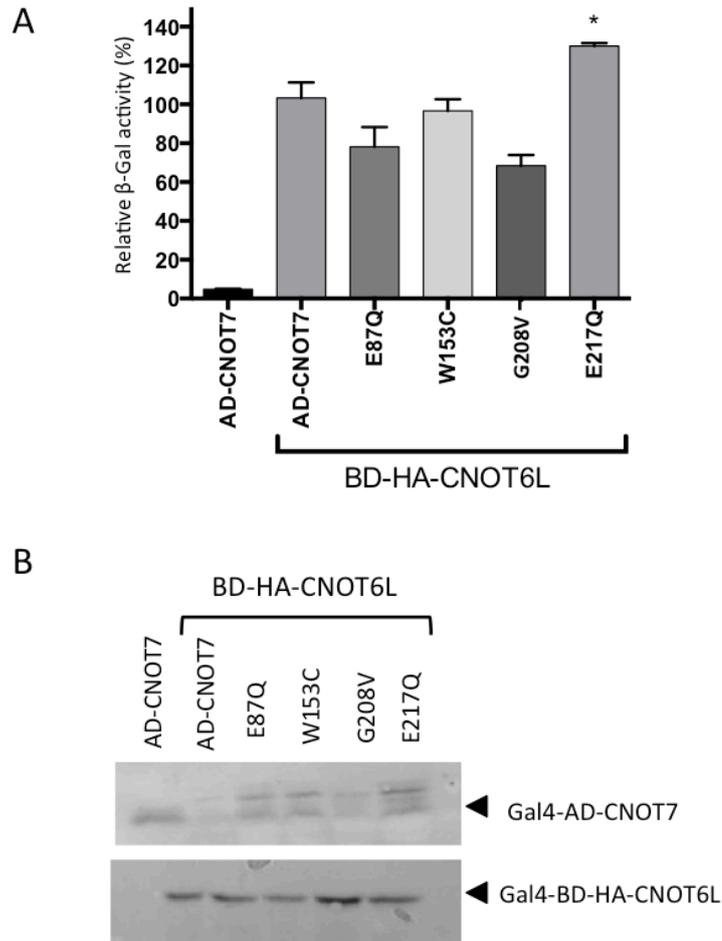
**Figure 5.12 Yeast two-hybrid analysis of the CNOT6L interaction with CNOT7 variants (Group I).** (A) YRG2 yeast cells were transformed with the indicated vectors; pAD-Gal4-CNOT7 and its variants, and pBD-Gal4-HA or pBD-Gal4-HA-CNOT6L.  $\beta$ -Galactosidase activity was normalised against total protein concentration. Error bars represent the standard error of the mean ( $n=3$ ) and the P value is indicated (\*\*  $\leq 0.01$ ). (B) Western blot analysis showing the expression of Gal4-BD-HA-CNOT1<sub>MIF4G</sub> and Gal4-AD-CNOT7. Anti-HA antibodies were used to detect Gal4-BD-HA-CNOT1<sub>MIF4G</sub>; anti-Gal4-AD antibodies were used to detect Gal4-AD-CNOT7.



**Figure 5.13 Yeast two-hybrid analysis of the CNOT6L interaction with CNOT7 variants (Group II).** **(A)** YRG2 yeast cells were transformed with the indicated vectors; pAD-Gal4-CNOT7 and its variants, and pBD-Gal4-HA or pBD-Gal4-HA-CNOT6L.  $\beta$ -Galactosidase activity was normalised against total protein concentration. Error bars represent the standard error of the mean (n=3). **(B)** Western blot analysis showing the expression of Gal4-BD-HA-CNOT6L and Gal4-AD-CNOT7. Anti-HA antibodies were used to detect Gal4-BD-HA-CNOT6L; anti-Gal4-AD antibodies were used to detect Gal4-AD-CNOT7.



**Figure 5.14 Yeast two-hybrid analysis of the CNOT6L interaction with CNOT7 variants (Group III).** **(A)** YRG2 yeast cells were transformed with the indicated vectors; pAD-Gal4-CNOT7 and its variants, and pBD-Gal4-HA or pBD-Gal4-HA-CNOT6L.  $\beta$ -Galactosidase activity was normalised against total protein concentration. Error bars represent the standard error of the mean ( $n=3$ ) and the P value is indicated ( $**** \leq 0.0001$ ). **(B)** Western blot analysis showing the expression of Gal4-BD-HA-CNOT6L and Gal4-AD-CNOT7. Anti-HA antibodies were used to detect Gal4-BD-HA-CNOT6L; anti-Gal4-AD antibodies were used to detect Gal4-AD-CNOT7.



**Figure 5.15 Yeast two-hybrid analysis of the CNOT6L interaction with CNOT7 variants (Group IV).** (A) YRG2 yeast cells were transformed with the indicated vectors; pAD-Gal4-CNOT7 and its variants, and pBD-Gal4-HA or pBD-Gal4-HA-CNOT6L.  $\beta$ -Galactosidase activity was normalised against total protein concentration. Error bars represent the standard error of the mean (n=3) and the P value is indicated (\*  $\leq 0.05$ ). (B) Western blot analysis showing the expression of Gal4-BD-HA-CNOT1<sub>MIF4G</sub> and Gal4-AD-CNOT7. Anti-HA antibodies were used to detect Gal4-BD-HA-CNOT1<sub>MIF4G</sub>; anti-Gal4-AD antibodies were used to detect Gal4-AD-CNOT7.

#### **5.4.4 Identification of CNOT7 residues required for the interaction with BTG1**

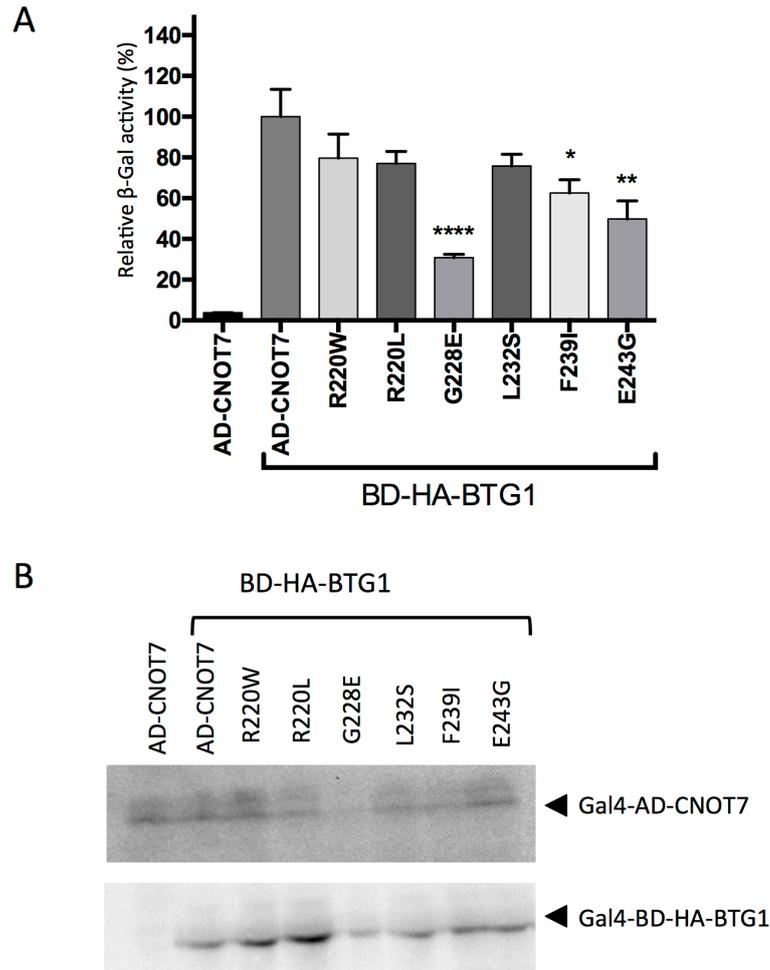
Finally, the impact of the CNOT7 mutations on the interaction with BTG1 was investigated.

Of the six mutations analysed in the Group I, G228E, F239I, and E243G resulted in a reduced  $\beta$ -galactosidase activity and thus were not able to interact with BTG1 (Figure 6.15B A). Cellular lysates were analysed by western blot to confirm the expression of BD-HA- BTG1 and the CNOT7 mutants (Figure 6.15 B).

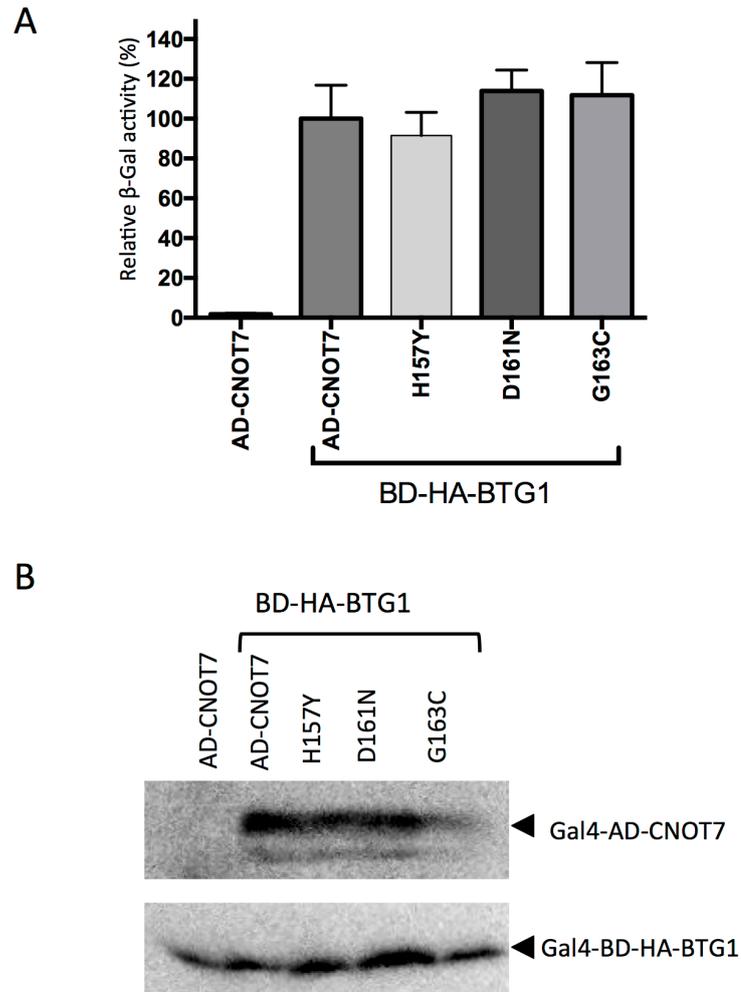
In Group II, no mutation showed any reduction in the  $\beta$ -galactosidase activity (Figure 5.17 A). The expression of BD-HA-CNOT6L and the CNOT7 variants was confirmed by western blot of yeast lysates (Figure 5.17 B).

Of the six CNOT7 variants of Group III, N19S and N19I showed a significant reduced expression of the  $\beta$ -galactosidase and hence were unable to interact with BD-BTG1 (Figure 5.18 A). Expression of BD-HA-BTG1 and CNOT7 variants was confirmed by western blot of yeast cellular lysates. As observed before, the two variants N19S and N19I showed lower expression levels (Figure 5.18 B).

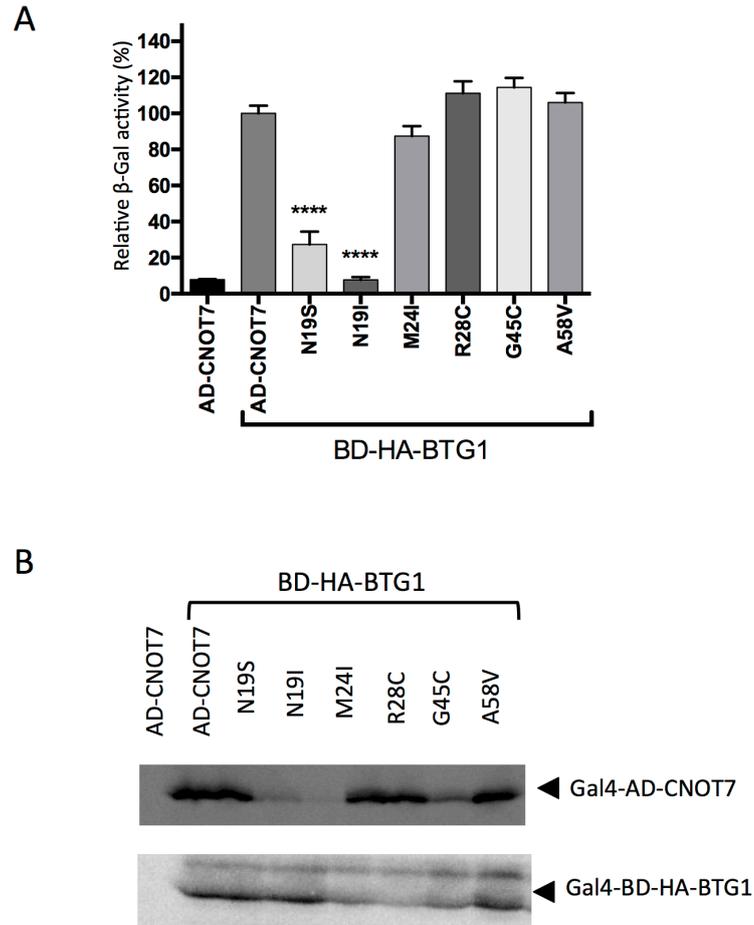
In the remaining Group IV, amino acid substitutions E87Q, G208V, and E217Q interfered with the CNOT7-CNOT6L interactions. (Figure 5.19 A). Cellular lysates were analysed by western blot to confirm the expression of BD-HA-CNOT6L and the CNOT7 mutants (Figure 5.19 B).



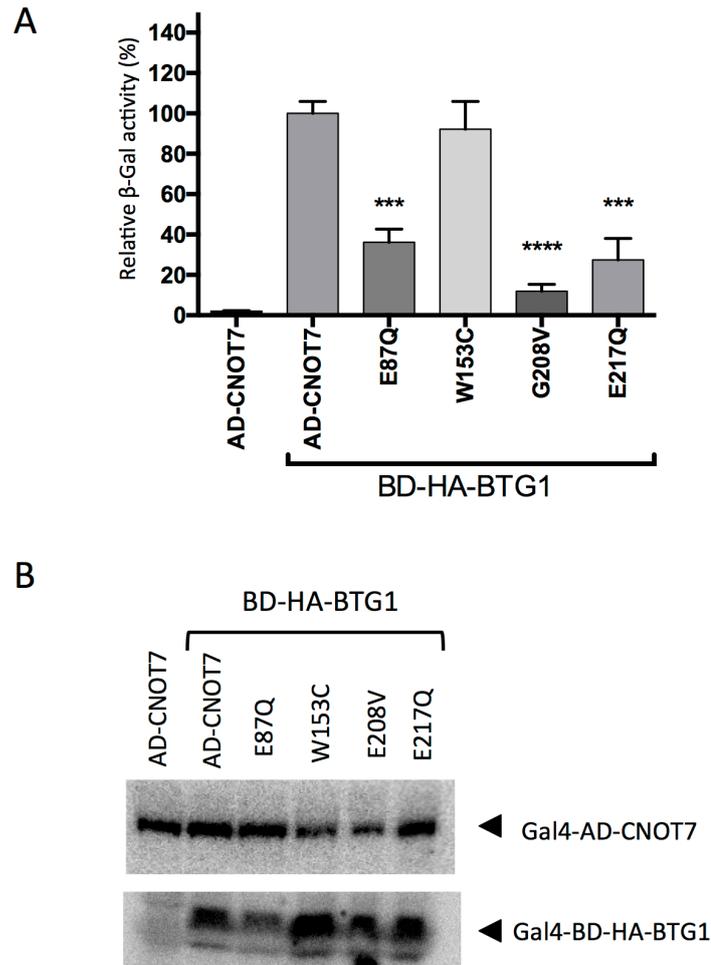
**Figure 5.16** Yeast two-hybrid analysis of the BTG1 interaction with CNOT7 variants (Group I). **(A)** YRG2 yeast cells were transformed with the indicated vectors; pAD-Gal4-CNOT7 and its variants, and pBD-Gal4-HA or pBD-Gal4-HA-BTG1  $\beta$ -Galactosidase activity was normalised against total protein concentration. Error bars represent the standard error of the mean (n=3) and the P value is indicated (\*  $\leq$  0.05, \*\*  $\leq$  0.01, \*\*\*\*  $\leq$  0.0001). **(B)** Western blot analysis showing the expression of Gal4-BD-HA-BTG1 and Gal4-AD-CNOT7. Anti-HA antibodies were used to detect Gal4-BD-HA-BTG1; anti-Gal4-AD antibodies were used to detect Gal4-AD-CNOT7.



**Figure 5.17 Yeast two-hybrid analysis of the BTG1 interaction with CNOT7 variants (Group II).** **(A)** YRG2 yeast cells were transformed with the indicated vectors; pAD-Gal4-CNOT7 and its variants, and pBD-Gal4-HA or pBD-Gal4-HA-BTG1  $\beta$ -Galactosidase activity was normalised against total protein concentration. Error bars represent the standard error (n=3). **(B)** Western blot analysis showing the expression of Gal4-BD-HA-BTG1 and Gal4-AD-CNOT7. Anti-HA antibodies were used to detect Gal4-BD-HA-BTG1; anti-Gal4-AD antibodies were used to detect Gal4-AD-CNOT7.



**Figure 5.18 Yeast two-hybrid analysis of the BTG1 interaction with CNOT7 variants (Group III).** **(A)** YRG2 yeast cells were transformed with the indicated vectors; pAD-Gal4-CNOT7 and its variants, and pBD-Gal4-HA or pBD-Gal4-HA-BTG1  $\beta$ -Galactosidase activity was normalised against total protein concentration. Error bars represent the standard error of the mean (n=3) and the P value is indicated (\*\*\*\*  $\leq 0.0001$ ). **(B)** Western blot analysis showing the expression of Gal4-BD-HA-BTG1 and Gal4-AD-CNOT7. Anti-HA antibodies were used to detect Gal4-BD-HA-BTG1; anti-Gal4-AD antibodies were used to detect Gal4-AD-CNOT7.



**Figure 5.19 Yeast two-hybrid analysis of the BTG1 interaction with CNOT7 variants (Group IV).** (A) YRG2 yeast cells were transformed with the indicated vectors; pAD-Gal4-CNOT7 and its variants, and pBD-Gal4-HA or pBD-Gal4-HA-BTG1  $\beta$ -Galactosidase activity was normalised against total protein concentration. Error bars represent the standard error of the mean (n=3) and the P value is indicated (\*\* $\leq 0.001$ , \*\*\*\* $\leq 0.0001$ ). (B) Western blot analysis showing the expression of Gal4-BD-HA-BTG1 and Gal4-AD-CNOT7. Anti-HA antibodies were used to detect Gal4-BD-HA-BTG1; anti-Gal4-AD antibodies were used to detect Gal4-AD-CNOT7.

#### 5.4.5 Identification of residues required for the activity of CNOT7

To assess the function of CNOT7 variants in mRNA degradation, RNA tethering assays were carried out. In order to perform the RNA tethering assay, plasmid pCI $\lambda$ N-CNOT7 and 19 variants were generated. CNOT7 and the corresponding mutant cDNAs were subcloned into the pCI $\lambda$ N by Gibson assembly using the pAD-GAL4-CNOT7 plasmids as a template (Figure 5.20). The sequences of the plasmid series were confirmed by restriction enzyme digestions and DNA sequencing.

HEK293 cells were co-transfected with the reporter pRL5boxB, expressing a luciferase reporter mRNA containing five boxB sequences in the 3' UTR, and the plasmid expressing the  $\lambda$ N fusion proteins. After transfection, total protein was extracted and luciferase activity determined. In parallel, protein expression was evaluated by western blot analysis.

In addition to the empty vector pCI $\lambda$ N, the D40A variant (Doidge et al., 2012) as used as a control. The Asp40 residue is required for binding of Mg<sup>2+</sup> ions and this variant is catalytically inactive.

All variants in Group I, composed of R220W, R220L, G228E, L232S, F239I and E243G, did not show any significant alteration in the function of the protein (Figure 5.21 A). HEK293 cellular lysates were analysed in order to confirm the expression of the proteins. The variants R220W, G228E, and L232S, showed a reduced level of expression but their function was not compromised (Figure 5.21 B).

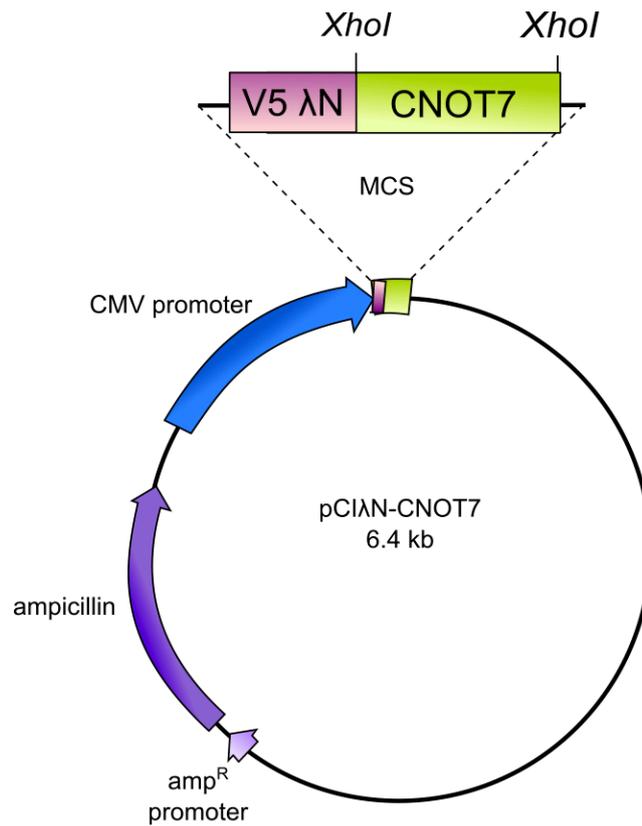
Group II, composed of H157Y, D161N and G163C, showed reduced activity only in case of the G163C variant (Figure 5.22 A). To confirm protein expression, cellular lysates were analysed by western blot. The G163C mutations though showed a reduced level of expression (Figure 5.22 B).

In Group III, no variant displayed reduced activity (Figure 5.23 A). Cellular lysates were analysed by western blot to confirm the protein expression. The mutation N19S and N19I resulted to have a reduced expression level, in agreements with the yeast two hybrids experiments findings (Figure 5.23 B).

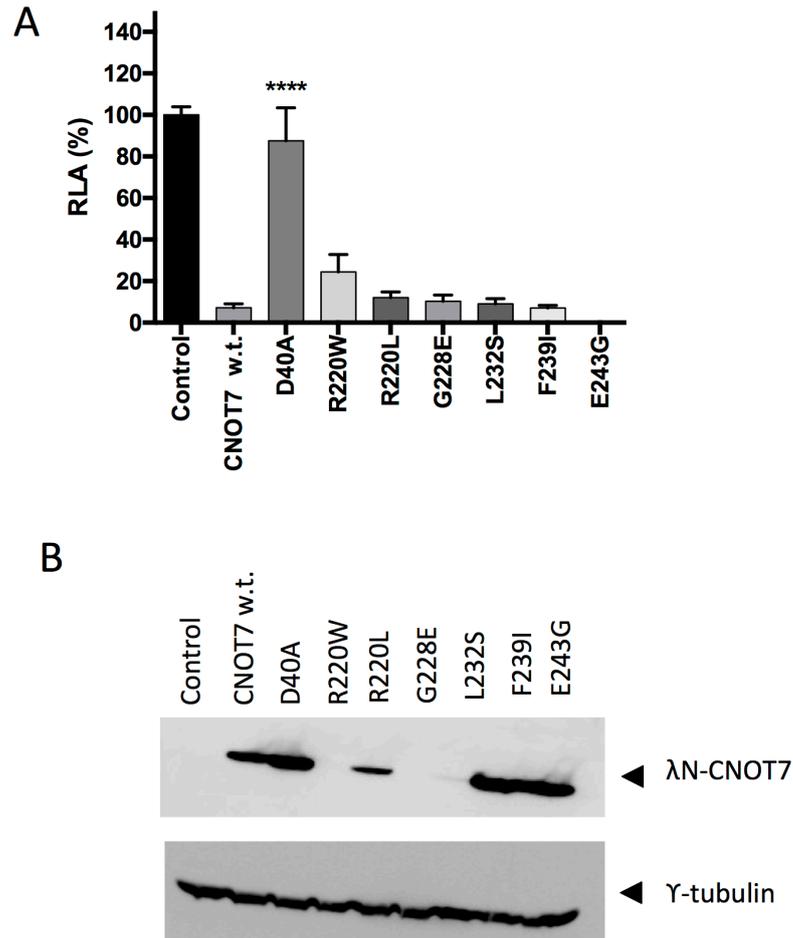
This was also the case for Group IV, which did not contain any variant whose function was compromised (Figure 5.24 A). The western blot of the

HEK293 cellular lysates confirmed protein expression although the variants E87Q and W153C appeared to have a lower expression level (Figure 5.24 B).

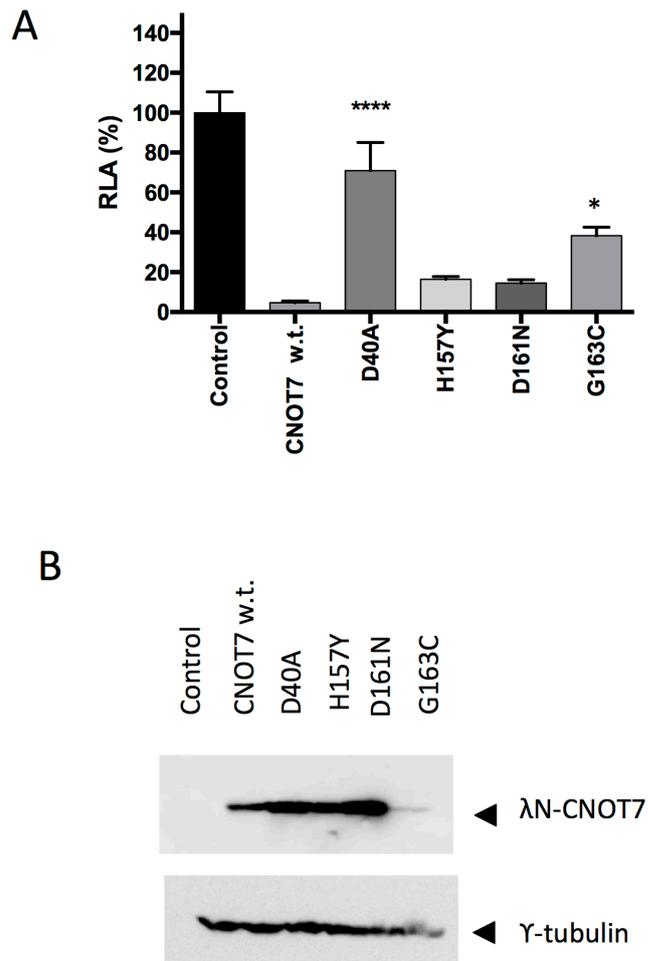
In conclusion, the RNA tethering assay identified only a single CNOT7 variant, G163C, with partially reduced activity.



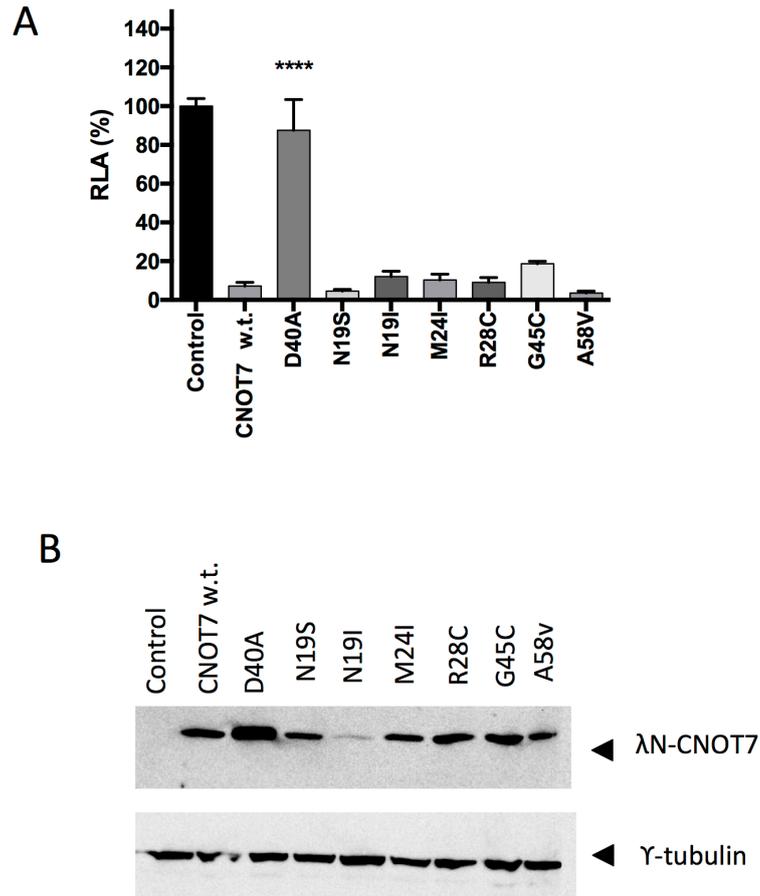
**Figure 5.20 Generation of the λN-CNOT7 expression plasmids for RNA tethering assays.** Expression of V5-tagged λN-CNOT7 fusion proteins is driven by a CMV promoter. CNOT7 variants were subcloned into the XhoI restriction site of the multiple cloning sequence (MCS) using Gibson assembly.



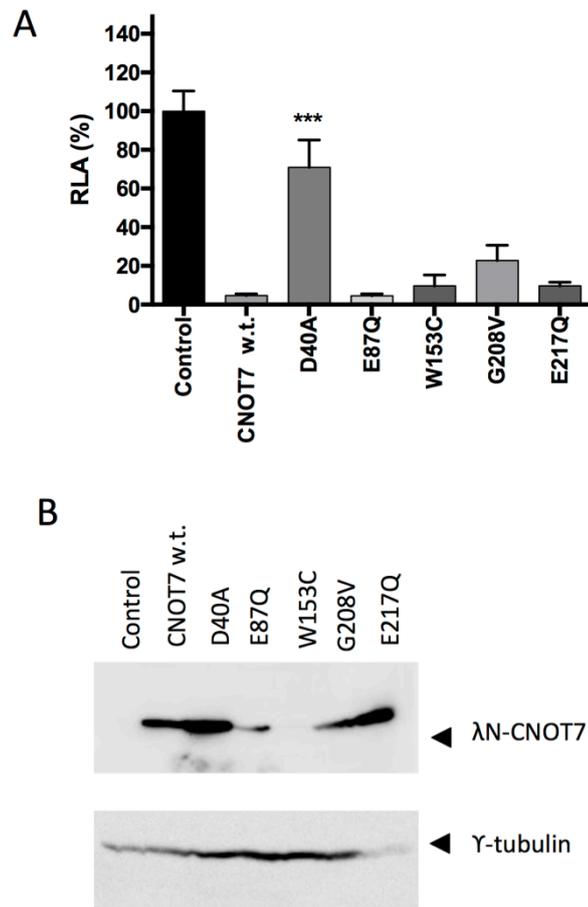
**Figure 5.21 Group I CNOT7 variants are functional in RNA degradation.** **(A)** Luciferase assay. HEK293 cells were transfected with pCIλN-CNOT7 plasmids or empty vector. The catalytically reduced variant CNOT7 D40A was included as a control. Error bars represent the standard error of the mean (n=3) and the P value is indicated (\*\*\*) ( $\leq 0.001$ ). **(B)** Cellular lysates were used to confirm protein expression. Antibody anti-V5 was used to detect λN-CNOT7 and its variants. Antibody anti-γ-tubulin was used as a loading control.



**Figure 5.22 CNOT7 variant G163C displays reduced activity in RNA degradation.** **(A)** Luciferase assay. HEK293 cells were transfected with pCIλN-CNOT7 plasmids or empty vector. The catalytically reduced variant CNOT7 D40A was included as a control. Error bars represent the standard error of the mean (n=3) and the P value is indicated (\* $\leq$  0.05 and \*\*\*\* $\leq$  0.001). **(B)** Cellular lysates were used to confirm protein expression. Antibody anti-V5 was used to detect λN-CNOT7 and its variants. Antibody anti-γ-tubulin was used as a loading control.



**Figure 5.23 Group III CNOT7 variants are functional in RNA degradation.** **(A)** Luciferase assay. HEK293 cells were transfected with pCIλN-CNOT7 plasmids or empty vector. The catalytically reduced variant CNOT7 D40A was included as a control. Error bars represent the standard error of the mean ( $n=3$ ) and the P value is indicated ( $****\leq 0.0001$ ). **(B)** Cellular lysates were used to confirm protein expression. Antibody anti-V5 was used to detect λN-CNOT7 and its variants. Antibody anti-γ-tubulin was used as a loading control.



**Figure 5.24 Group IV CNOT7 variants are functional in RNA degradation.** **(A)** Luciferase assay. HEK293 cells were transfected with pCI $\lambda$ N-CNOT7 plasmids or empty vector. The catalytically reduced variant CNOT7 D40A was included as a control. Error bars represent the standard error of the mean (n=3) and the P value is indicated (\*\*\*) ( $\leq 0.001$ ). **(B)** Cellular lysates were used to confirm protein expression. Antibody anti-V5 was used to detect  $\lambda$ N-CNOT7 and its variants. Antibody anti- $\gamma$ -tubulin was used as a loading control.

## 5.5 Discussion

In this chapter, *in silico* and *in vitro* approaches were combined to study non-synonymous mutations found in the CNOT7 ribonuclease in different cancer tissues.

Initially, a list of somatic mutations was retrieved from the COSMIC (Catalogue Of Somatic Mutations In Cancer) database and the Ensemble genome browser (Bamford et al., 2004; Forbes et al., 2008, 2011; Flicek et al., 2014; Chen et al., 2010; Rios et al., 2010). Using this approach, 46 non-synonymous point mutations from re-sequencing data, literature, and the Cancer Genome Project at Sanger Institute were identified. Most of the mutations were found in lung (eleven) and endometrium (seven) cancer tissues (Figure 5.4)

### Prediction of mutation affecting protein function

The effect of the mutations on the protein function was predicted by the Sorting Intolerant From Tolerant (SIFT) algorithm (Ng and Henikoff, 2001, 2002, 2003; Kumar et al., 2009). These analyses predicted seventeen mutations to affect the protein function (Table 5.1). Four additional mutations predicted to be tolerated were included as control of the prediction method sensitivity.

In addition to SIFT, SuSPect was also used to predict damaging mutations (Yates et al., 2014). Discrepancies between the two software predictions were found. Out of the 21 mutation selected from the SIFT prediction, the variant M24I was predicted to be tolerated by SIFT but resulted to be damaging and disease-associated by SuSPect. On the contrary, the mutations H157Y, G163C, R220W, R220L, F239I, and E243G were predicted to be damaging by SIFT but not disease-associated by SusPect. Thus, SIFT may be more sensitive in identifying damaging mutations.

To obtain more detailed insight into possible defects caused by the amino acid substitutions, two approaches were used.

First, the possible effects on protein stability were investigated. Secondly, the structure of the CNOT7 ribonuclease was investigated in detail to evaluate whether the amino acid substitutions could be accommodated. To investigate the stability of the protein following amino acid substitution, Site Directed Mutator was used (Worth et al., 2011). The mutations were ranked in highly destabilising (two), destabilising (four), neutral (five), slightly stabilising (eight) and highly stabilising (one) (Table 5.2). Interestingly the destabilising mutations L232S, E243G, and L79Q and the highly stabilising mutation G163C were predicted to contribute to the malfunction of the protein.

Furthermore changes in the overall structure of the protein may affect not only its function but also its interaction with other proteins. Studies revealed CNOT7 to contact several proteins. The ribonuclease subunit was shown to be anchored on the MIF4G domain of the central scaffold of the complex CNOT1 (Bai et al., 1999; Basquin et al., 2012; Petit et al., 2012). Moreover, CNOT7 interacts with the LRR domain of the second ribonuclease subunit of the Ccr4-Not complex CNOT6L (Malvar et al., 1992; Dupressoir et al., 2001; Basquin et al., 2012; Clark et al., 2004). Studies assessed the interaction of BTG/Tob family proteins BTG1 and BTG2 with CNOT7 via their BTG domain. BTG1 and BTG2 are highly similar proteins and have a regulatory role on the shortening of the poly(A) tail (Mauxion et al., 2008; Winkler, 2010; Xu et al., 2014; Doidge et al., 2012; Stupfler et al., 2016). In this context, accommodation of the mutations in the CNOT7 structure was investigated to identify mutations causing perturbation in the structure of the protein or in the interactions with the MIF4G domain of CNOT1, with CNOT6L, and with BTG1. For this purpose, UCSF Chimera was used to study the accommodation of amino acid substitutions into the CNOT7 model (Pettersen et al., 2004). In this analysis every residue rotamer possible with probability >5% was inspected and accommodation into the structure was evaluated. Although six amino acid substitutions were found to interfere with the protein structure and could not be accommodated, none of them directly affected the interactions with the other components of the sub-complex. G45C

substitution could not be accommodated within the structure in 93.5% of the cases, R220W with the probability 81.1%, D161N and G163C with 33%, G228E with 24.1%, and H157Y with 8.7% (Figure 5.6).

#### Validation of protein-protein interactions and function in RNA degradation

To investigate the effects of the mutations further, two approaches were used. The impact of the mutations on the interactions with CNOT1, CNOT6L, and BTG1 was investigated by yeast two hybrid assays, whereas the impact on the function of the protein was inspected by RNA tethering assays.

The yeast two hybrid assays identified CNOT7 residues required for the interactions with the other proteins. Eight variants N19S, N19I, W153C, G208V, R220W, L232S, F239I, and E243G were found to affect the interaction with the MIF4G domain of CNOT1 (Figure 5.8-5.11). Three variants N19I, G45C, and G228E were found to affect the interaction with CNOT6L whereas E217Q showed an increased affinity for the ribonuclease protein (Figure 5.12-5.15). The study of CNOT7 variants identified eight mutations N19S, N19I, E87Q, G228E, F239I, E243G, G208V, and E217Q to be critical for the interaction with BTG1 (Figure 5.16-5.20). Interestingly, the mutation N19I was found to disrupt the interaction with all three studied partners CNOT1, CNOT6L, and BTG1. Moreover, the two variants E87Q and E217Q that were predicted to be tolerated were unable to interact with BTG1.

The impact of the mutations on the function of CNOT7 was investigated by RNA tethering assays. The amino acid substitution G163C was the only variant with reduced activity (Figure 5. 21- 5. 24).

#### Combining predictive and experimental approaches

This study combined *in silico* and *in vitro* analyses to explore the effect of mutations found in cancer on the function, stability, structure, and interactions of the CNOT7 ribonuclease subunit. The profile of the CNOT7 variants obtained from these approaches revealed residues specific for the

interaction with specific proteins and residues whose role involved the interaction with more than one partner.

Previous studies in *S. cerevisiae* showed that double substitution of Caf1 A215E and F219E, equivalent to the human CNOT7 residues C67 and L71, abolished the interaction with Ccr4 and reduced the affinity for Not1 (Basquin et al., 2012). Similarly, substitution of Asp19 (N19S and N19I) indicated that this residue is involved in the interaction with CNOT1<sub>MIF4G</sub>, CNOT6L and BTG1. Interestingly, the impact of the two different substitutions was different. The serine substitution impacted only the interaction with CNOT1 and BTG1, although was predicted to be destabilising by Site Directed Mutator (SDM). The substitution with isoleucine, found in lung cancer, was found to impact on the interaction with all three proteins whilst the stability of the protein was predicted to be not affected. Moreover structural analysis did not show any clashes with other amino acids and no functional role was found for this residue. Interestingly, these variants showed a reduced level of expression in yeast and HEK293 cells.

CNOT7 variants found in lung cancer, G208V and F239I, and E243G, found in central nervous system cancer, affected the interaction with both CNOT1 and BTG1. The variants F293I and E243G were predicted to be damaging for the protein function by SIFT but not disease-associated according to SuSpect. Structural analysis did not indicate any clashes with other residues. The substitution of the Glu243 with a glycine was predicted to be highly destabilising causing malfunction of the protein according to SDM.

CNOT7 variant G228E found in large intestine cancer showed a loss of affinity for both CNOT6L and BTG1. The substitution of Gly228 with glutamic acid rotamers present in 18.1% of the cases created clashes with Gln96 and Ile12. Interestingly the E217Q variant predicted to be tolerated was demonstrated to be unable to interact with BTG1 but to increase the affinity for CNOT6L. The mutation was predicted to be slightly stabilising. CNOT7 variants R220W, L232S, and W153C only affected the interaction with the MIF4G domain of CNOT7.

Double mutants of Caf1 M290K and M296K, equivalent to the human CNOT7 residues M141 and L147, were shown to disrupt the interaction with Not1 but not with Ccr4 (Basquin et al., 2012).

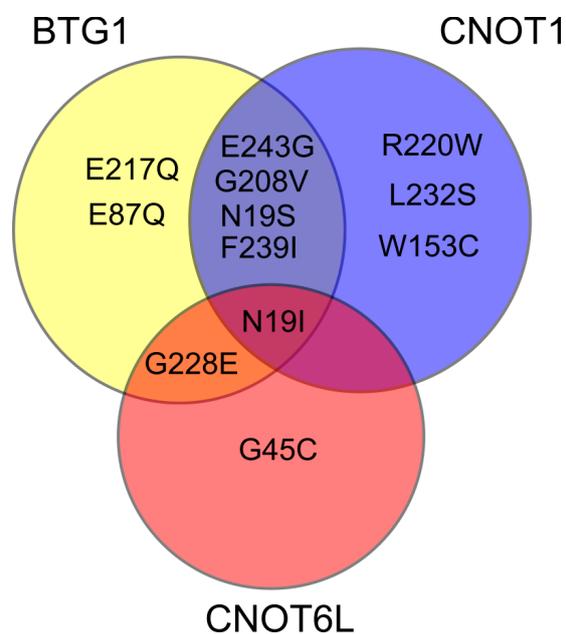
Similarly, the mutation L232S found in lung cancer tissue was predicted to be highly destabilising, W153C destabilising whereas R220W, found in lung cancer, was predicted to be slightly stabilising.

Moreover, substitution of the Arg220 with tryptophan could not be accommodated in the structure without major changes with rotamers present in the 71% of the cases and clashes with Gln210, Gln224, and Gly222 were found.

Gly45 of CNOT7 was found to be specifically required for the interaction with CNOT6L. Substitution of this residue with cysteine negatively affected the affinity between the two ribonucleases. The substitution was predicted to be destabilising for the protein and structural analysis showed that in the 93.5% of the cases, cysteine rotamers could not be accommodated in the structure without major structural changes. Although no correlation with malfunction of the protein was predicted, E87Q, together with E217Q, was found to affect the interaction with BTG1. These findings are summarised in Figure 5.25.

Importantly, experimental studies did not show any correlation between the analysis of protein-protein interactions using yeast-two hybrid assays functional analysis using the RNA tethering assays. The twelve mutations impacting the interaction with other components of the nuclease module did not affect the deadenylase activity of the protein *in vitro*. Although the CNOT7 deadenylase activity is not compromised, the reduced affinity of CNOT7 mutants for CNOT1<sub>MIF4G</sub>, CNOT6L or BTG1 may results in their incapacity to be recruited on the complex and impact the mRNA degradation *in vivo*. The only mutation that was found to have an intrinsic reduced activity was G163C. This variant was predicted to be highly stabilising resulting in malfunction and disease-association. The substitution of Gly163 with cysteine could not be accommodated in the 33% of the cysteine rotamers clashing with the Leu174.

This study combined *in silico* and *in vitro* analyses to explore the effect of mutations found in cancer on the function, stability, structure, and interactions of the CNOT7 ribonuclease subunit. Residues required for the interaction with CNOT1, CNOT6L, and BTG1 were identified and characterised. From RNA tethering assays, the variant G163C was found to be the only analysed mutation to have a functional impact on the CNOT7 deadenylase activity.



**Figure 5.25 Overview of the CNOT7 variants and their impact on the interaction with CNOT6L, CNOT1, and BTG1.** The Venn diagram summarises the outcome of the interaction studies using the yeast two hybrid assay. In the yellow circle, the variants are grouped that interfered with the interaction with BTG1; in the blue circle, the variants interfering with the interaction with CNOT1; in the red circle, the variants interfering with the interaction with CNOT6L. Overlaps of the circles highlight variants interfering with more two or more protein.

# **Chapter 6**

## **Phylogenetic and variability analyses of the Ccr4-Not deadenylase complex**

## **Chapter 6. Phylogenetic and variability analysis of the Ccr4-Not complex**

### **6.1 Introduction**

According to modern evolutionary theory, one molecule undergoes diversification into many variations and one or more of those variants can be selected to be reproduced or amplified throughout a population over many generations. Such variations at the molecular level can be caused by mutations such as deletions, insertions, inversions, or substitutions at the nucleotide level, which in turn affect protein structure and biological function (Hahn, 2008). Phylogenesis tries to recover the order of evolutionary events and represent them in an evolutionary tree that graphically describes the relationship among species or genes.

Phylogenetic reconstruction is a problem of statistical inference and involves the use of models of evolution and assumptions about the process of nucleotide or aminoacid substitution (Lemey et al., 2009) . In this study, the phylogenetic reconstruction was performed using two different approaches and subsequent comparison. The first approach used was the Neighbour-Joining (NJ) (Saitou N, 1987). This method operates on a matrix of pairwise distances between taxa with a minimum-evolution criterion selecting the tree with the shortest branch length as optimal. The second approach used was the Maximum Likelihood (ML) discrete-character method that operates on a matrix of discrete characters assigning one or more attributes or character states to each taxon. This method infers probabilities to all the possible trees and selects the one with the highest probability as optimal.

Furthermore I performed variability analysis in order to detect selective pressure acting on the proteins. Adaptation by natural selection is the most important process in biology. Mutations are the source of variability and are subjected to selective pressures. When a mutation has no effect on the protein function it may be fixed in the population without altering its fitness (neutral selective pressure). When the mutation has damaging effects on the protein, hence on the organism, it is subjected to negative

selective pressure because it cannot be fixed in the population reducing the fitness of the organisms carrying it. A mutation that is convenient and increases the fitness of an organism is fixed in the population and is likely to be linked to acquiring of new function of the protein (positive selective pressures) (Lemey et al., 2009). To detect selective pressure, we used two kinds of models: the site and branch-site models. The site model allows the parameter describing the selective pressure ( $\omega$ ) to vary among sites (codons) (Nielsen and Yang 1998; Yang et al 200b). The branch-site model allows  $\omega$  to vary both among sites in the protein and across branches of the tree with the intent of detecting positive selection acting on few sites along particular lineages (Yang et al., 2005; Zhang et al., 2005).

In this study we reconstructed the phylogeny of each subunit of the Ccr4-Not complex and inspected their sequences in order to detect positive selection acting among the 15 mammalian species listed in Table 6.1.

<b>Species</b>	<b>Common name</b>	<b>Taxonomy ID</b>
<i>Pan troglodytes</i>	Chimpanzee	9598
<i>Homo sapiens</i>	Human	9606
<i>Gorilla gorilla gorilla</i>	Western lowland gorilla	9595
<i>Nomascus leucogenys</i>	Northern white-cheeked gibbon	61853
<i>Macaca mulatta</i>	Rhesus macaque	9544
<i>Papio anubis</i>	Olive baboon	9555
<i>Oryctolagus cuniculus</i>	European Rabbit	9986
<i>Bos taurus</i>	Cattle	9913
<i>Sus scrofa</i>	Pig	9823
<i>Equus caballus</i>	Horse	9796
<i>Felis catus</i>	Cat	9685
<i>Rattus norvegicus</i>	Rat	10116
<i>Mus musculus</i>	Mouse	10090
<i>Canis lupus familiaris</i>	Dog	9615
<i>Pongo abelii</i>	Sumatran orangutan	9601

**Table 6.1 Species analysed.** The table reports the latin name, the common name and the taxonomy identifier of the 15 analysed mammal species..

## 6.2 Phylogenetic and variability analysis of CNOT1

### 6.2.1 Phylogenetic reconstruction of CNOT1

The CNOT1 homologous sequences were retrieved using the human CNOT1 protein as query by DELTA-BLAST (Boratyn et al., 2012) from the Reference protein database (refseq\_protein) and are listed in Supplementary Table S.I. 65.1.

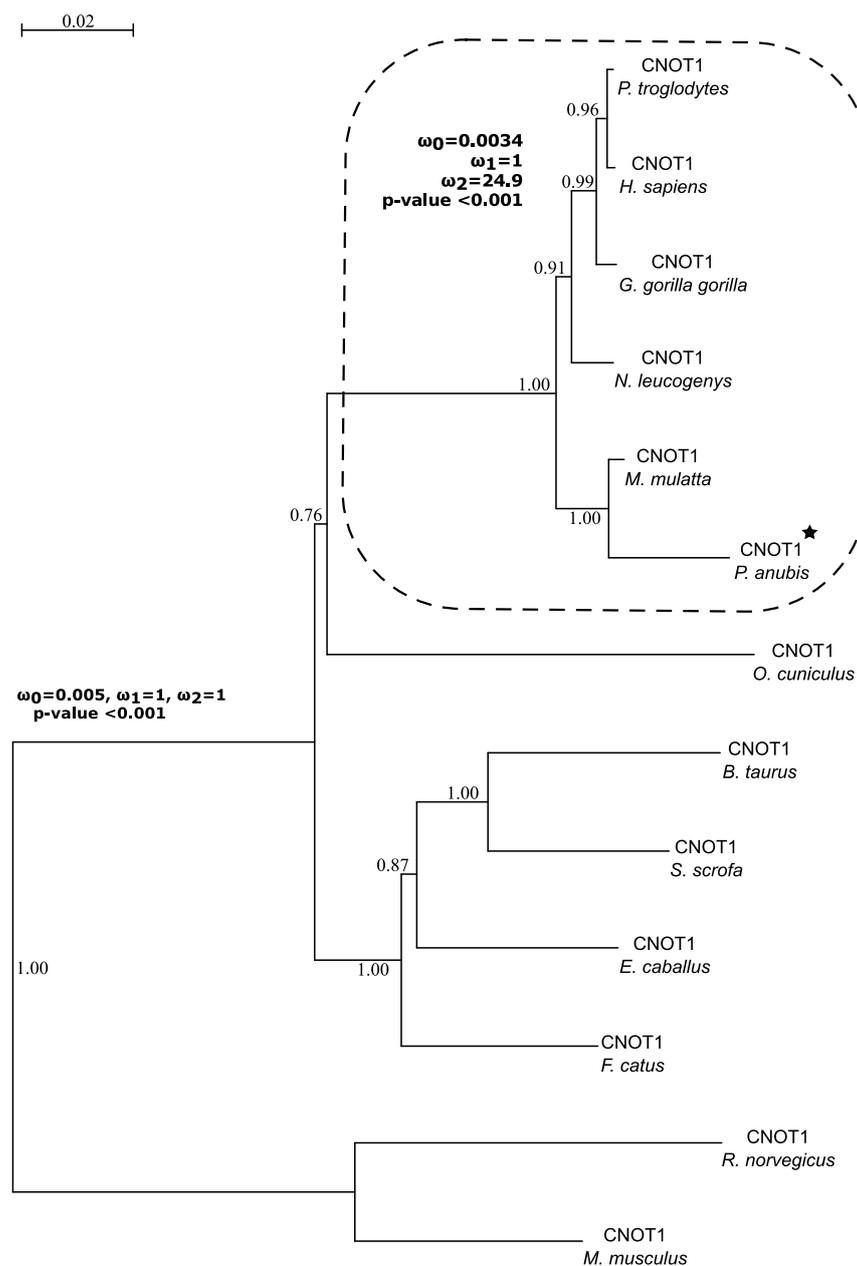
The *Canis lupus familiaris* and *Pongo abelii* homologues were discarded from the analyses because of their *low quality protein* annotation status. Homologous proteins were aligned by MAFFT v.7 (Katoh et al., 2002; Katoh and Standley, 2013) using the E-ins-I algorithm to preserve the domains alignments. The multi-aligned nucleotide sequences were then used to infer a phylogenetic tree with the Neighbour-Joining (NJ) method using MEGA v6.06 (Tamura et al., 2013) with a bootstrap value of 1000. The resulting tree, visualised with Seaview v.4.5.4 (Gouy et al., 2010), is shown in Supporting Figure 6.1. Branches were supported from strong bootstrap values allowing a robust tree topology.

In order to reconstruct a phylogenetic tree with the Maximum Likelihood method, nucleotide substitution model choice was evaluated using jModelTest (Posada, 2008; Darriba et al., 2015). The nucleotide substitution model chosen according to the Akaike Information Criterion (AIC) (Akaike, 1974) was the Tamura-Nei model (Tamura and Nei, 1993) (TrN+I+G, 10020) indicating that there was a great proportion of invariable sites (+I) and there was rate heterogeneity among sites (+G). The phylogenetic tree with the Maximum Likelihood method was constructed based on the nucleotide substitution model using phyML 3.0 (Guindon et al., 2010). Branches were tested using the non-parametric Shimodaira-Hasegawa-like (SH-like) (Shimodaira and Hasegawa, 1999) branch support test. The reconstructed Maximum Likelihood tree is very similar to the one obtained with the NJ method with no substantial differences and is shown in Figure 6.1.

### 6.2.2 Variability analysis of the CNOT1

Selective pressures acting on the CNOT1 proteins among the selected mammals were detected using the Phylogenetic Analysis by Maximum Likelihood (PAML) software v. 4.8a (Yang, 2007).

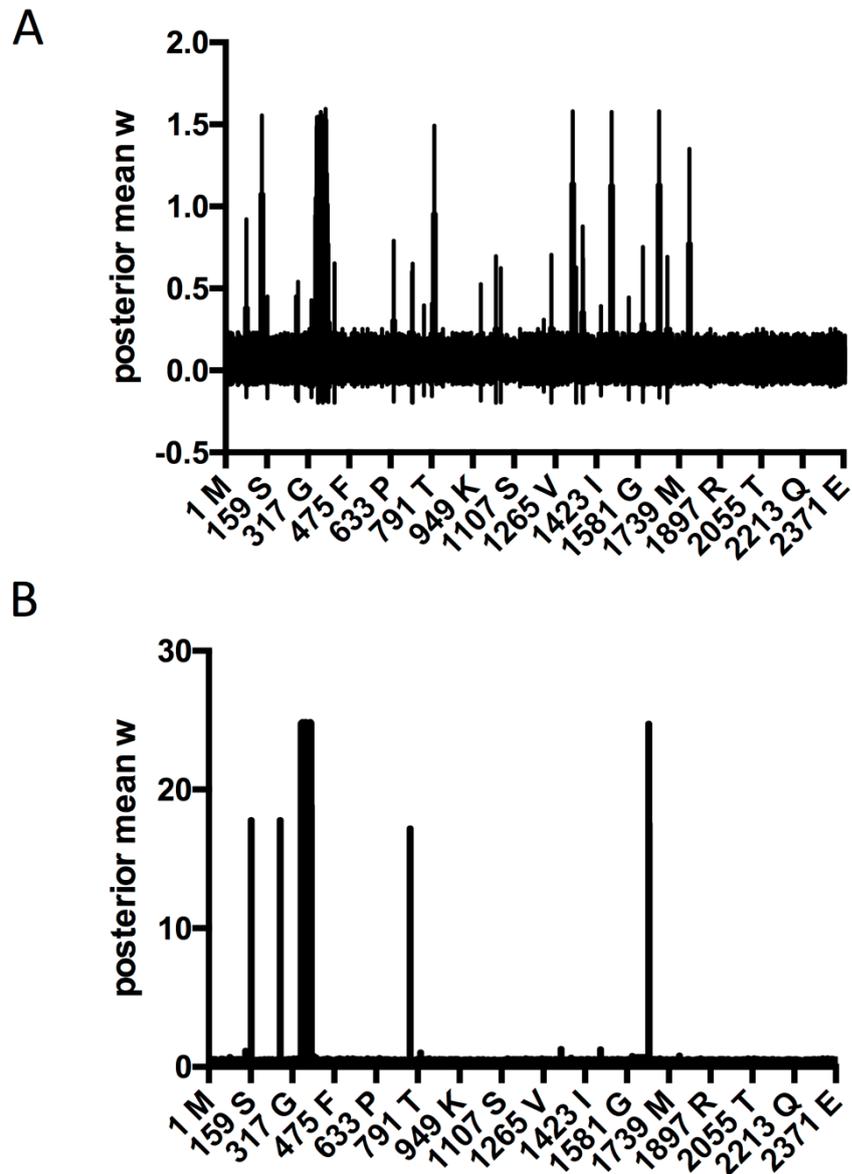
In order to detect if positive selective pressures were acting in the CNOT1 among all lineages, we performed a comparison between the positive selection model m2a and its null hypothesis model m1 (Nielsen and Yang, 1998; Yang et al., 2005). The null hypothesis was rejected according Likelihood Ratio Test (LRT) (Felsenstein, 1981; Huelsenbeck and Crandall, 2003; Huelsenbeck et al., 1997) with the P value <0.001 allowing us to infer positive selective pressures acting on the proteins, as shown in Table 6.2. Unfortunately, no sites were significantly selected by Bayes Empirical Bayes (BEB) analysis (Yang et al., 2005). I therefore try to detect selective pressures acting only on some of the branches of the tree (model mA) (Yang et al., 2005; Zhang et al., 2005). In particular I focused on the primates lineage. The branch-site model was applied and LRT calculations rejected the null hypothesis with a P value <0.001. Posterior mean  $\omega$  values were calculated for both positive selection models m2a and mA, and are shown in Figure 6.2 A and B respectively. The BEB analysis identified 18 sites positively selected within the primates lineage with a P value <0.05 shown in Table 6.3. The residues positively selected are in a hot-spot spanning from the amino acid 349 to the amino acid 386. These residues are located in the HEAT-repeat domain in the CNOT1 N-terminus. In particular *Papio anubis* resulted to be the species diverging significantly from the other primates for most of the sites, meaningful of positive selective pressures acting on the CNOT1 of this species.



**Figure 6.1 CNOT1 Maximum Likelihood phylogenetic tree.** The Maximum Likelihood (ML) tree was inferred using the Tamura-Nei model (Tamura and Nei, 1993) including the proportion of invariable sites parameter (I) and the shape parameter of the gamma distribution (G). Shimodaira-Hasegawa-like (SH-like) branch support values are reported. Primates lineage is included in the dashed line box and estimated  $\omega$  parameter for the branch-site model are reported. The species with more than three positively selected sites whose codons differ from other primates is labelled with a star.

Lineage	Model	lnL	LRT	P value	$\omega$ parameters
A	m1	-17608.95			
	m2a	-17564.83	88.23	< 0.001	$\omega_0=0.005$ , $\omega_1=1, \omega_2=1$
P	mA <sub>0</sub>	-17475.69			$\omega_0=0.00349, \omega_1=1$
	mA	-17430.44	90.49	< 0.001	$\omega_0=0.00349$ , $\omega_1=1, \omega_2=24.9$

**Table 6.2CNOT1 Likelihood Ratio Tests.** The table summarises the models tested for CNOT1. The two site models (m1 and m2a) were tested among the whole CNOT1 tree (A). The two branch-site models (mA and its null hypothesis mA<sub>0</sub>) were tested assuming positive selection in the primates lineage and neutral selective pressure for the remaining lineages (P). The log Likelihood, Likelihood Ratio Tests, P values and estimated  $\omega$  parameters are reported in the table for the models.



**Figure 6.2 CNOT1 posterior mean  $\omega$  graphs.** Posterior mean  $\omega$  values were calculated for each analysed site. **(A)** Posterior mean  $\omega$  for the m2a site model are shown. **(B)** Posterior mean  $\omega$  values calculated for the branch-site model applied to the primates lineage (mA) are represented.

Site	<i>H. sapiens</i>		<i>G. gorilla gorilla</i>		<i>N. leucogenys</i>		<i>P. troglodytes</i>		<i>P. anubis</i>		<i>M. mulatta</i>	
349	AAT	(N)	AAT	(N)	AAT	(N)	AAT	(N)	ATG	(M)	AAT	(N)
350	CCA	(P)	CCA	(P)	CCG	(P)	CCA	(P)	AAG	(K)	CCG	(P)
351	AGT	(S)	AGT	(S)	AGT	(S)	AGT	(S)	TCT	(S)	AGT	(S)
352	TTG	(L)	TTG	(L)	TTG	(L)	TTG	(L)	CAC	(H)	TTG	(L)
353	AAT	(N)	AAT	(N)	AAT	(N)	AAT	(N)	TCT	(S)	AAT	(N)
355	AAG	(K)	AAG	(K)	AAG	(K)	AAG	(K)	GGC	(G)	AAG	(K)
363	CAT	(H)	CAT	(H)	CAT	(H)	CAT	(H)	GAA	(E)	CAT	(H)
365	GGA	(G)	GGA	(G)	GGA	(G)	GGA	(G)	AAT	(N)	GGA	(G)
366	TTT	(F)	TTT	(F)	TTT	(F)	TTT	(F)	GAA	(E)	TTT	(F)
367	CAA	(Q)	CAA	(Q)	CAA	(Q)	CAA	(Q)	AAC	(N)	CAA	(Q)
369	CGT	(R)	CGT	(R)	CGT	(R)	CGT	(R)	TCA	(S)	CGT	(R)
371	AGT	(S)	AGT	(S)	AGT	(S)	AGT	(S)	TTT	(F)	AGT	(S)
376	AAT	(N)	AAT	(N)	AAT	(N)	AAT	(N)	ACA	(T)	AAT	(N)
381	ATT	(I)	ATT	(I)	ATT	(I)	ATT	(I)	GGT	(G)	ATT	(I)
383	AGG	(R)	AGG	(R)	AGG	(R)	AGG	(R)	CAT	(H)	AGG	(R)
384	GGT	(G)	GGT	(G)	GGT	(G)	GGT	(G)	TTT	(F)	GGT	(G)
386	GGT	(G)	GGT	(G)	GGT	(G)	GGT	(G)	TCA	(S)	GGT	(G)
1663	CAA	(Q)	CAA	(Q)	CAA	(Q)	CAA	(Q)	GTA	(V)	CAA	(Q)

**Table 6.3 CNOT1 positively selected sites in the primates lineage.**

PAML returned 18 positively selected sites among the primates lineage according to the Bayes Empirical Bayes (BEB) analyses with a P value <0.05.

## **6.3 Phylogenetic and variability analysis of CNOT2**

### **6.3.1 Phylogenetic reconstruction of CNOT2**

The CNOT2 homologous sequences were retrieved by DELTA-BLAST from the reference sequences database using the human CNOT2 as query sequence (Supporting Table 6.2).

The CNOT2 phylogenetic tree with the Neighbour-Joining (NJ) was reconstructed as described in Chapter 2 and is shown in Supporting Figure 6.2. Except for few branches, the branch support values showed a low quality tree topology due to short distances between the proteins.

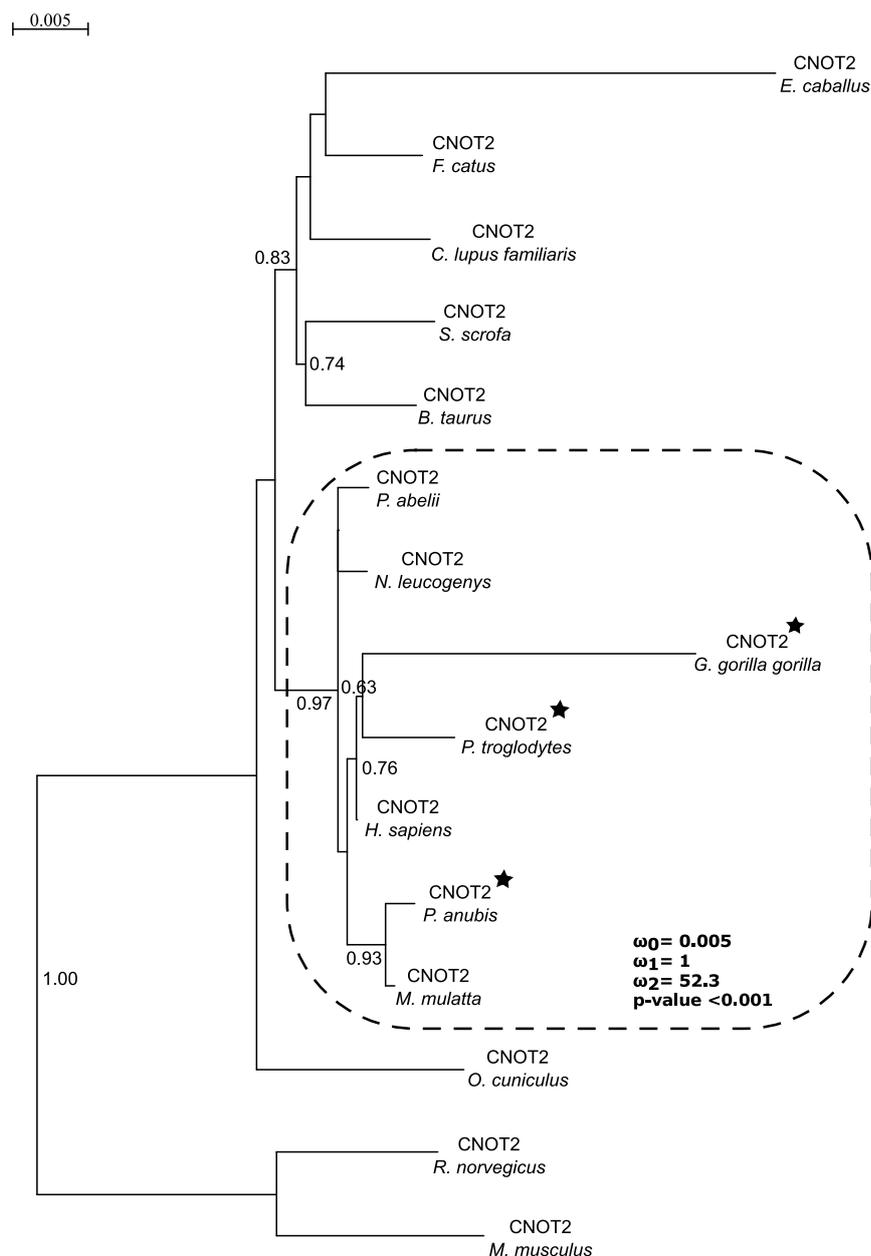
In order to reconstruct a phylogenetic tree with the Maximum Likelihood method, jModelTest (Posada, 2008; Darriba et al., 2015) was used to choose the nucleotide substitution model with the best fit according to the Akaike Information Criterion (AIC) (Akaike, 1974). The model with the lowest AIC resulted to be the Tamura-Nei model (Tamura and Nei, 1993) (TrN+I+G, 10020) indicating that there was a great proportion of invariable sites (+I) and there was rate heterogeneity among sites (+G). The resulting Maximum Likelihood tree, reconstructed by phyML 3.0 (Guindon et al., 2010), is shown in Figure 6.3 and its topology was better supported than in the NJ phylogenetic tree.

### **6.3.2 Variability analysis of the CNOT2 component**

CNOT2 phylogenetic tree was inspected to detect positive selection using PAML (Yang, 2007). As shown in Table 6.4 the null hypothesis of the site model, m1a (Yang et al., 1998, 2005) was accepted meaningful of neutral selective pressures acting on the whole CNOT2 tree. Branch-site models, assuming positive selective pressures only on the primates lineage, were tested (Yang et al., 2005; Zhang et al., 2005). The null hypothesis was rejected with a P value <0.001 according to LRT calculations.

Posterior mean  $\omega$  values were calculated for each site and are shown in Figure 6.4. From the Bayes Empirical Bayes (BEB) analyses 9 sites resulted to be positively selected with a P value <0.05% and are shown in Table 6.5. These sites are mostly localised in the N-terminal domain of CNOT2. Sites

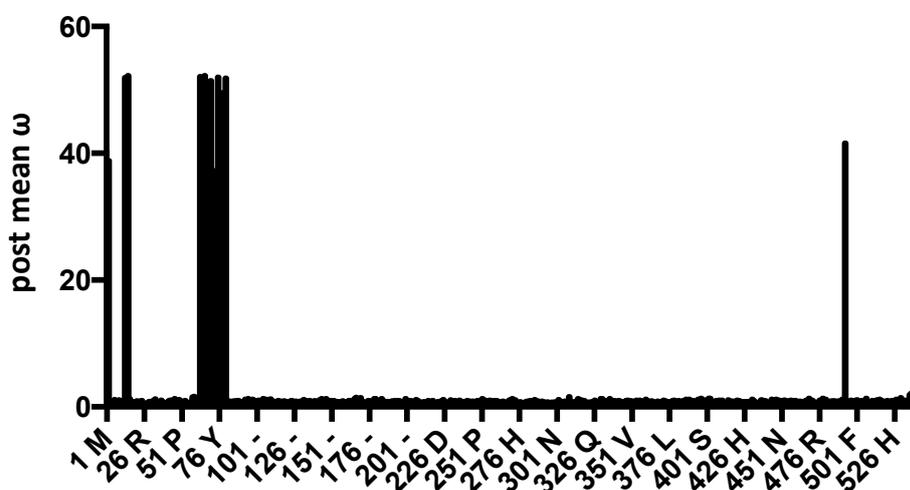
13, 14 and 15 resulted to be divergent from the other primates in *Pan troglodytes* and in *P. anubis* which presented deletion of these sites. The remaining sites resulted to be divergent in *Gorilla gorilla gorilla*.



**Figure 6.3 CNOT2 Maximum Likelihood phylogenetic tree.** The Maximum Likelihood (ML) tree was inferred using the Tamura-Nei nucleotide substitution model (Tamura and Nei, 1993) with proportion of invariable sites (I) and shape parameter of the gamma distribution (G). Support branches values are reported as Shimodaira-Hasegawa-like (SH-like). Primates lineage is highlighted by the dashed box and  $\omega$  parameter for the branch-site model are reported. The species with more than three positively selected sites which codons differ from other primates are labelled with stars.

Lineage	Model	lnL	LRT	P value	$\omega$ parameters
A	m1	-3518.82			$\omega_0 = 0.004, \omega_1 = 1,$
	m2a	-3518.82	0		$\omega_2 = 1$
P	mA <sub>0</sub>	-3497.85			$\omega_0 = 0.00532, \omega_1 = 1,$
	mA	-3472.36	50.98	<0.001	$\omega_2 = 52.29,$

**Table 6.4 CNOT2 Likelihood Ratio Test table.** The table summaries the selection models for the CNOT2 protein. The site models were tested among all the CNOT2 tree branches (A); the branch-site models (P) assumed positive selection only in the primates lineage and neutral selection in the remaining lineages. The log Likelihood (lnL), Likelihood Ratio Test (LRT), P values and estimated  $\omega$  parameters are shown in the table.



**Figure 6.4 CNOT2 posterior mean  $\omega$  graph.** Posterior mean  $\omega$  values relative to the primates lineage were calculated for each site.

Site	<i>H. Sapiens</i>	<i>P. abelii</i>	<i>M. mulatta</i>	<i>G. gorilla gorilla</i>	<i>N. leucogenys</i>	<i>P. troglodytes</i>	<i>P. anubis</i>
13	AGA (R)	AGA (R)	AGA (R)	AGA (R)	AGA (R)	GAA (E)	---
14	AAC (N)	AAC (N)	AAC (N)	AAC (N)	AAC (N)	GTT (V)	---
15	TAC (Y)	TAC (Y)	TAC (Y)	TAC (Y)	TAC (Y)	GCA (A)	---
63	TCT (S)	TCT (S)	TCT (S)	AGC (S)	TCT (S)	TCT (S)	TCT (S)
64	ACA (T)	ACA (T)	ACA (T)	AAC (N)	ACA (T)	ACA (T)	ACA (T)
66	GGT (G)	GGT (G)	GGT (G)	AAA (K)	GGT (G)	GGT (G)	GGT (G)
70	CAG (Q)	CAG (Q)	CAG (Q)	CCT (P)	CAG (Q)	CAG (Q)	CAG (Q)
75	TTA (L)	TTA (L)	TTA (L)	GGA (G)	TTA (L)	TTA (L)	TTA (L)
80	AGT (S)	AGT (S)	AGT (S)	AAA (K)	AGT (S)	AGT (S)	AGT (S)

**Table 6.5 CNOT2 positively selected sites in primates.** Variability analysis performed by PAML returned 9 positively selected sites among primates lineage with a P value >95%, according to the Bayes Empirical Bayes (BEB) analyses

## 6.4 Phylogenetic and variability analysis of CNOT3

### 6.4.1 Phylogenetic reconstruction of CNOT3

CNOT3 homologous sequences were retrieved by DELTA-BLAST (Boratyn et al., 2012) from the Reference sequences database using the human CNOT3 sequence as query and are listed in Supporting Table 6.3.

The *P. abelii* homologue was discarded for the analyses because of its *low quality protein* annotation status.

Homologous proteins were multi-aligned by MAFFT and their phylogeny was reconstructed using the Neighbour-Joining method by MEGA (Supporting Figure 6.3). The NJ method did not allow to obtain a robust tree topology especially for the primates lineage.

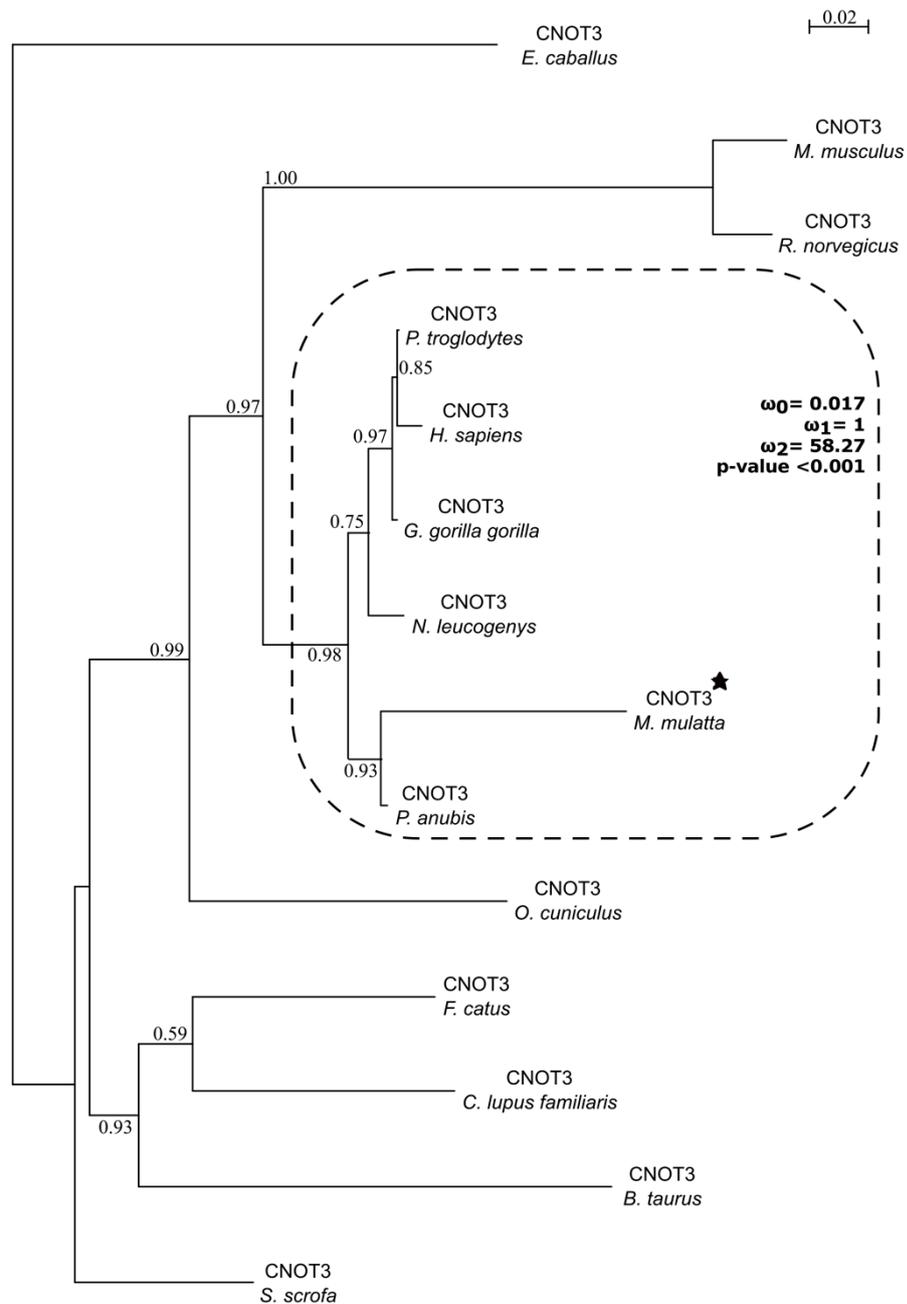
We therefore reconstructed the phylogenetic tree with the Maximum Likelihood method using the transition nucleotide substitution model (TIM1+I+G) (Posada, 2003) chosen according to Akaike Information Criterion (AIC). This latter method allowed a phylogenetic reconstruction with more robust tree topology compared to the tree obtained by the NJ method (Figure 6.5).

### 6.4.2 Variability analysis of the CNOT3

Phylogenetic Analysis by Maximum Likelihood (PAML) software was used to detect selective pressures acting on the CNOT3 mammal proteins.

As shown in Table 6.6, the null hypothesis of the site model m2a, m1a, was accepted addressing general natural selection pressures acting among all CNOT3 homologues (Yang et al., 1998, 2005). We therefore applied the branch-site model (mA) on the primates branches in order to detect positive selection in this lineage (Yang et al., 2005; Zhang et al., 2005). The Likelihood Ratio Test rejected the null hypothesis for this model with a LRT=94.91 and P value <0.001 as shown in Table 6.6. Posterior mean  $\omega$  calculations were calculated for each site and are shown in Figure 6.6 Sites with a  $\omega$  value  $\gg 1$  are mostly clustered in the C-terminal of the protein.

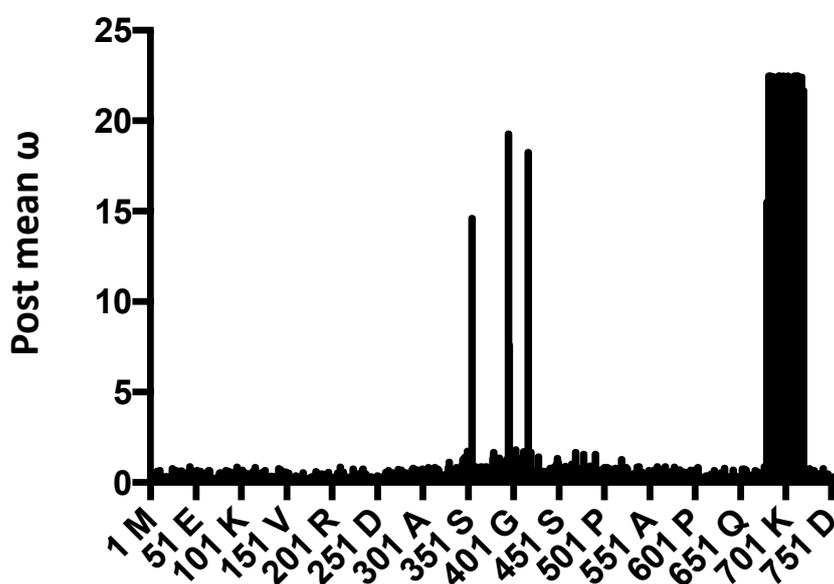
The Bayes Empirical Bayes analyses on the selected sites identified 22 positively selected sites with P value 0.05 shown in Table 6.7. In particular the *M. mulatta* CNOT3 resulted to be divergent from the other primates. Interestingly the positively selected region, spanning from the residue 682 to the residue 720, mapped on the human NOT box involved in protein-protein interaction with CNOT2.



**Figure 6.5 CNOT3 Maximum Likelihood (ML) phylogenetic tree.** The CNOT3 ML tree was inferred using the transition model (TIM1) including the proportion of invariable sites parameter (I) and the shape parameter of the gamma distribution (G). Branches support values are shown as SH-like support values. Primates lineage is included in the dashed box and  $\omega$  parameter for the branch model are reported. The species with more than three positively selected sites, which codons differ from other primates, is labelled with a star.

Lineage	Model	lnL	LRT	P value	$\omega$ parameters
A	m1	-7984.17			
	m2a	-7984.17	0		
P	mA <sub>0</sub>	-7663.03			$\omega_0=0.017, \omega_1=1,$
	mA	-7615.57	94.91	<0.001	$\omega_2=22.50,$

**Table 6.6 CNOT3 Likelihood Ratio test table.** The table summaries the models tested for the CNOT3 proteins. The site models tested among the whole CNOT3 tree (A) was rejected giving LRT 0. The branch-site models tested (P) assumed positive selection pressure on the primates lineage and neutral selection pressure for the remaining lineages. The log Likelihood, Likelihood Ratio Tests, P values and estimated  $\omega$  parameters are reported in the table for the models.



**Figure 6.6 CNOT3 posterior mean  $\omega$  graph.** Posterior mean  $\omega$  values were calculated for each analysed site by branch-model in the primates lineage.

Site	<i>H. sapiens</i>		<i>M. mulatta</i>		<i>P. anubis</i>		<i>P. troglodytes</i>		<i>G. gorilla gorilla</i>		<i>N. leucogenys</i>	
682	AAG	(K)	GCC	(A)	AAG	(K)	AAG	(K)	AAG	(K)	AAG	(K)
684	CAG	(Q)	GTC	(V)	CAG	(Q)	CAG	(Q)	CAG	(Q)	CAG	(Q)
685	TAT	(Y)	CCC	(P)	TAC	(Y)	TAC	(Y)	TAC	(Y)	TAC	(Y)
687	GCA	(A)	CGG	(R)	GCA	(A)	GCA	(A)	GCA	(A)	GCA	(A)
690	GCC	(A)	GGG	(G)	GCC	(A)	GCC	(A)	GCC	(A)	GCC	(A)
691	CTA	(L)	GCC	(A)	CTA	(L)	CTA	(L)	CTA	(L)	CTA	(L)
692	AAG	(K)	GAC	(D)	AAG	(K)	AAG	(K)	AAG	(K)	AAG	(K)
693	AAG	(K)	CCC	(P)	AAG	(K)	AAG	(K)	AAG	(K)	AAG	(K)
694	CAG	(Q)	ACT	(T)	CAG	(Q)	CAG	(Q)	CAG	(Q)	CAG	(Q)
696	TGG	(W)	CTC	(L)	TGG	(W)	TGG	(W)	TGG	(W)	TGG	(W)
697	CGA	(R)	GTC	(V)	CGA	(R)	CGA	(R)	CGA	(R)	CGA	(R)
698	TTC	(F)	GGC	(G)	TTC	(F)	TTC	(F)	TTC	(F)	TTC	(F)
702	TAC	(Y)	GCC	(A)	TAC	(Y)	TAC	(Y)	TAC	(Y)	TAC	(Y)
703	ATG	(M)	TCC	(S)	ATG	(M)	ATG	(M)	ATG	(M)	ATG	(M)
706	TTC	(F)	CCC	(P)	TTC	(F)	TTC	(F)	TTC	(F)	TTC	(F)
708	AGG	(R)	AAC	(N)	AGG	(R)	AGG	(R)	AGG	(R)	AGG	(R)
710	GAG	(E)	CCT	(P)	GAG	(E)	GAG	(E)	GAG	(E)	GAG	(E)
711	GAG	(E)	CCA	(P)	GAG	(E)	GAG	(E)	GAG	(E)	GAG	(E)
713	AAG	(K)	GCC	(A)	AAG	(K)	AAG	(K)	AAG	(K)	AAG	(K)
714	ACC	(T)	CAG	(Q)	ACC	(T)	ACC	(T)	ACC	(T)	ACC	(T)
718	GAG	(E)	CAC	(H)	GAG	(E)	GAG	(E)	GAG	(E)	GAG	(E)
720	GAG	(E)	GGC	(G)	GAG	(E)	GAG	(E)	GAG	(E)	GAG	(E)

**Table 6.7 CNOT3 positively selected sites in the primates lineage.**

PAML returned 22 positively selected sites among the primates lineage with a P value <0.05 according to the Bayes Empirical Bayes (BEB) analyses

## 6.5 Phylogenetic and variability analysis of the Ccr4 subunits

### 6.5.1 Phylogenetic reconstruction of the Ccr4 subunits

The Ccr4 subunit is one of the two catalytically active subunits of the Ccr4-Not complex. Studies showed that this deadenylase enzyme is highly conserved among animals and vertebrates (Winkler and Balacco, 2013). In mammals two paralogous proteins, encoded by the CNOT6 and CNOT6L genes, are present (Supporting Table 6.4). The *O. cuniculus*, *P. Anubis*, and *P. abelii* CNOT6 homologous proteins were discarded for the analyses because of their annotation status as *low quality protein* and *partial CDS*. The Neighbour-Joining tree (Supporting Figure 6.4) shows the orthologues separating in two distinct branches with less distance between the CNOT6L paralogues compared to the CNOT6 branches. Support branch values showed a robust tree topology especially in the CNOT6L lineage. The Maximum Likelihood phylogenetic tree of the Ccr4 homologues (Figure 6.7) was constructed using the General Time Reversible nucleotide substitution model (Lanave et al., 1984; Rodríguez et al., 1990) including the gamma shape parameter (GTR+G, 012345) chosen by AIC calculations. Branches were tested using the non-parametric SH-like branch support test and addressed a robust tree topology despite less distances in the CNOT6 lineage.

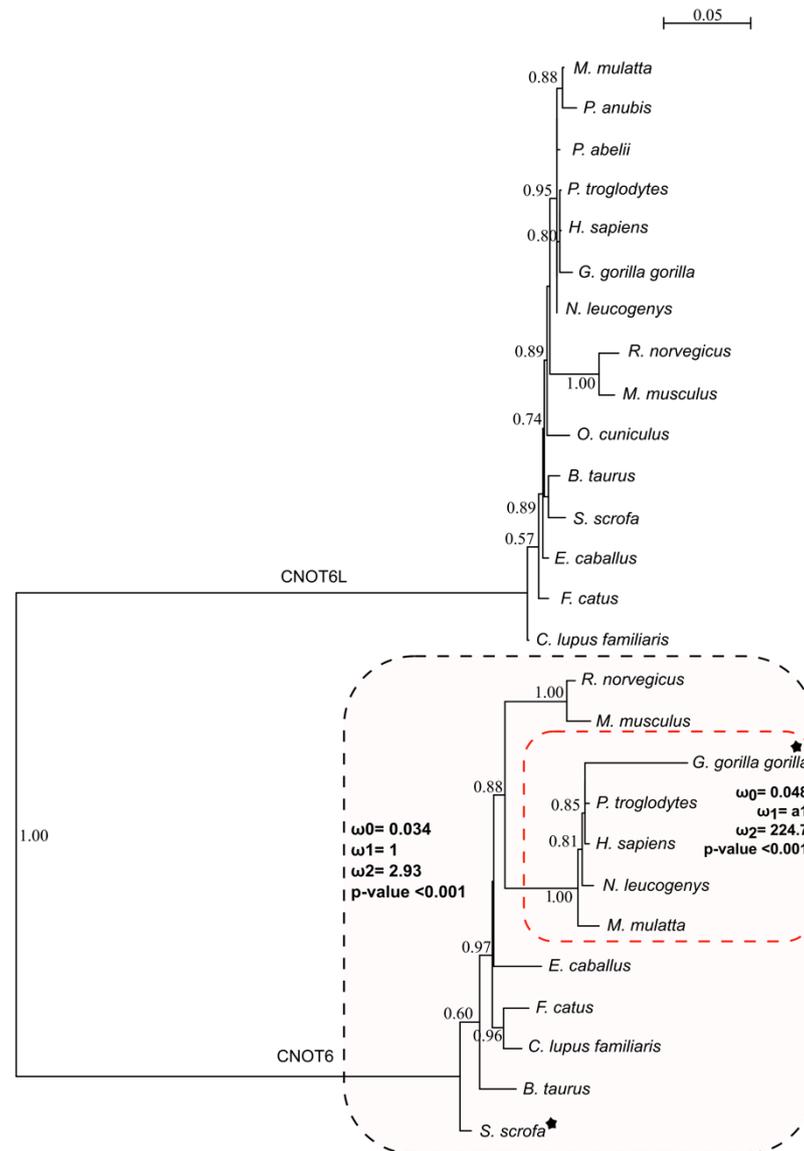
The phylogenetic trees, reconstructed using both methods, show CNOT6 and CNOT6L proteins forming two distinct branches with less divergence between the CNOT6L orthologues than between the CNOT6 ones.

### 6.5.2 Variability analysis of the Ccr4 homologues

Ccr4 homologues were investigated in order to detect selective pressures. Comparison between the site model m2a and its null hypothesis m1 in order to detect positive selection acting on both the CNOT6 and CNOT6L branches was performed (Yang et al., 1998, 2005). The null hypothesis was accepted by LRT calculation allowing to infer general neutral selection acting on both CNOT6 and CNOT6L branches as shown in Table 6.8 and from post mean  $\omega$  values shown in Figure 6.8.

I therefore try to detect selective pressures acting in the two orthologous branches separately (Yang et al., 2005; Zhang et al., 2005). The Likelihood Ratio test on the branch-site model applied to the CNOT6 orthologues rejected the null hypothesis inferring positive selection with a P value <0.001 (Figure 6.8). Ten sites resulted to be positively selected with a P value <0.05 as shown in posterior mean  $\omega$  graph (Figure 6.8 B). The selected sites, listed in Table 6.9, are localised in the amino terminal of the protein of the *S. scrofa* and *G. gorilla gorilla*. Moreover the CNOT6 branch was inspected focusing on the primates lineage, assuming positive selection acting only within this lineage. As shown in Table 6.8,, LRT calculation accepted our hypothesis. Posterior mean  $\omega$  calculations (Figure 6.8 C) and BEB analysis, returned 17 positively selected sites with P value <0.005. Once more the sites were localised in the N-terminus. In particular it is clear how the *G. gorilla gorilla* N-terminal underwent to positive selective pressures diverging from the other primate species.

The models accepted inferred positive selective pressure acting on the CNOT6 branch and more specifically among primates assuming neutral selection on the remaining lineage. As confirmation I performed a Likelihood Ratio Test for the branch model assuming positive selection among all the CNOT6L lineage. This model was rejected and the null hypothesis accepted. Furthermore the CNOT6L primates lineage was inspected. LRT calculation rejected the null hypothesis and positive selective pressures were detected in this lineage. No sites were selected with a P value <0.05.



**Figure 6.7 Ccr4 Maximum Likelihood phylogenetic tree.** The Maximum Likelihood (ML) tree was inferred using the General Time Reversible model including the shape parameter of the gamma distribution (GTR+G). Branches values show the SH-like test support values. The tree shows how the paralogues separates in two distinct branches with less distance in the CNOT6L branch comparing to the CNOT6 one. Branch-site models are highlighted by dashed boxes. In the black dashed box is highlighted the branch-site model relative to the CNOT6 lineage, in the red dashed box is highlighted the branch-site model relative to the primates lineage. Estimated  $\omega$  parameters for the models are reported in the respective boxes. The species with more than 3 positively selected sites are labelled with a star.

Protein	Lineage	Model	lnL	LRT	P value	$\omega$ parameters
CNOT6- CNOT6L	A	m1	-6084.43			
		m2a	-6084.43	0	Null hypothesis accepted	
CNOT6	A	H <sub>0</sub>	-6553.87			
		H <sub>1</sub>	-6546.65	14.46	<0.001	$\omega_0=0.034, \omega_1=1, \omega_2=2.93, \omega_3=2.93$
	P	H <sub>0</sub>	-6539.28			
		H <sub>1</sub>	-6504.66	69.25	<0.001	$\omega_0=0.033, \omega_1=1, \omega_2=22.5, \omega_3=22.5$
CNOT6L	A	H <sub>0</sub>	-6584.93			
		H <sub>1</sub>	-6577.59	-14.69	Null hypothesis accepted	
	P	H <sub>0</sub>	-6499.06			
		H <sub>1</sub>	-6567.49	136.8	<0.001	$\omega_0=0.048, \omega_1=1, \omega_2=224.78, \omega_3=224.78$

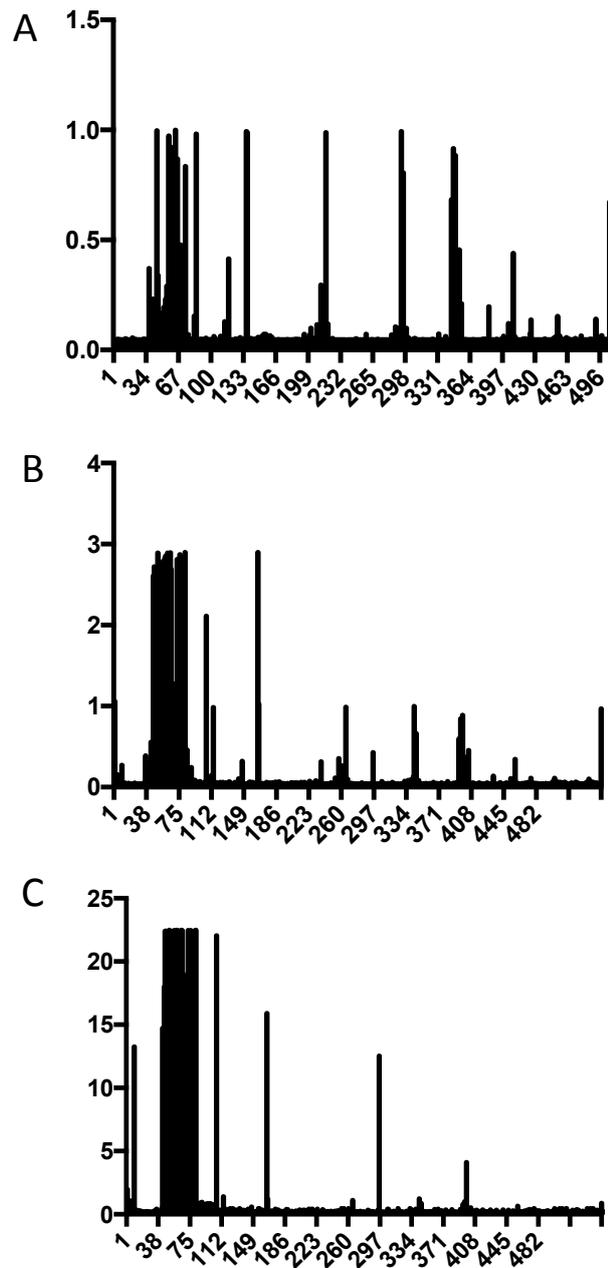
**Table 6.8 Ccr4 Likelihood Ratio Test table.** The table summaries the models tested on the Ccr4 homologue proteins. The site models were tested among the whole Ccr4 tree (CNOT6-CNOT6L). The branch-site models were testes on the CNOT6 whole (A) and primates (P) lineage and CNOT6L whole (A) and primates (P) lineage. The log likelihood, Likelihood Ratio tests, P values, and estimated  $\omega$  parameters are reported for each model applied.

Amino acid	Site	<i>CNOT6 H. sapiens</i>	<i>CNOT6 G. gorilla gorilla</i>	<i>CNOT6 N. leucogenys</i>	<i>CNOT6 P. troglodytes</i>	<i>CNOT6 M. mulatta</i>	<i>CNOT6 M. musculus</i>	<i>CNOT6 C. lupus familiaris</i>	<i>CNOT6 S. scrofa</i>	<i>CNOT6 R. norvegicus</i>	<i>CNOT6 F. catus</i>	<i>CNOT6 B. Taurus</i>	<i>CNOT6 E. caballus</i>
46	51	TCT (S)	GGG (G)	TCT (S)	TCT (S)	TCT (S)	TCT (S)	TCT (S)	TCT (S)	TCT (S)	TCT (S)	TCT (S)	TCT (S)
52	57	CAC (H)	ACT (T)	CAC (H)	CAC (H)	CAC (H)	CAC (H)	CAC (H)	CAC (H)	CAC (H)	CAC (H)	CAC (H)	CAC (H)
54	59	ACA (T)	TTG (L)	ACA (T)	ACA (T)	ACA (T)	ACA (T)	ACA (T)	ACA (T)	ACA (T)	ACA (T)	ACA (T)	ACA (T)
55	60	GCT (A)	TTA (L)	GCT (A)	GCT (A)	GCT (A)	GCT (A)	GCT (A)	GCT (A)	GCT (A)	GCT (A)	GCT (A)	GCT (A)
57	62	CAT (H)	ATT (D)	CAT (H)	CAT (H)	CAT (H)	CAT (H)	CAC (H)	TAT (Y)	CAT (H)	CAT (H)	TAT (Y)	CAT (H)
58	63	TTG (L)	GGG (G)	TTG (L)	TTG (L)	TTG (L)	CTG (L)	CTG (L)	CTA (L)	CTG (L)	CTG (L)	CTG (L)	CTG (L)
60	65	GAC (D)	ATA (D)	GAC (D)	GAC (D)	GAC (D)	GAC (D)	GAC (D)	GAC (D)	GAC (D)	GAC (D)	GAC (D)	GAC (D)
71	76	GCC (A)	CTG (L)	GCC (A)	GCC (A)	GCC (A)	GCC (A)	GCC (A)	GCC (A)	GCC (A)	GCC (A)	GCC (A)	GCC (A)
77	82	CTG (L)	AGT (S)	CTG (L)	CTG (L)	CTG (L)	CTG (L)	CTG (L)	CTG (L)	CTG (L)	CTG (L)	CTG (L)	CTG (L)
164	165	ATT (I)	GTT (V)	GTT (V)	ATT (I)	ATT (I)	GTT (V)	ATT (I)	GTT (V)	ATT (I)	ATT (I)	ATT (I)	ATT (I)

**Table 6.9 CNOT6 positively selected sites.** PAML returned 10 positively selected sites among the CNOT6 orthologues according to the BEB analyses with a P value <0.05.

Amino Acid	Site	<i>CNOT6 H. sapiens</i>	<i>CNOT6 G. gorilla</i>	<i>CNOT6 N. leucogenys</i>	<i>CNOT6 P. troglodytes</i>	<i>CNOT6 M. mulatta</i>
41	46	AGA (R)	GAG (E)	AGA (R)	AGA (R)	AGA (R)
42	47	AGC (S)	GAT (D)	AGC (S)	AGC (S)	AGC (S)
45	50	GCA (A)	AGC (S)	GCA (A)	GCA (A)	GCA (A)
46	51	TCT (S)	GGG (G)	TCT (S)	TCT (S)	TCT (S)
51	56	ACT (T)	TGT (C)	ACT (T)	ACT (T)	ACT (T)
52	57	CAC (H)	ACT (T)	CAC (H)	CAC (H)	CAC (H)
53	58	CTG (L)	CAC (H)	CTG (L)	CTG (L)	CTG (L)
54	59	ACA (T)	TTG (L)	ACA (T)	ACA (T)	ACA (T)
55	60	GCT (A)	TTA (L)	GCT (A)	GCT (A)	GCT (A)
58	63	TTG (L)	GGG (G)	TTG (L)	TTG (L)	TTG (L)
60	65	GAC (D)	ATA (I)	GAC (D)	GAC (D)	GAC (D)
61	66	AAT (N)	TCC (S)	AAT (N)	AAT (N)	AAT (N)
68	73	TCA (S)	AGT (S)	TCA (S)	TCA (S)	TCA (S)
71	76	GCC (A)	CTG (L)	GCC (A)	GCC (A)	GCC (A)
75	81	AAT (N)	TCC (S)	AAT (N)	AAT (N)	AAT (N)
76	82	CTG (L)	AGT (S)	CTG (L)	CTG (L)	CTG (L)
100	106	AGG (R)	CAG (Q)	AGG (R)	AGG (R)	AGG (R)

**Table 6.10 CNOT6 positively selected sites in the primate lineage.** PAML returned 17 positively selected sites among the primates lineage according to the BEB analysis with a P value <0.05.



**Figure 6.8 CNOT6 and CNOT6L posterior  $\omega$  graph. (A)** Selective pressures were investigated on the CNOT6-CNOT6L phylogenetic tree. No site resulted to be under positive selective pressures. Posterior mean  $\omega$  were calculated for each site. **(B)** Posterior mean  $\omega$  were calculated for the branch-model applied for the CNOT6L lineage highlighting sites under positive selective pressures. **(C)** Posterior mean  $\omega$  were calculated for the branch-model applied for the CNOT6L primates lineage highlighting sites under positive selective pressures.

## 6.6 Phylogenetic and variability analysis of the Caf1 subunit

### 6.6.1 Phylogenetic reconstruction of the Caf1 subunit

The Caf1 subunit is the DEED type deadenylase of the complex. As for the case of the Ccr4 subunit, vertebrates present two paralogues of the Caf1 gene i.e. CNOT7 and CNOT8 (Winkler and Balacco, 2013). Homologues sequences were retrieved as previously described from the Refseq database and multi-aligned using the MAFFT software. Retrieved protein sequences are listed in Supporting Table 6.5.

The *E. caballus* CNOT8 sequence was not used for further analysis because of its *low quality protein* annotation status. Phylogenetic reconstruction of the CNOT7 and CNOT8 proteins was performed with the Neighbour-Joining (NJ) method using the MEGA software and using bootstrap 1000 as branch support statistics. The resulting tree, visualised by Seaview, is shown in Supporting Figure 6.5. Orthologues clustered in two separate branches with less distance in the CNOT7 branch compared to the CNOT8 one. Unfortunately low distances between the CNOT7 orthologues did not allow a robust tree topology.

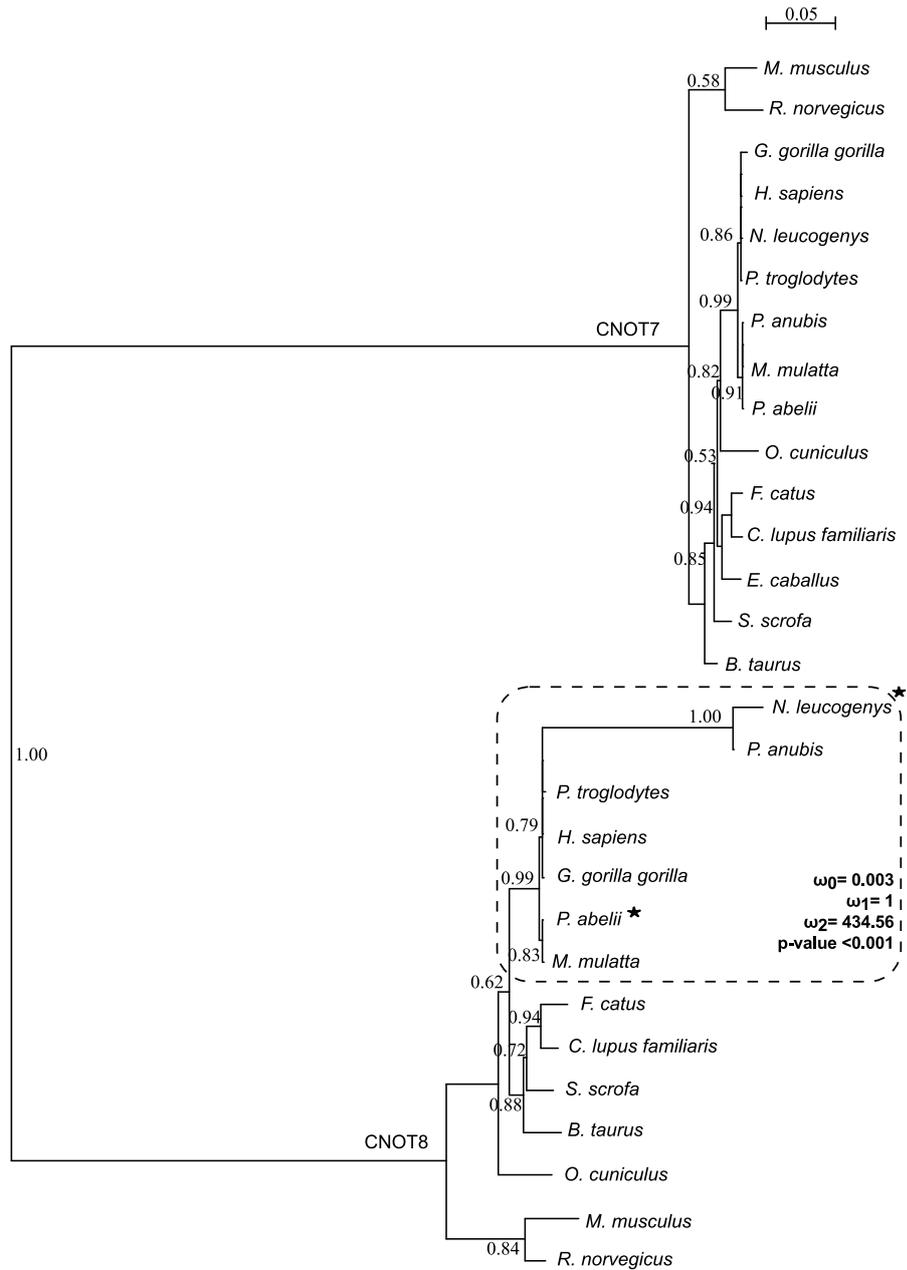
In order to reconstruct the phylogeny of the Caf1 homologues with the Maximum Likelihood method, nucleotide substitution model choice was evaluated using jModelTest. The nucleotide substitution model chosen according to the AIC calculations was the transition model (TIM1+G) with rate heterogeneity among sites (+G) including the gamma shape parameter (TIM1+G, 012230) (Posada, 2003).

The phylogenetic tree was then reconstructed using the TIM1+G nucleotide substitution model using phyML. Branches were tested using the non-parametric SH-like branch support test. The Caf1 phylogenetic tree is shown in Figure 6.9.

The phylogenetic trees, reconstructed using the two different methods, show how CNOT7 and CNOT8 paralogues form two clear distinct branches.

### 6.6.2 Variability analysis of the Caf1 homologues

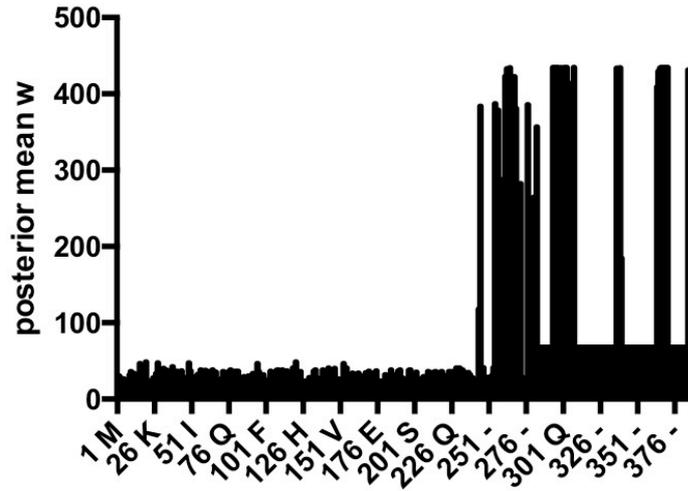
Caf1 homologues were investigated in order to detect variability patterns within the mammals. Firstly, the comparison between the site-model m2a and its null hypothesis m1 was performed as described in Chapter 2 (Yang et al., 1998, 2005). The null hypothesis was accepted by LRT calculations (Table 6.11) inferring neutral positive selection acting on the CNOT7 and CNOT8 branches. I therefore inspected the two paralogous branches independently. The branch-site null hypothesis was accepted assuming neutral selection acting on these branches (Yang et al., 2005; Zhang et al., 2005). Moreover I applied the branch-site model to the primate branches in CNOT8. Unfortunately it was not possible to apply the branch-site model for the CNOT7 primate lineage because the distances of the orthologous proteins were too low. Lastly, the branch-site model was applied on the CNOT8 primate lineage. In this case, the null hypothesis was rejected with a P value <0.01 and posterior mean  $\omega$  values were calculated (Figure 6.10). Eighteen sites resulted to be positively selected according to the BEB analyses with a P value <0.05%. The sites mapped between residues 140 and 180. Codons in *P. abelii* and *N. leucogenys* resulted to differ significantly from the other primate species (Table 6.12).



**Figure 6.9 Caf1 Maximum Likelihood phylogenetic tree.** The Maximum Likelihood (ML) tree was inferred using the General Time Reversible model including the shape parameter of the gamma distribution (GTR+G). Branches values show the SH-like test support values. The tree shows how the orthologues clearly clustered in two separate branches with less distance in the CNOT7 branch comparing to the CNOT8 one. Estimated  $\omega$  parameters for the model are reported in the dashed box. The species with more than three positively selected sites are labelled with a star.

Protein	Lineage	Model	lnL	LRT	P value	$\omega$ parameters
CNOT7- CNOT8	A	m1	-4045.6			
		m2a	-4045.6	0	Null hypothesis accepted	
CNOT8	A	mA <sub>0</sub>	-4045.6			
		mA	-4045.6	0	Null hypothesis accepted	
	P	mA <sub>0</sub>	- 4023.79			$\omega_0=0.003, \omega_1=1, \omega_2=434.56$
		mA	- 3913.23	221.12	<0.001	
CNOT7	A	mA <sub>0</sub>	-4045.6			
		mA	-4045.6	0	Null hypothesis accepted	

**Table 6.11 Caf1 Likelihood Ratio Test table.** The table summarises the models tested on the Caf1 homologue proteins. The site models were tested among the whole Caf1 tree (CNOT7-CNOT8). The branch-site models were tested on the CNOT8 whole (A) and primates lineage (P) and CNOT7 whole lineage (A). The log likelihood, Likelihood Ratio tests, P values, and estimated  $\omega$  parameters are reported for each model applied.



**Figure 6.10 CNOT8 posterior mean  $\omega$  graph.** Posterior mean  $\omega$  values were calculated for each analysed site by branch-site model in the priates lineage.

	P.									
AA Site	<i>H. sapiens</i>	<i>M. mulatta</i>	<i>anubis</i>	<i>G. gorilla</i>	<i>gorilla</i>	<i>P. troglodytes</i>	<i>P. abelii</i>	<i>N. leucogenys</i>		
142	262	GAC (D)	GAC (D)	GAC (D)	GAC (D)	GAC (D)	AAG (K)	AAG (K)		
143	263	AGC (S)	AGC (S)	AGC (S)	AGC (S)	AGC (S)	GAG (E)	GAG (E)		
145	265	GAT (D)	GAT (D)	GAT (D)	GAT (D)	GAT (D)	GTG (V)	GCG (A)		
148	268	AAG (K)	AAG (K)	AAG (K)	AAG (K)	AAG (K)	CTG (L)	CTG (L)		
157	294	GGC (G)	GGC (G)	GGC (G)	GGC (G)	GGC (G)	---	TCT (S)		
158	295	ACA (T)	ACA (T)	ACA (T)	ACA (T)	ACA (T)	---	GTC (V)		
159	296	GGA (G)	GGA (G)	GGA (G)	GGA (G)	GGA (G)	---	ACC (T)		
160	297	GTG (V)	GTG (V)	GTG (V)	GTG (V)	GTG (V)	---	CAG (Q)		
162	299	CAG (Q)	CAG (Q)	CAG (Q)	CAG (Q)	CAG (Q)	---	GGA (G)		
163	300	AAG (K)	AAG (K)	AAG (K)	AAG (K)	AAG (K)	---	GTG (V)		
165	302	AAT (N)	AAT (N)	AAT (N)	AAT (N)	AAT (N)	---	TGG (W)		
166	303	GAG (E)	GAG (E)	GAG (E)	GAG (E)	GAG (E)	---	TGC (C)		
168	305	GTG (V)	GTG (V)	GTG (V)	GTG (V)	GTG (V)	---	CTT (L)		
169	306	GAC (D)	GAC (D)	GAC (D)	GAC (D)	GAC (D)	---	GGC (G)		
171	308	GCC (A)	GCC (A)	GCC (A)	GCC (A)	GCC (A)	---	CTG (L)		
173	337	GAG (E)	GAG (E)	GAG (E)	GAG (E)	GAG (E)	---	GCT (A)		
175	339	ATG (M)	ATG (M)	ATG (M)	ATG (M)	ATG (M)	---	TTT (F)		
180	365	ATT (I)	ATT (I)	ATT (I)	ATT (I)	ATT (I)	---	GCC (A)		

**Table 6.12 CNOT8 positively selected sites in the primate lineage.** PAML returned 18 positively selected sites among

the primates lineage according to the BEB analysis with a P value <0.05.

## 6.7 Phylogenetic and variability analysis of CNOT9, CNOT10 and CNOT11

### 6.7.1 Phylogenetic reconstruction of CNOT9, CNOT10 and CNOT11

The CNOT9, CNOT10 and CNOT11 homologues sequences were retrieved using the human sequences as query by DELTA-BLAST from the Reference sequences protein database and are listed in Supporting Table 6.6, Supporting Table 6.7, and Supporting Table 6.8 respectively.

The *M. mulatta* CNOT9 and the *S. scrofa* CNOT10 were not included into the analyses because of their annotation status. Multi-aligned codon sequences were then used to reconstruct the phylogenetic tree with the Neighbour-Joining (NJ) method. The resulting trees are shown in Supporting Figures 6.6, S Supporting Figures 6.7, Supporting Figures 6.8 respectively for CNOT9, CNOT10 and CNOT11.

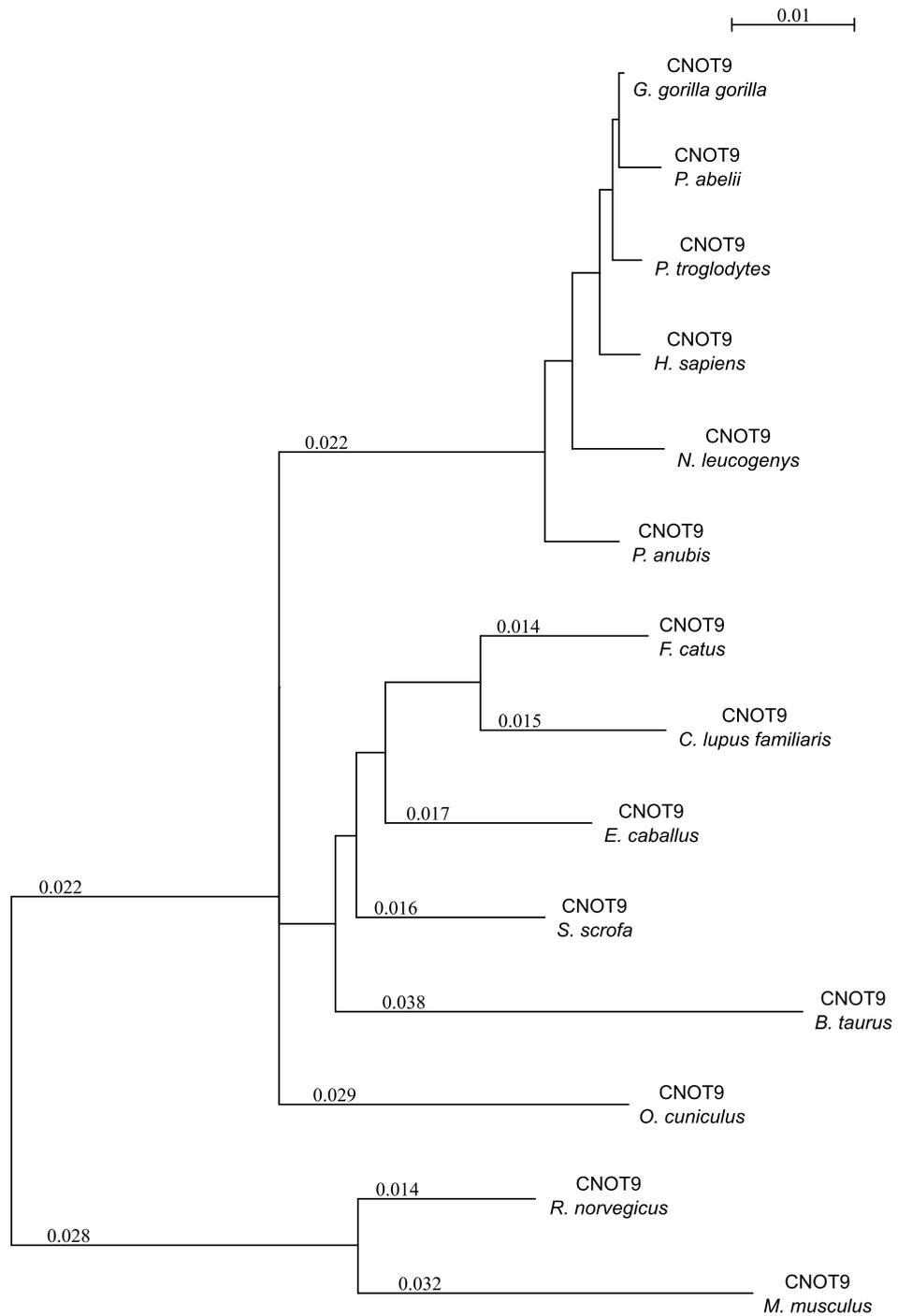
In order to reconstruct the phylogenetic trees for with the Maximum likelihood method, nucleotide substitution models were chosen for each protein using jModelTest. The nucleotide substitution model with the best fit for CNOT9 resulted to be the Tamura-Nei model (Tamura and Nei, 1993) including the gamma shape parameter (TrNef+G, 10020). The nucleotide substitution model with the best fit for CNOT10 and for CNOT11 resulted to be in both cases the general time reversible model including the gamma shape parameter (GTR+G, 012345). The phylogenetic trees were constructed using the appropriate nucleotide substitution model by phyML. The reconstructed Maximum Likelihood phylogenetic trees for CNOT9, CNOT10, and CNOT11 are shown in Figure 6.11, Figure 6.12, and Figure 6.13 respectively. Branches were tested using the non-parametric Shimodaira-Hasegawa-like (SH-like) (Shimodaira and Hasegawa, 1999) branch support test. The CNOT9 tree resulted to have a weak topology whereas CNOT10 and CNOT11 resulted to have a more robust tree topology.

### **6.7.2 Variability analysis of the CNOT9, CNOT10, and CNOT11**

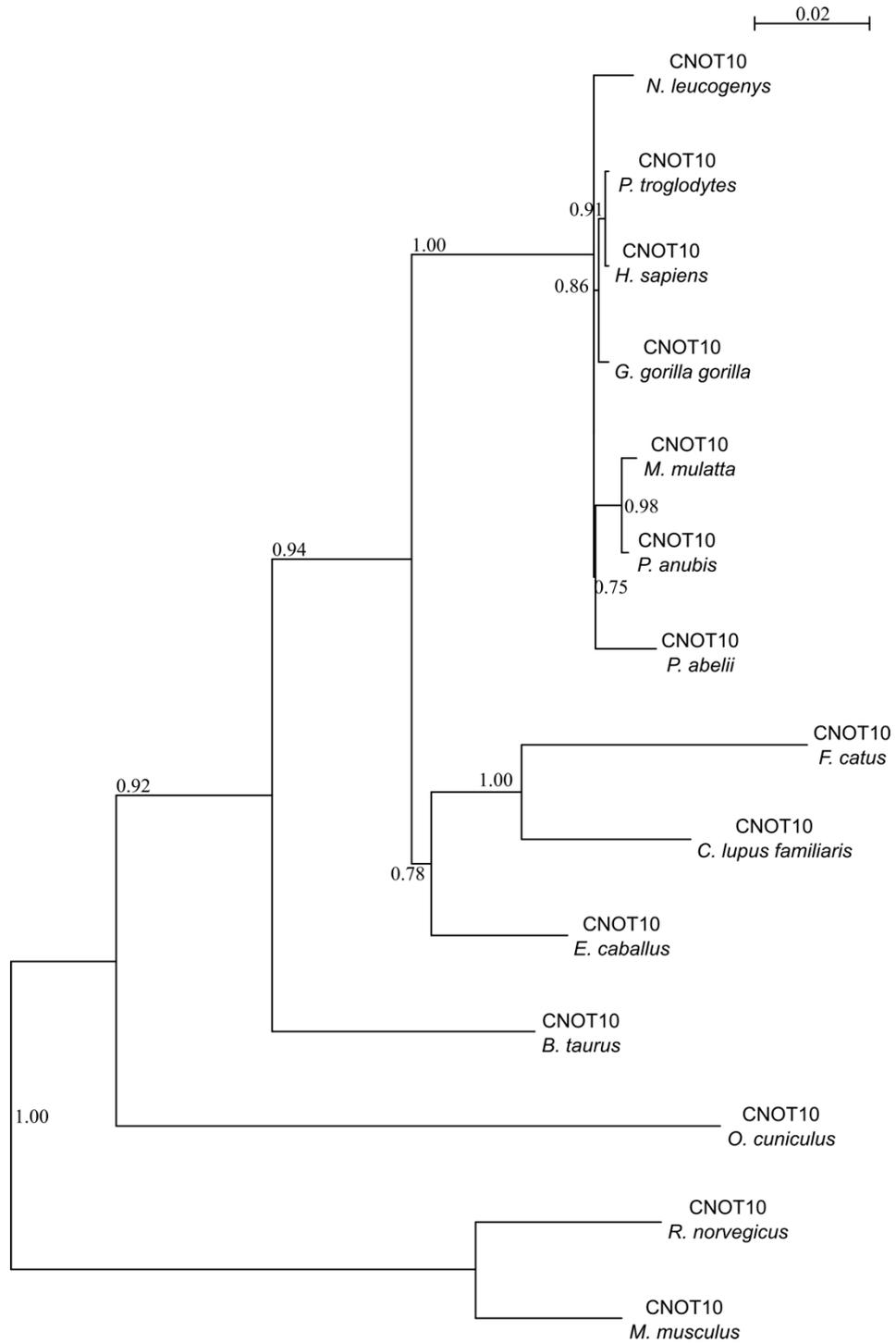
#### **subunits**

Phylogenetic Analysis by Maximum Likelihood (PAML) software was used to detect positive selection acting on the CNOT9, CNOT10 and CNOT11 mammal proteins.

We performed a comparison between the m2a site model and its null hypothesis m1 site model on the whole trees (Yang et al., 1998, 2005). As shown in Table 6.13, the null hypothesis of the model m2a, m1a, was accepted addressing general natural selection pressures acting among all the homologues in all the three cases. We therefore applied the branch-site model (mA) on the primates branches in order to detect positive selection in this lineage (Yang et al., 2005; Zhang et al., 2005). The Likelihood Ratio Test rejected the null hypothesis for this model with a P value <0.05 as shown in Table 6.13. Even though models inferring positive selection acting on the mammals were accepted, sites under positive selection were not detected in any of the three proteins.

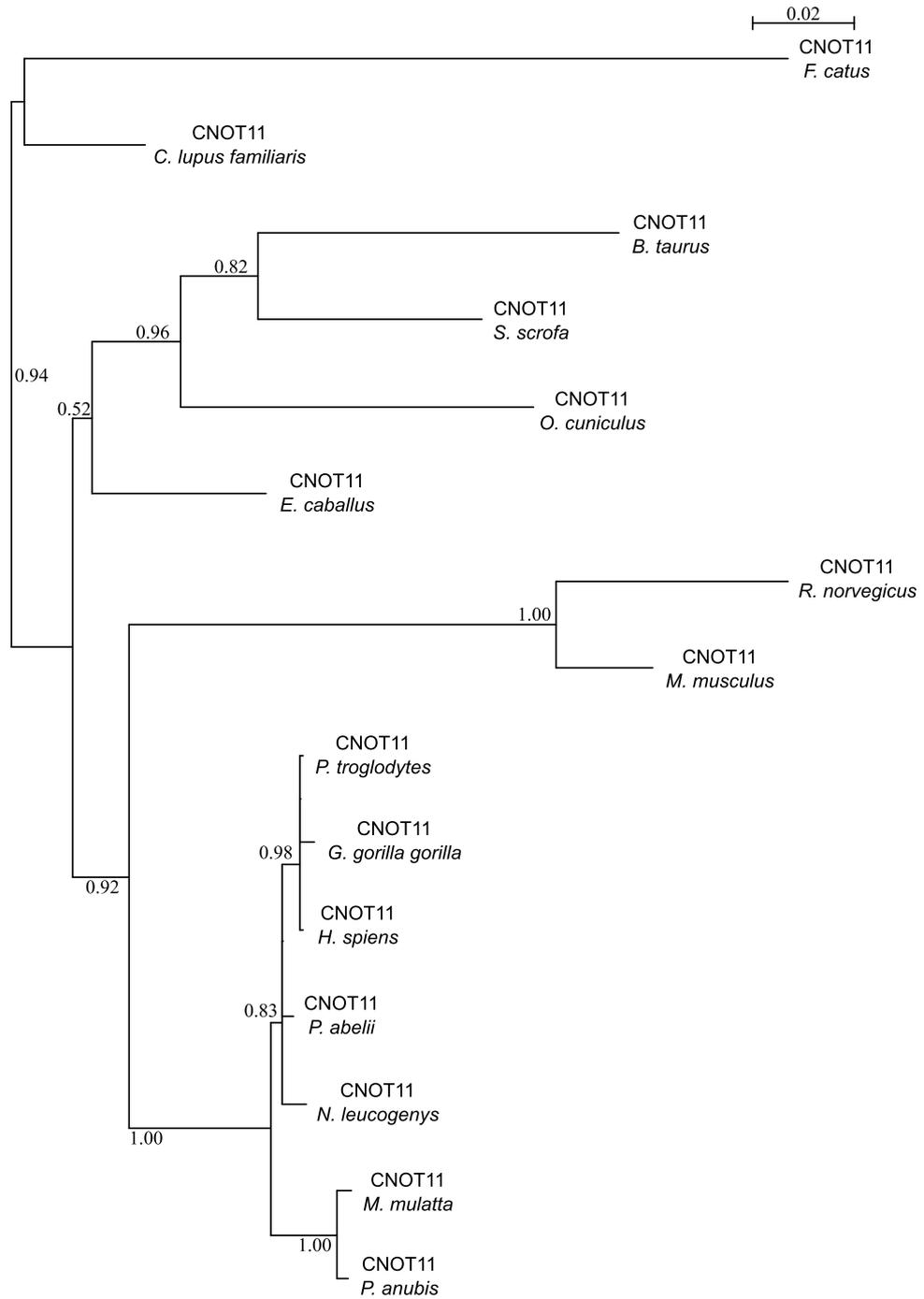


**Figure 6.11 CNOT9 Maximum Likelihood (ML) phylogenetic tree.** The CNOT9 ML phylogenetic tree was inferred using the Tamura-Nei model (TrNef) including and shape parameter of the gamma distribution (G). Branches values show the SH-like support values.



**Figure 6.12 CNOT10 Maximum Likelihood (ML) phylogenetic tree.**

The CNOT10 ML phylogenetic tree was inferred using the general time reversible model (GTR) including and shape parameter of the gamma distribution (G). Branches values show the SH-like support values.



**Figure 6.13 CNOT11 Maximum Likelihood (ML) phylogenetic tree.**

The CNOT11 ML phylogenetic tree was inferred using the general time reversible model (GTR) including and shape parameter of the gamma distribution (G). Branches values show the SH-like support values.

Protein	Lineage	model	lnL	LRT	P value
CNOT9	A	m1	-2203.99		
		m2a	-2203.99	0	Null hypothesis accepted
	P	H <sub>0</sub>	-2206.71		
		H <sub>1</sub>	-2202.6	8.22	<0.05
CNOT10	A	m1	-7554.13		
		m2a	-7554.13	0	Null hypothesis accepted
	P	H <sub>0</sub>	-7552.64		
		H <sub>1</sub>	-7549.31	6.66	<0.05
CNOT11	A	m1	-2765.69		
		m2a	-2765.69	0	Null hypothesis accepted
	P	H <sub>0</sub>	-2765.69		
		H <sub>1</sub>	-2757.63	16.12	<0.001

**Table 6.13 CNOT9, CNOT10, and CNOT11 Likelihood Ratio test table.**

The table summarises the models tested for the CNOT9, CNOT10 and CNOT11 proteins. The first model tested (A) assuming positive selection acting among all the trees, was rejected in all the cases. The second model tested assumed positive selection pressure on the primate lineage (P) and neutral selection pressure for the remaining lineages and was accepted with significant P value in all the three cases. The log Likelihood, Likelihood Ratio Tests, P values and estimated  $\omega$  parameters are reported in the table for the models applied.

### 6.7.3 Discussion

This chapter describes a comprehensive evolutionary analysis on the basis of the currently available data of the Ccr4-Not complex. The complex is composed by ten subunits that are organised in modules i.e. the deadenylase module, the NOT module and the CNOT10/11 module which are anchored on the central scaffold CNOT1 each with a specific function (Collart and Panasenko, 2012; Wahle and Winkler, 2013; Winkler and Balacco, 2013).

Phylogeny of each subunit of the deadenylase complex was reconstructed and natural positive selection was investigated in order to identify whether particular subunits have diverged significantly in specific species.

#### CNOT1

CNOT1 is the central scaffold of the complex where all the other modules are anchored (Bai et al., 1999). Studies on mice demonstrated a direct interaction of CNOT1 with NANOS2 a key role protein of the male germ cell development (Suzuki et al., 2015a). The phylogenetic reconstruction of the protein showed that *P. anubis* and *M. mulatta* diverged from other primates. The phylogenetic tree was investigated in order to detect natural selection pressures. The positive selection model, applied to the whole tree, did not identify any site under selection in any species. Investigation of the primate lineage, applying the branch-site model, identified 18 sites under positive selective pressures. Inspection of the codons revealed that these sites are subjected to positive selection in *P. anubis*. The sites resulted to be in a hotspot of positively selected sites spanning from the Asp349 to the Gly386 located in the HEAT-repeats domain in the CNOT1 N-terminal region,.

#### The NOT-module

CNOT1 interacts with CNOT2 and CNOT3 forming the NOT module that is involved in several cellular processes (Boland et al., 2013).

CNOT2 is a fundamental component of the Ccr4-Not complex. Its depletion has been shown to lead to apoptotic cell death in a caspase-dependent way (Ito et al., 2011a). The CNOT2 phylogeny showed that the *G. gorilla gorilla* and *P. troglodytes* proteins are divergent from the other primates.

Investigation of natural selection in the phylogenetic tree detected positive selection pressures only in the primate lineage. The inspection of the codons revealed two groups of positively selected sites mapping to the N-terminal of the protein. The first group, composed of Arg-13, Asn-14, and Tyr-15, are positively selected in *P. troglodytes* and in *P. anubis*. Interestingly, this latter species presented a deletion in this region. The second group was formed by the remaining six positively selected sites Ser-63, Thr-64, Gly-66, Gln-70, Leu-75, and Ser-80. These sites were only significantly changed in *G. gorilla gorilla* and may be linked to a gain of function in this species as adaptive evolution.

The other component of the NOT module, CNOT3, was shown to regulate mitotic progression (MAD1) and metabolism (Morita et al., 2011; Takahashi et al., 2012). The phylogenic reconstruction of CNOT3 deviated from the precedent tree phylogeny of the above-mentioned proteins. In particular, the *M. musculus* and *R. norvegicus* lineage separated from the primates lineage in contrast to the previous phylogenies where it diverged from all the other species. Moreover The *E. caballus* sequence diverged significantly from all the other sequences. *M. mulatta* was the most distant species in the primates lineage was.

As for the case of the other NOT module components, positive selection was detected only in the primate lineage. The 22 positively selected sites were identified in a hotspot spanning from Lys-682 to Glu-720. Inspection of the codons revealed the *M. mulatta* to be the only species with divergent codons. Intriguingly the positively selected sites mapped in the NOT box of the CNOT3, in particular in the domain responsible for the hetero-dimerization with CNOT2 (Boland et al., 2013). The changing in this important domain may be significant of function gaining in the macaque.

### The nuclease module

The Ccr4-NOT complex contains two components that are associated with the deadenylase activity: the Ccr4 and the Caf1 (Winkler and Balacco, 2013). In vertebrates two paralogues of the Ccr4 are present and are encoded by the CNOT6 and CNOT6L genes (Wang et al., 2010; Cooke et al., 2010). The phylogeny of the Ccr4 proteins showed the two paralogues to form two separate branches congruently with what previously observed (Winkler and Balacco, 2013).

Interestingly, the CNOT6L branch showed to be more distant than the CNOT6 branch from the hypothetical ancestor protein. The distances between the CNOT6 orthologues resulted to be bigger than the distances between the CNOT6L orthologues. It can be speculated that CNOT6L proteins diverged significantly from the ancestor with an equal rate whereas orthologues that resulted to highly conserved. CNOT6, by contrast, is more closely related to the ancestor and the orthologues appear to have diversified further. Investigation of natural selection pressure among the phylogenetic tree identified positive selective pressures in the CNOT6 branch but not in the CNOT6L one. The analysis detected ten positively selected sites. Inspection of the codons revealed that sites 62, 63 and 165 are divergent in *S. scrofa* and *G. gorilla gorilla* compared to the other species. The remaining sites have diverged only in *G. gorilla gorilla*. We focused then only on the CNOT6 primate lineage identifying ten additional positively selected sites. These were sites under positive pressure exclusively between primates and, once again, the *G. gorilla gorilla* is the diverging primate.

Data suggest that the two paralogues are under two different selective evolutionary pressures. The CNOT6 protein is at the same time more closely related to the ancestor sequence but more heterogeneous in the orthologous sequences. Natural selection acted positively selecting important mutations with possible gain of function. The CNOT6L proteins, in contrast, were identified to be more distant from the ancestor sequence but more homogenous in the orthologous sequences. Positive selection was not detected.

As for the case of Ccr4, Caf1 is present in vertebrates as two closely related paralogues encoded by CNOT7 and CNOT8. The phylogenetic reconstruction showed paralogues separating in two different branches in agreement with previous findings (Winkler and Balacco, 2013). As for the case of the Ccr4 homologues, one of the two Caf1 paralogues, CNOT7, resulted to be more distant from the ancestor sequence but more homogeneous in the orthologous sequences.

The CNOT8 branch is more heterogeneous allowing a better branch separation. In this lineage, two species appeared to be more distant than the others: *N. leucogenys* and *P. anubis*. The tree was inspected in order to identify natural selective pressures. In contrast with the Ccr4 case, no positive selection was found in CNOT8 lineage. When focusing on the CNOT8 primate lineage, though, positive selection was detected. This was an expected result because distances in the CNOT8 primate sequences were very low except for the two above-mentioned species. This analysis returned 18 positively selected sites localised in a region between Asp-262 and Ile-365. Codon investigations revealed that the *N. leucogenys* presents non-synonymous mutations that were selected. *P. abelii*, presented non-synonymous mutations for the 262, 263, 265, and 268 sites, but contained a deletion covering the remaining 14 sites. This suggested that the more heterogeneous status of the orthologues in the CNOT8 branch allowed the protein sequences to positively select non-synonymous mutation keeping the CNOT7 counterpart more homogeneous and less variable.

#### RQCD-1 and the N-terminal module

RQCD1 (CNOT9) is a component of the Ccr4-Not complex that is involved in miRNA-mediated gene silencing by recruitment of TRNC6/GW182 proteins. The phylogenetic reconstruction of this protein resulted in a tree with robust topology. Unfortunately, even though positive selective pressures were detected in the primate lineage, no sites were found to be positively selected.

The CNOT10/CNOT11 module is the last component of the Ccr4-Not complex. CNOT10 is an essential component for the integrity of the Ccr4

complex but the function of the CNOT10/CNOT11 module is still not clear (Mauxion et al., 2013). As for the case of CNOT9, the phylogenetic reconstruction returned trees with robust topology for both proteins. The proteins were found to be under positive selection in the primate lineage but unfortunately no sites were positively selected.

In summary, the analysis described in this chapter reconstructed the phylogeny of each component of the Ccr4-Not complex and identified selective pressures acting on it. The study demonstrated that primates are undergoing to positive selection for most of the components of the Ccr4-Not complex between the analysed species. analysed In summary *P. anubis* CNOT1 diverged from the other primates' proteins; CNOT2 had positive selected sites in *P. troglodytes*, *P. anubis* and *G. gorilla gorilla*; the CNOT3 NOT box was found to have positively selected mutations in *M. mulatta*; CNOT6 was positively selected in *S. scrofa* and *G. gorilla gorilla*; CNOT8 had mutation selected positively in *N. leucogenys* and *P. abelii*; CNOT9, CNOT10, CNOT11 did not have any positively selected sites. Deadenylase subunits showed different evolutionary forces acting on the paralogues. CNOT6 resulted to be the more variable compared to CNOT6L which sequences, even if more distant from the ancestral sequence, are more conserved and homogeneous identifying purifying selection acting on this branches and positive selection acting on the CNOT6 branch. Moreover, the CNOT8 resulted to be more variable than CNOT7 which sequences were more conserved between orthologues. Positive selection was identified acting on the CNOT8 primate lineage only. Paralogues of deadenylase enzymes resulted to be subjected to different natural selection pressures and this may explain the redundancy of the proteins. We can speculate that one of the proteins is more conservative in its sequence (CNOT6L and CNOT7) compared to the other two (CNOT6 and CNOT8) which are under positive selective pressures often synonymous of gaining of new function.

## **Chapter 7**

### **Concluding Remarks and Future Outlook**

## **Chapter 7. Concluding Remarks and Future Outlook**

In this study bioinformatics and biochemical approaches were adopted to study and characterise the Ccr4-Not deadenylase complex. The combined *in silico* and *in vitro* approaches identified new biological pathways involving the recruitment of the Ccr4-Not complex by the study of the deadenylase subunits Ccr4 and Caf1 interactome (Chapter 3).

Bioinformatics analyses were subsequently used to identify amino acid residues important for protein-protein interactions of the Ccr4-Not complex components CNOT1, CNOT2, and CNOT3 (Chapter 4). Moreover, the study focussed on mutations found in cancer and identified variants of the deadenylase subunit CNOT7 whose ability to assemble in the nuclease sub-complex was disrupted (Chapter 5). Finally, a phylogenetic reconstruction and variability analyses of the components of the complex in 15 different mammalian species identified selective pressures acting among primates (Chapter 6).

### **7.1 The interactome of the Ccr4 and Caf1 subunit revealed by SILAC-RIME**

Degradation of mRNAs plays a pivotal role in the regulation of the gene expression. Deadenylation is in most cases the first step of cytoplasmic mRNA degradation. Two complexes are involved in the shortening of the poly(A) tail: the Pan2-Pan3 complex and the Ccr4-Not complex.

The Ccr4-Not complex is the major deadenylase enzyme involved in the shortening of the poly(A) tail and is composed by two catalytic subunits Ccr4 (CNOT6/CNOT6L) and Caf1 (CNOT7/CNOT8) (Wahle and Winkler, 2013). Regulatory proteins can directly or indirectly recruit the Ccr4-Not complex and trigger deadenylation of specific mRNA leading to mRNA degradation and down-regulation of protein expression (Goldstrohm et al., 2006, 2007; Sandler and Stoecklin, 2008; Joly et al., 2013; Raisch et al., 2016). The study of protein-protein interactions of the Pan2-Pan3 and

Ccr4-Not complexes has led to the characterisation of the deadenylation role in different regulatory pathways (Inada and Makino, 2014; Collart, 2016). To further investigate protein-protein interactions of the deadenylase complexes, Stable Isotope Labelling with Amino acids in Cell culture (SILAC)-Rapid Immunoprecipitations and Mass spectrometry of Endogenous protein (RIME) was used together with bioinformatic tools for the generation and exploration of protein-protein interaction networks (Chapter 3). This approach identified either direct and indirect protein-protein interactions with the Caf1 (CNOT7 and CNOT8) and Ccr4 (CNOT6 and CNOT6L) subunits. The SILAC-RIME data was further explored as a protein-protein network using Cytoscape and its plug-ins Bisogenet, MCODE and ClusterOne (Shannon et al., 2003; Martin et al., 2010; Bader et al., 2003; Zhang et al., 2014). The enrichment of the protein-protein interactions between the network components combined with the clustering analyses, identified important biological pathways that may involve the recruitment of the Ccr4-Not complex.

The clustering analysis identified clusters of known protein-protein interactions and new partners such as the mini-chromosome maintenance complex (MCM) components of the eukaryotic pre-replication complex, polypyrimidine binding proteins involved in pre-mRNA processing and YTH domain containing protein family members (YTHDF1-3) that specifically bind N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) and are involved in alternative splicing and mRNA stability (Wang et al., 2014).

This study explored the interactome of the deadenylase subunits of the Ccr4-Not complex emphasising new interaction partners and suggesting the recruitment of the complex in biological pathways not yet described. Interestingly, the data presented in this study suggest a recruitment of the Ccr4-NOT complex by the YTHDC1, YTHDF1, YTHDF2, and YTHDF3 proteins in agreement with the recent findings that demonstrated the recruitment of the Ccr4-NOT complex by YTHDF2 via CNOT1 (Du et al., 2016). Further approaches such as co-immunoprecipitation and yeast two hybrid analysis could be used to better characterise the recruitment of the

Ccr4-NOT complex by the YTH domain containing proteins and the other new partners identified by this study.

## **7.2 Bioinformatic approaches to predict residues critical for the assembly and function of the NOT-module**

The Ccr4-Not complex consists of two catalytic and six non-catalytic subunits. The largest subunit of the complex, CNOT1, acts as a central scaffold for the anchoring of several protein modules (Wahle and Winkler, 2013; Villanyi and Collart, 2015). The nuclease module, composed of Caf1 and Ccr4, binds the CNOT1 MIF4G domain, the CNOT10/CNOT11 module binds its N-terminal region of CNOT1, CNOT1 binds the central DUF3819 domain, and the NOT module, composed of CNOT2 and CNOT3, binds the C-terminal domain (Winkler and Balacco, 2013; Collart, 2016).

The NOT module components are involved in several cellular processes and their depletion may cause important changes in cell development and survival also phenotypically afflicting the organisms.

In Chapter 4, a bioinformatic approach were used to identify and characterise amino acid residues critical for the trimeric assembly and function of the NOT module. Variants of CNOT1, CNOT2 and CNOT3 were retrieved from the Catalogue Of Somatic Mutation In Cancer (COSMIC) database (Bamford et al., 2004; Forbes et al., 2008, 2011) and the Ensemble genome browser (Chen et al., 2010; Rios et al., 2010; Flicek et al., 2014). Prediction of their effects on protein function was performed using the Sorting Intolerant From Tolerant (SIFT) algorithm (Ng and Henikoff, 2003; Kumar et al., 2009). Surface characterisation of the NOT-module performed by Protein Interface, Surface, and Assembly (PISA) (Krissinel and Henrick, 2007) identified residues that participate in the assembly of the NOT-module by interacting with the other proteins. The interface between the NOT-module components was further investigated by constructing a residue-residue interaction network by RINalyzer and by the exploration of the resulting network by Cytoscape. These analyses supplied a general profile of the complex interacting residues allowing to predict which mutations that could affect the interactions between the hetero-trimeric components. Further studies are

required to validate the computational predictions. The interactions between the component of the NOT module may be studied using recombinant proteins and their mutants as described in Boland 2013 or by using isothermal titration calorimetry and surface plasmon resonance.

### **7.3 *In silico* and *in vitro* approaches identified CNOT7 variants unable to interact with CNOT6L, CNOT1, and BTG1.**

The nuclease module composed of the Caf1 homologues CNOT7 and CNOT8, and the Ccr4 homologues CNOT6 and CNOT6L is anchored on the central MIF4G domain of the CNOT1 subunit (Bai et al., 1999; Basquin et al., 2012; Petit et al., 2012). Only one homologue of each subunit assembles on the CNOT1 central scaffold. Moreover, the interaction between Ccr4 and Caf1 via the LRR domain of Ccr4 is essential for the stable recruitment and activity of Ccr4 to the complex (Draper et al., 1995; Dupressoir et al., 2001; Mittal et al., 2011; Basquin et al., 2012). Caf1 can also interact with BTG/TOB proteins, which act as regulator for the recruitment of the Ccr4-Not complex involved in inhibition of the cell progression (Prévôt et al., 2001; Winkler, 2010).

In Chapter 5, the analysis of sequence variants of the CNOT7 ribonuclease subunit of the Ccr4-Not complex was investigated to study the effect of mutations in the context of the sub-complex composed of the CNOT1 MIF4G domain, BTG1, and CNOT6L. The approach used combined *in silico* and *in vitro* methods to identify CNOT7 variants with impaired function and defective interactions with the other components of the nuclease sub-complex. Variants of CNOT7 were retrieved from the COSMIC database and the Ensemble genome browser. The impact of each amino acid substitution on protein function was predicted using SIFT algorithm (Ng and Henikoff, 2003; Kumar et al., 2009) and the Disease-Susceptibility-based SAV Phenotype Prediction package (SuSPect) (Yates et al., 2014). Investigations of the impact of the amino acid substitutions on the stability and the structure of CNOT7 were carried out using Site Directed Mutator (SMD) and UCSF Chimera.

The impact of the CNOT7 variants predicted to be damaging for its function was further investigated by RNA tethering assay *in vitro*.

Interestingly, none of the analysed mutations resulted in altered function of CNOT7 except for the G163C variant that resulted to have reduced activity. Furthermore, the impact of the selected CNOT7 variants on the assembly components of the sub-complex composed by CNOT1, CNOT6L and BTG1 was carried out by yeast-two hybrid assay. This study identified CNOT7 variants that are unable to assemble in the sub-complex. The CNOT7 variant G45C resulted to abolish the interaction with CNOT6L. CNOT7 variants E217Q and E87Q specifically disrupted the interaction with BTG1. Variants R220W, L232S, and W153C resulted to disrupt the interaction with the central scaffold of the Ccr4-NOT complex CNOT1. The mutation G228E compromised the interaction with both BTG1 and CNOT6L, whereas variants E243G, G208V, N19S, and F239I abolished the interaction with both BTG1 and CNOT1. Interestingly the mutation N19I negatively affected the interaction with BTG1, CNOT1, and CNOT6L. The study identified variants that are not affected in their intrinsic deadenylase activity but affect the assembly of the CNOT7 subunit in the Ccr4-Not complex.

#### **7.4 Variability analyses identified different evolutionary rate on the orthologous protein of Ccr4 and Caf1**

Modern molecular evolution theory asserts that molecules undergo many variations such as deletions, insertions, inversions, or single nucleotide substitutions. These variants, if not deleterious, can be reproduced and amplified within a population (positive selection) and are likely to be linked to confer new functions to proteins (Hahn, 2008; Lemey et al., 2009).

In Chapter 6, I reconstructed the phylogeny of each component of the Ccr4-Not complex in 15 different mammalian species and identified selective pressures acting on its components. The study demonstrated that primates are undergoing positive selection for most of the components of the Ccr4-Not complex.

The *P. anubis* CNOT1 diverged from other proteins of primates (Figure 6.1); CNOT2 had positive selected sites in *P. troglodytes*, *P. anubis* and *G. gorilla gorilla* (Figure 6.3); the CNOT3 NOT box was found to have

positively selected mutations in *M. mulatta* (Figure 6.5); CNOT6 was positively selected in *S. scrofa* and *G. gorilla gorilla* (Figure 6.7); CNOT8 had mutation selected positively in *N. leucogenys* and *P. abelii* (Figure 6.9); CNOT9, CNOT10, CNOT11 did not have any positively selected sites (Figure 6.11-13). Interestingly, paralogues of deadenylase enzymes were found to be subjected to different natural selection pressures and this may explain the redundancy of the proteins. We can speculate that CNOT6L and CNOT7 are more conservative in its sequence compared to their paralogues CNOT6 and CNOT8 respectively and are under positive selective pressures often synonymous of gaining of new function.

## 7.5 Future Outlook

The work described in this thesis presents a dual approach to study the complexity and heterogeneity of the Ccr4-Not deadenylase complex. I used SILAC-RIME and Cytoscape to detect and explore the interactome of the deadenylase subunits of the complex identified new interacting partners (Chapter 3). A further *in vitro* investigation of the new interacting partner identified will confirm and describe the recruitment of the complex in new biological pathways. Moreover, the characterisation of the interactome of Pan2-Pan3, the other deadenylase complex involved in the deadenylation process, will clarify the role of the two deadenylase complexes and identify specific protein interacting with Pan2-Pan3.

Bioinformatic approaches were used to study the effect of mutation found in cancer of both the NOT-module components (Chapter 4) and the CNOT7 deadenylase subunit. Whereas the investigation of the NOT-module variants requires *in vitro* validation of the data, the CNOT7 analyses allowed the identification of variants found in cancer that affect the function and structure of the protein. CNOT7 variants were investigated by predicting the effects of single amino acid substitutions on the function, the structure, stability and interaction with CNOT6L, BTG1, and CNOT1. Studies of the deadenylase activity using a fluorescence-based assay could identify CNOT7 variants as potential therapeutic targets (Maryati et al., 2014). Moreover, phenotypic characterisation using cell based studies may clarify the effect of the identified mutations at the cellular level. The phylogenetic analyses identified a difference in the evolution rate of the CNOT7/CNOT8 and CNOT6/CNOT6L orthologues suggesting a gain of new function. Future studies could investigate the codons variability including more species to better understand the evolutionary pressures acting on the subunits of the Ccr4-Not complex.

## BIBLIOGRAPHY

- Adhikari, S, Xiao, W, Zhao, Y-L, Yang, Y-G. 2016. m(6)A: Signaling for mRNA splicing. *RNA Biol.* 13: 756–9.
- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19: 716–723.
- Albert, TK, Hanzawa, H, Legtenberg, YIA, de Ruwe, MJ, van den Heuvel, FAJ, Collart, MA, Boelens, R, Timmers, HTM. 2002. Identification of a ubiquitin-protein ligase subunit within the CCR4-NOT transcription repressor complex. *EMBO J.* 21: 355–64.
- Albert, TK, Lemaire, M, van Berkum, NL, Gentz, R, Collart, MA, Timmers, HT. 2000. Isolation and characterization of human orthologs of yeast CCR4-NOT complex subunits. *Nucleic Acids Res.* 28: 809–17.
- Altschul, SF. 1998. Generalized affine gap costs for protein sequence alignment. *Proteins* 32: 88–96.
- Amrani, N, Ganesan, R, Kervestin, S, Mangus, DA, Ghosh, S, Jacobson, A. 2004. A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* 432: 112–8.
- Andersen, KR, Jonstrup, AT, Van, L a NB. 2009. The activity and selectivity of fission yeast Pop2p are affected by a high affinity for Zn<sup>2+</sup> and Mn<sup>2+</sup> in the active site. 850–861.
- Arimbasseri, AG, Maraia, RJ. 2016. RNA Polymerase III Advances: Structural and tRNA Functional Views. *Trends Biochem. Sci.* 41: 546–59.
- Arribas-Layton, M, Wu, D, Lykke-Andersen, J, Song, H. 2013. Structural and functional control of the eukaryotic mRNA decapping machinery. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1829: 580–589.
- Aslam, A, Mittal, S, Koch, F, Andrau, J-C, Winkler, GS. 2009. The Ccr4-NOT deadenylase subunits CNOT7 and CNOT8 have overlapping roles and modulate cell proliferation. *Mol. Biol. Cell* 20: 3840–50.
- Azzouz, N, Panasencko, OO, Deluen, C, Hsieh, J, Theiler, G, Collart, MA. 2009. Specific roles for the Ccr4-Not complex subunits in expression of the genome. *RNA* 15: 377–83.
- Bader, GDG, Hogue, CWCCW V, Fields, S, Uetz, P, Giot, L, Cagney, G, Mansfield, T, Judson, R, Knight, J, Ito, T, Chiba, T, Ozawa, R, Yoshida, M, Hattori, M, Sakaki, Y, Drees, B, Sundin, B, Brazeau, E, Caviston, J, Chen, G, Guo, W, Fromont-Racine, M, Mayes, A, Brunet-Simon, A, Rain, J, Colley, A, Dix, I, Ho, Y, Gruhler, A, Heilbut, A, Bader, GDG, Moore, L, Adams, S, Gavin, A, Bosche, M, Krause, R, Grandi, P, Marzioch, M, Bauer, A, Christendat, D, Yee, A, Dharamsi, A, Kluger, Y, Savchenko, A, Cort, J, Kim, SS, Lund, J, Kiraly, M, Duke, K, Jiang, M, Stuart, J, Tong, A, Drees, B, Nardelli, G, Bader, GDG, Brannetti, B, Castagnoli, L, Winzeler, E, Shoemaker, D, Astromoff, A, Liang, H, Anderson, K, Andre, B, Chervitz, S, Hester, E, Ball, C, Dolinski, K, Dwight, S, Harris, M, Mewes, H, Frishman, D, Gruber, C, Geier, B, Haase, D, Kaps, A, Costanzo, M, Crawford, M, Hirschman, J, Kranz, J, Olsen, P, Robertson, L, Bader, GDG, Donaldson, I, Wolting, C, Ouellette, B, Pawson, T, Hogue, CWCCW

- V, Xenarios, I, Salwinski, L, Duan, X, Higney, P, Kim, SS, Eisenberg, D, Takai-Igarashi, T, Nadaoka, Y, Kaminuma, T, Wingender, E, Chen, X, et al. 2003. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4: 2.
- Bai, Y, Salvatore, C, Chiang, YC, Collart, MA, Liu, HY, Denis, CL. 1999. The CCR4 and CAF1 proteins of the CCR4-NOT complex are physically and functionally separated from NOT2, NOT4, and NOT5. *Mol. Cell. Biol.* 19: 6642–51.
- Baines, RA. 2005. Neuronal homeostasis through translational control. *Mol. Neurobiol.* 32: 113–21.
- Bamford, S, Dawson, E, Forbes, S, Clements, J, Pettett, R, Dogan, A, Flanagan, A, Teague, J, Futreal, PA, Stratton, MR, Wooster, R. 2004. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* 91: 355–8.
- Barreau, C, Paillard, L, Osborne, HB. 2005. AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res.* 33: 7138–50.
- Bartlam, M, Yamamoto, T. 2010. The structural basis for deadenylation by the CCR4-NOT complex. *Protein Cell* 1: 443–52.
- Basquin, J, Roudko, V V., Rode, M, Basquin, C, Séraphin, B, Conti, E. 2012. Architecture of the nuclease module of the yeast ccr4-Not complex: The not1-caf1-ccr4 interaction. *Mol. Cell* 48: 207–218.
- Bawankar, P, Loh, B, Wohlbold, L, Schmidt, S, Izaurralde, E. 2013. NOT10 and C2orf29/NOT11 form a conserved module of the CCR4-NOT complex that docks onto the NOT1 N-terminal domain. *RNA Biol.* 10: 228–44.
- Behm-Ansmant, I, Rehwinkel, J, Doerks, T, Stark, A, Bork, P, Izaurralde, E. 2006. mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev.* 20: 1885–98.
- Bell, SD, Botchan, MR. 2013. The minichromosome maintenance replicative helicase. *Cold Spring Harb. Perspect. Biol.* 5: a012807.
- Bentley, DL. 2005. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr. Opin. Cell Biol.* 17: 251–6.
- Berndt, H, Harnisch, C, Rammelt, C, Stöhr, N, Zirkel, A, Dohm, JC, Himmelbauer, H, Tavanez, J-P, Hüttelmaier, S, Wahle, E. 2012. Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA* 18: 958–72.
- Bessman, MJ, Frick, DN, O’Handley, SF. 1996. The MutT proteins or “Nudix” hydrolases, a family of versatile, widely distributed, “housecleaning” enzymes. *J. Biol. Chem.* 271: 25059–62.
- Bhaskar, V, Basquin, J, Conti, E. 2015. Architecture of the ubiquitylation module of the yeast Ccr4-Not complex. *Structure* 23: 921–8.
- Bhaskar, V, Roudko, V, Basquin, J, Sharma, K, Urlaub, H, Séraphin, B, Conti, E. 2013. Structure and RNA-binding properties of the Not1-Not2-Not5 module of the yeast Ccr4-Not complex. *Nat. Struct. Mol. Biol.* 20: 1281–8.
- Bhuvanagiri, M, Schlitter, AM, Hentze, MW, Kulozik, AE. 2010. NMD: RNA biology meets human genetic medicine. *Biochem. J.* 430.

- Bianchin, C, Mauxion, F, Sentis, S, Séraphin, B, Corbo, L. 2005. Conservation of the deadenylase activity of proteins of the Caf1 family in human. *RNA* 11: 487–494.
- Boeck, R, Tarun, S, Rieger, M, Deardorff, JA, Mu, S, Sachs, AB. 1996a. The Yeast Pan2 Protein Is Required for Poly ( A ) -binding Protein-stimulated Poly ( A ) -nuclease Activity \*. *J. Biol. Chem.* 271: 432–438.
- Boeck, R, Tarun, S, Rieger, M, Deardorff, JA, Müller-Auer, S, Sachs, AB. 1996b. The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. *J. Biol. Chem.* 271: 432–8.
- Boiko, AD, Porteous, S, Razorenova, O V, Krivokrysenko, VI, Williams, BR, Gudkov, A V. 2006. A systematic search for downstream mediators of tumor suppressor function of p53 reveals a major role of BTG2 in suppression of Ras-induced transformation. *Genes Dev.* 20: 236–52.
- Boland, A, Chen, Y, Raisch, T, Jonas, S, Kuzuoğlu-Öztürk, D, Wohlbold, L, Weichenrieder, O, Izaurralde, E. 2013. Structure and assembly of the NOT module of the human CCR4-NOT complex. *Nat. Struct. Mol. Biol.* 20: 1289–97.
- Boratyn, GM, Schäffer, AA, Agarwala, R, Altschul, SF, Lipman, DJ, Madden, TL. 2012. Domain enhanced lookup time accelerated BLAST. *Biol. Direct* 7: 12.
- Boulon, S, Bertrand, E, Pradet-Balade, B. 2012. HSP90 and the R2TP co-chaperone complex: Building multi-protein machineries essential for cell growth and gene expression. *RNA Biol.* 9: 148–154.
- Braun, KA, Young, ET. 2014. Coupling mRNA synthesis and decay. *Mol. Cell Biol.* 34: 4078–87.
- Brooks, SA, Blackshear, PJ. Tristetraprolin (TTP): interactions with mRNA and proteins, and current thoughts on mechanisms of action. *Biochim. Biophys. Acta* 1829: 666–79.
- Brown, CE, Sachs, AB. 1998. Poly(A) tail length control in *Saccharomyces cerevisiae* occurs by message-specific deadenylation. *Mol. Cell Biol.* 18: 6548–59.
- Brown, CE, Tarun, SZ, Boeck, R, Sachs, AB. 1996. PAN3 encodes a subunit of the Pab1p-dependent poly(A) nuclease in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* 16: 5744–53.
- Buratowski, S. 2009. Progression through the RNA polymerase II CTD cycle. *Mol. Cell* 36: 541–6.
- Bushati, N, Cohen, SM. 2007. microRNA functions. *Annu. Rev. Cell Dev. Biol.* 23: 175–205.
- Carmody, SR, Wenthe, SR. 2009. mRNA nuclear export at a glance. *J. Cell Sci.* 122: 1933–7.
- Catalanotto, C, Cogoni, C, Zardo, G. 2016. MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions. *Int. J. Mol. Sci.* 17: 1712.
- Cevher, MA, Zhang, X, Fernandez, S, Kim, S, Baquero, J, Nilsson, P, Lee, S, Virtanen, A, Kleiman, FE. 2010. Nuclear deadenylation/polyadenylation factors regulate 3' processing in response to DNA damage. *EMBO J.* 29: 1674–87.
- Chan, S, Choi, E-A, Shi, Y. 2011. Pre-mRNA 3'-end processing complex

- assembly and function. *Wiley Interdiscip. Rev. RNA* 2: 321–35.
- Chang, H, Lim, J, Ha, M, Kim, VN. 2014. TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol. Cell* 53: 1044–52.
- Chekulaeva, M, Mathys, H, Zipprich, JT, Attig, J, Colic, M, Parker, R, Filipowicz, W. 2011. miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. *Nat. Struct. Mol. Biol.* 18: 1218–26.
- Chen, C-YA, Shyu, A-B. 2011. Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip. Rev. RNA* 2: 167–83.
- Chen, CA, Ezzeddine, N, Shyu, A. 2008. Chapter 17 Messenger RNA Half-Life Measurements in Mammalian Cells. *Methods Enzymol.* 448: 335–357.
- Chen, CY, Shyu, AB. 1994. Selective degradation of early-response-gene mRNAs: functional analyses of sequence features of the AU-rich elements. *Mol. Cell. Biol.* 14: 8471–82.
- Chen, CY, Xu, N, Shyu, AB. 1995. mRNA decay mediated by two distinct AU-rich elements from c-fos and granulocyte-macrophage colony-stimulating factor transcripts: different deadenylation kinetics and uncoupling from translation. *Mol. Cell. Biol.* 15: 5777–88.
- Chen, J, Chiang, YC, Denis, CL. 2002. CCR4, a 3'-5' poly(A) RNA and ssDNA exonuclease, is the catalytic component of the cytoplasmic deadenylase. *EMBO J.* 21: 1414–1426.
- Chen, J, Rappsilber, J, Chiang, YC, Russell, P, Mann, M, Denis, CL. 2001. Purification and characterization of the 1.0 MDa CCR4-NOT complex identifies two novel components of the complex. *J. Mol. Biol.* 314: 683–94.
- Chen, Y, Boland, A, Kuzuoğlu-Öztürk, D, Bawankar, P, Loh, B, Chang, C-T, Weichenrieder, O, Izaurralde, E. 2014. A DDX6-CNOT1 complex and W-binding pockets in CNOT9 reveal direct links between miRNA target recognition and silencing. *Mol. Cell* 54: 737–50.
- Chen, Y, Cunningham, F, Rios, D, McLaren, WM, Smith, J, Pritchard, B, Spudich, GM, Brent, S, Kulesha, E, Marin-Garcia, P, Smedley, D, Birney, E, Flicek, P. 2010. Ensembl variation resources. *BMC Genomics* 11: 293.
- Chendrimada, TP, Gregory, RI, Kumaraswamy, E, Norman, J, Cooch, N, Nishikura, K, Shiekhattar, R. 2005. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* 436: 740–4.
- Cheung, ACM, Cramer, P. 2012. A movie of RNA polymerase II transcription. *Cell* 149: 1431–7.
- Choi, HS, Hwang, CK, Song, KY, Law, P-Y, Wei, L-N, Loh, HH. 2009. Poly(C)-binding proteins as transcriptional regulators of gene expression. *Biochem. Biophys. Res. Commun.* 380: 431–6.
- Chou, W-L, Huang, L-F, Fang, J-C, Yeh, C-H, Hong, C-Y, Wu, S-J, Lu, C-A. 2014. Divergence of the expression and subcellular localization of CCR4-associated factor 1 (CAF1) deadenylase proteins in *Oryza sativa*. *Plant Mol. Biol.* 85: 443–58.
- Chowdhury, A, Mukhopadhyay, J, Tharun, S. 2007. The decapping activator

- Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. *RNA* 13: 998–1016.
- Christie, M, Boland, A, Huntzinger, E, Weichenrieder, O, Izaurralde, E, Boeck, R, Tarun, S, Rieger, M, Deardorff, JA, Müller-Auer, S, Sachs, AB, Braun, JE, Huntzinger, E, Fauser, M, Izaurralde, E, Brown, CE, Tarun, SZ, Boeck, R, Sachs, AB, Chekulaeva, M, Mathys, H, Zipprich, JT, Attig, J, Colic, M, Parker, R, Filipowicz, W, Fabian, MR, Cieplak, MK, Frank, F, Morita, M, Green, J, Srikumar, T, Nagar, B, Yamamoto, T, Raught, B, Duchaine, TF, Sonenberg, N, Hammet, A, Pike, BL, Heierhorst, J, Huntzinger, E, Kuzuoglu-Öztürk, D, Braun, JE, Eulalio, A, Wohlbold, L, Izaurralde, E, Huse, M, Kuriyan, J, Jinek, M, Fabian, MR, Coyle, SM, Sonenberg, N, Doudna, JA, Kozlov, G, Safaee, N, Rosenauer, A, Gehring, K, Lowell, JE, Rudner, DZ, Sachs, AB, Mangus, DA, Evans, MC, Agrin, NS, Smith, M, Gongidi, P, Jacobson, A, Min, X, Lee, BH, Cobb, MH, Goldsmith, EJ, Munchel, SE, Shultzaberger, RK, Takizawa, N, Weis, K, Schirle, NT, MacRae, IJ, Siddiqui, N, Mangus, DA, Chang, TC, Palermino, JM, Shyu, AB, Gehring, K, Taylor, SS, Kornev, AP, Tompa, P, Tucker, M, Valencia-Sanchez, MA, Staples, RR, Chen, J, Denis, CL, Parker, R, Uchida, N, Hoshino, S, Katada, T, Wahle, E, Winkler, GS, Weill, L, Belloc, E, et al. 2013. Structure of the PAN3 Pseudokinase Reveals the Basis for Interactions with the PAN2 Deadenylation and the GW182 Proteins. *Mol. Cell* 51: 360–373.
- Clark, LB, Viswanathan, P, Quigley, G, Chiang, YC, McMahon, JS, Yao, G, Chen, J, Nelsbach, A, Denis, CL. 2004. Systematic Mutagenesis of the Leucine-rich Repeat (LRR) Domain of CCR4 Reveals Specific Sites for Binding to CAF1 and A Separate Critical Role for the LRR in CCR4 Deadenylation Activity. *J. Biol. Chem.* 279: 13616–13623.
- Colgan, DF, Manley, JL. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* 11: 2755–66.
- Collart, MA. 2016. The Ccr4-Not complex is a key regulator of eukaryotic gene expression. *Wiley Interdiscip. Rev. RNA* 7: 438–54.
- Collart, MA, Panasenko, OO. 2012. The Ccr4-Not complex. *Gene* 492: 42–53.
- Collart, MA, Struhl, K. 1994. NOT1(CDC39), NOT2(CDC36), NOT3, and NOT4 encode a global-negative regulator of transcription that differentially affects TATA-element utilization. *Genes Dev.* 8: 525–537.
- Cooke, A, Prigge, A, Wickens, M. 2010. Translational repression by deadenylases. *J. Biol. Chem.* 285: 28506–28513.
- Couttet, P, Fromont-Racine, M, Steel, D, Pictet, R, Grange, T. 1997. Messenger RNA deadenylation precedes decapping in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 94: 5628–33.
- Cowling, VH. 2009. Regulation of mRNA cap methylation. *Biochem. J.* 425: 295–302.
- Culjkovic-Kraljacic, B, Borden, KLB. 2013. Aiding and abetting cancer: mRNA export and the nuclear pore. *Trends Cell Biol.* 23: 328–35.
- Cullen, BR. 2003. Nuclear mRNA export: insights from virology. *Trends Biochem. Sci.* 28: 419–24.

- Darriba, D, Taboada, GL, Doallo, R, Posada, D. 2015. Europe PMC Funders Group jModelTest 2 : more models , new heuristics and high-performance computing. *Nat. Methods* 9: 6–9.
- Daxinger, L, Kanno, T, Bucher, E, van der Winden, J, Naumann, U, Matzke, AJM, Matzke, M. 2009. A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. *EMBO J.* 28: 48–57.
- Decker, CJ, Parker, R. 1993. A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes Dev.* 7: 1632–43.
- Dehlin, E, Wormington, M, Körner, CG, Wahle, E, Astrom, J, Astrom, A, Virtanen, A, Astrom, J, Astrom, A, Virtanen, A, Audic, Y, Omilli, F, Osborne, H, Bashkirov, V, Scherthan, H, Solinger, J, Buerstedde, J, Heyer, W, Beelman, C, Parker, R, Boeck, R, Tarun, S, Rieger, M, Deardorff, J, Müller-Auer, S, Sachs, A, Bouvet, P, Omilli, F, Arlot-Bonnemains, Y, Legagneux, V, Roghi, C, Bassez, T, Osborne, H, Brown, C, Sachs, A, Brown, C, Tarun, S, Boeck, R, Sachs, A, Buiting, K, Körner, C, Ulrich, B, Wahle, E, Horsthemke, B, Caponigro, G, Parker, R, Caponigro, G, Parker, R, Chen, C, Shyu, A, Chernokalskaya, E, Dubell, A, Cunningham, K, Hanson, M, Dompenciel, R, Schoenberg, D, Coller, J, Gray, N, Wickens, M, Couttet, P, Fromont-Racine, M, Steel, D, Pictet, R, Grange, T, Gregorio, E De, Preiss, T, Hentze, M, Drummond, D, Armstrong, J, Colman, A, Fox, C, Wickens, M, Gillian-Daniel, D, Gray, N, Aström, J, Barkoff, A, Wickens, M, Gingras, A, Raught, B, Sonenberg, N, Haghigat, A, Sonenberg, N, Imataka, H, Gradi, A, Sonenberg, N, Izaurralde, E, Lewis, J, McGuigan, C, Jankowska, M, Darzynkiewicz, E, Mattaj, I, Anderson, JJ, Parker, R, Jacobson, A, Peltz, S, Körner, C, Wahle, E, Körner, C, et al. 2000. Cap-dependent deadenylation of mRNA. *EMBO J.* 19: 1079–86.
- Denis, CL. 1984. Identification of new genes involved in the regulation of yeast alcohol dehydrogenase II. *Genetics* 108: 833–844.
- van Dijk, E, Cougot, N, Meyer, S, Babajko, S, Wahle, E, Séraphin, B, Bai, R, Koester, C, Ouyang, T, Hahn, S, Hammerschmidt, M, Peschel, C, Duyster, J, Bashkirov, V, Scherthan, H, Solinger, J, Buerstedde, J, Heyer, W, Beelman, C, Stevens, A, Caponigro, G, LaGrandeur, T, Hatfield, L, Fortner, D, Parker, R, Bessman, M, Frick, D, O’Handley, S, Bousquet-Antonelli, C, Presutti, C, Tollervey, D, Bouveret, E, Rigaut, G, Shevchenko, A, Wilm, M, Séraphin, B, Brogna, S, Sato, T, Rosbash, M, Brown, J, Johnson, A, Butler, J, Caponigro, G, Parker, R, Chen, C, Couttet, P, Fromont-Racine, M, Steel, D, Pictet, R, Grange, T, Czaplinski, K, Ruiz-Echevarria, M, Gonzalez, C, Peltz, S, Daugeron, M, Mauxion, F, Séraphin, B, Dunckley, T, Parker, R, Eystathioy, T, Chan, E, Tenenbaum, S, Keene, J, Griffith, K, Fritzler, M, Fischer, N, Weis, K, Frischmeyer, P, Hoof, A van, O’Donnell, K, Guerrerio, A, Parker, R, Dietz, H, Gao, M, Wilusz, C, Peltz, S, Wilusz, J, He, F, Jacobson, A, He, F, Peltz, S, Donahue, J, Rosbash, M, Jacobson, A, Higman, M, Christen, L, Niles, E, Iborra, F, Jackson, D, Cook, P, Ishigaki, Y, Li, X, Serin, G, Maquat, L, LaGrandeur, T, Parker, R, Larimer, F, Hsu, C, et al. 2002. Human Dcp2: a catalytically active mRNA decapping enzyme located

- in specific cytoplasmic structures. *EMBO J.* 21: 6915–24.
- Dimitrova, LN, Kuroha, K, Tatematsu, T, Inada, T. 2009. Nascent peptide-dependent translation arrest leads to Not4p-mediated protein degradation by the proteasome. *J. Biol. Chem.* 284: 10343–52.
- Dlakić, M. 2000. Functionally unrelated signalling proteins contain a fold similar to Mg<sup>2+</sup>-dependent endonucleases. *Trends Biochem. Sci.* 25: 272–3.
- Doidge, R, Mittal, S, Aslam, A, Winkler, GS. 2012. The Anti-Proliferative Activity of BTG/TOB Proteins Is Mediated via the Caf1a (CNOT7) and Caf1b (CNOT8) Deadenylation Subunits of the Ccr4-Not Complex. *PLoS One* 7.
- Doma, MK, Parker, R. 2006. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* 440: 561–4.
- Doma, MK, Parker, R. 2007. RNA quality control in eukaryotes. *Cell* 131: 660–8.
- Doncheva, NT, Klein, K, Domingues, FS, Albrecht, M, O'Donoghue, SI, al., et, Morris, JH, al., et, Csermely, P, Vishveshwara, S, al., et, Vendruscolo, M, al., et, Krishnan, A, al., et, Swint-Kruse, L, Swint-Kruse, L, al., et, Amitai, G, al., et, Süel, GM, al., et, Sol, A del, al., et, Reichmann, D, al., et, Sathyapriya, R, al., et, Cheng, TM, al., et, Welsch, C, al., et, Susser, P, al., et, Prongay, AJ, al., et, Zhou, P, al., et, Shannon, P, al., et, Morris, JH, al., et, Pettersen, EF, al., et, Zhou, Y, al., et, Yoon, SI, al., et, Jones, BC, al., et, Word, JM, al., et. 2011. Analyzing and visualizing residue networks of protein structures. *Trends Biochem. Sci.* 36: 179–82.
- Draper, MP, Salvatore, C, Denis, CL. 1995. Identification of a mouse protein whose homolog in *Saccharomyces cerevisiae* is a component of the CCR4 transcriptional regulatory complex. *Mol. Cell. Biol.* 15: 3487–3495.
- Du, H, Zhao, Y, He, J, Zhang, Y, Xi, H, Liu, M, Ma, J, Wu, L. 2016. YTHDF2 destabilizes m(6)A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nat. Commun.* 7: 12626.
- Dubois, M-L, Bastin, C, Lévesque, D, Boisvert, F-M. 2016. Comprehensive Characterization of Minichromosome Maintenance Complex (MCM) Protein Interactions Using Affinity and Proximity Purifications Coupled to Mass Spectrometry. *J. Proteome Res.* 15: 2924–2934.
- Dunbrack, RL. 2002. Rotamer libraries in the 21st century. *Curr. Opin. Struct. Biol.* 12: 431–40.
- Dunckley, T, Parker, R. 1999. The DCP2 protein is required for mRNA decapping in *Saccharomyces cerevisiae* and contains a functional MutT motif. *EMBO J.* 18: 5411–22.
- Dupressoir, A, Morel, A-P, Barbot, W, Loireau, M-P, Corbo, L, Heidmann, T. 2001. Identification of four families of yCCR4- and Mg<sup>2+</sup>-dependent endonuclease-related proteins in higher eukaryotes, and characterization of orthologs of yCCR4 with a conserved leucine-rich repeat essential for hCAF1/hPOP2 binding. *BMC Genomics* 2: 9.
- Eberle, AB, Stalder, L, Mathys, H, Orozco, RZ, Mühlemann, O. 2008. Posttranscriptional gene regulation by spatial rearrangement of the 3' untranslated region. *PLoS Biol.* 6: e92.
- Eckmann, CR, Rammelt, C, Wahle, E. 2011. Control of poly(A) tail length.

- Wiley Interdiscip. Rev. RNA 2: 348–61.
- Eick, D, Geyer, M. 2013. The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. *Chem. Rev.* 113: 8456–8490.
- Eisenberg-Lerner, A, Ciechanover, A, Merbl, Y. 2016. Post-translational modification profiling - A novel tool for mapping the protein modification landscape in cancer. *Exp. Biol. Med.* 241: 1475–1482.
- Eulalio, A, Huntzinger, E, Izaurralde, E. 2008. GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nat. Struct. Mol. Biol.* 15: 346–353.
- Eulalio, A, Tritschler, F, Büttner, R, Weichenrieder, O, Izaurralde, E, Truffault, V. 2009. The RRM domain in GW182 proteins contributes to miRNA-mediated gene silencing. *Nucleic Acids Res.* 37: 2974–83.
- Ezzeddine, N, Chang, T-C, Zhu, W, Yamashita, A, Chen, C-YA, Zhong, Z, Yamashita, Y, Zheng, D, Shyu, A-B. 2007. Human TOB, an antiproliferative transcription factor, is a poly(A)-binding protein-dependent positive regulator of cytoplasmic mRNA deadenylation. *Mol. Cell. Biol.* 27: 7791–801.
- Ezzeddine, N, Chen, C-YA, Shyu, A-B. 2012. Evidence providing new insights into TOB-promoted deadenylation and supporting a link between TOB's deadenylation-enhancing and antiproliferative activities. *Mol. Cell. Biol.* 32: 1089–98.
- Fabian, MR, Cieplak, MK, Frank, F, Morita, M, Green, J, Srikumar, T, Nagar, B, Yamamoto, T, Raught, B, Duchaine, TF, Sonenberg, N. 2012. miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4–NOT. *Nat. Struct. Mol. Biol.* 19: 1211–1217.
- Fabian, MR, Frank, F, Rouya, C, Siddiqui, N, Lai, WS, Karetnikov, A, Blackshear, PJ, Nagar, B, Sonenberg, N. 2013. Structural basis for the recruitment of the human CCR4–NOT deadenylase complex by tristetraprolin. *Nat. Struct. Mol. Biol.* 20: 735–9.
- Fang, L, Wang, X, Yamoah, K, Chen, P, Pan, Z-Q, Huang, L. 2008. Characterization of the Human COP9 Signalosome Complex Using Affinity Purification and Mass Spectrometry. *J. Proteome Res.* 7: 4914–4925.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17: 368–76.
- Fenger-Grøn, M, Fillman, C, Norrild, B, Lykke-Andersen, J. 2005. Multiple processing body factors and the ARE binding protein TTP activate mRNA decapping. *Mol. Cell* 20: 905–15.
- Fischer, N, Weis, K, Akao, Y, Marukawa, O, Morikawa, H, Nakao, K, Kamei, M, Hachiya, T, Tsujimoto, Y, Ayscough, K, Drubin, D, Beelman, C, Stevens, A, Caponigro, G, LaGrandeur, T, Hatfield, L, Fortner, D, Parker, R, Bender, A, Pringle, J, Boeck, R, Lapeyre, B, Brown, C, Sachs, A, Bonnerot, C, Boeck, R, Lapeyre, B, Bouveret, E, Rigaut, G, Shevchenko, A, Wilm, M, Seraphin, B, Chang, M, French-Cornay, D, Fan, H, Klein, H, Denis, C, Jaehning, J, Coller, J, Tucker, M, Sheth, U, Valencia-Sanchez, M, Parker, R, Czapinski, K, Ruiz-Echevarria, M, Gonzalez, C, Peltz, S, Dageron, M, Mauxion, F, Seraphin, B, Decker, C, Parker, R, Cruz, J de la, Kressler, D, Linder, P, Valoir, T de, Tucker, M, Belikoff, E, Camp, L,

- Bolduc, C, Beckingham, K, Dunckley, T, Parker, R, Dunckley, T, Tucker, M, Parker, R, Gallie, D, Guthrie, C, Fink, G, Hata, H, Mitsui, H, Liu, H, Bai, Y, Denis, C, Shimizu, Y, Sakai, A, Hatfield, L, Beelman, C, Stevens, A, Parker, R, He, W, Parker, R, He, F, Peltz, S, Donahue, J, Rosbash, M, Jacobson, A, Hentze, M, Kulozik, A, Hilleren, P, Parker, R, Ho, J, Kallstrom, G, Johnson, A, Hsu, C, Stevens, A, Jacobs, J, Anderson, A, et al. 2002. The DEAD box protein Dhh1 stimulates the decapping enzyme Dcp1. *EMBO J.* 21: 2788–97.
- Flicek, P, Amode, MR, Barrell, D, Beal, K, Billis, K, Brent, S, Carvalho-Silva, D, Clapham, P, Coates, G, Fitzgerald, S, Gil, L, Girón, CG, Gordon, L, Hourlier, T, Hunt, S, Johnson, N, Juettemann, T, Kähäri, AK, Keenan, S, Kulesha, E, Martin, FJ, Maurel, T, McLaren, WM, Murphy, DN, Nag, R, Overduin, B, Pignatelli, M, Pritchard, B, Pritchard, E, Riat, HS, Ruffier, M, Sheppard, D, Taylor, K, Thormann, A, Trevanion, SJ, Vullo, A, Wilder, SP, Wilson, M, Zadissa, A, Aken, BL, Birney, E, Cunningham, F, Harrow, J, Herrero, J, Hubbard, TJP, Kinsella, R, Muffato, M, Parker, A, Spudich, G, Yates, A, Zerbino, DR, Searle, SMJ. 2014. Ensembl 2014. *Nucleic Acids Res.* 42: D749-55.
- Floch, AG, Palancade, B, Doye, V. 2014. Fifty years of nuclear pores and nucleocytoplasmic transport studies: multiple tools revealing complex rules. *Methods Cell Biol.* 122: 1–40.
- Folco, EG, Lee, C-S, Dufu, K, Yamazaki, T, Reed, R. 2012. The proteins PDIP3 and ZC11A associate with the human TREX complex in an ATP-dependent manner and function in mRNA export. *PLoS One* 7: e43804.
- Forbes, SA, Bhamra, G, Bamford, S, Dawson, E, Kok, C, Clements, J, Menzies, A, Teague, JW, Futreal, PA, Stratton, MR. 2008. The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr. Protoc. Hum. Genet.* Chapter 10: Unit 10.11.
- Forbes, SA, Bindal, N, Bamford, S, Cole, C, Kok, CY, Beare, D, Jia, M, Shepherd, R, Leung, K, Menzies, A, Teague, JW, Campbell, PJ, Stratton, MR, Futreal, PA. 2011. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* 39: D945-50.
- Foster, MW, Thompson, JW, Forrester, MT, Sha, Y, McMahon, TJ, Bowles, DE, Moseley, MA, Marshall, HE. 2013. Proteomic analysis of the NOS2 interactome in human airway epithelial cells. *Nitric Oxide* 34: 37–46.
- Fromm, SA, Truffault, V, Kamenz, J, Braun, JE, Hoffmann, NA, Izaurralde, E, Sprangers, R, Albrecht, M, Lengauer, T, Badis, G, Saveanu, C, Fromont-Racine, M, Jacquier, A, Bahler, J, Wu, J, Longtine, M, Shah, N, McKenzie, A, Steever, A, Wach, A, Philippsen, P, Pringle, J, Boeck, R, Lapeyre, B, Brown, C, Sachs, A, Bonnerot, C, Boeck, R, Lapeyre, B, Borja, M, Piotukh, K, Freund, C, Gross, J, Bouveret, E, Rigaut, G, Shevchenko, A, Wilm, M, Seraphin, B, Callebaut, I, Coller, J, Parker, R, Cornilescu, G, Delaglio, F, Bax, A, Decker, C, Parker, R, Decker, C, Teixeira, D, Parker, R, Decourty, L, Saveanu, C, Zemam, K, Hantraye, F, Frachon, E, Rousselle, J, Fromont-Racine, M, Jacquier, A, Eulalio, A, Behm-Ansmant, I, Izaurralde, E, Fenger-Gron, M, Fillman, C, Norrild, B, Lykke-Andersen, J, Fischer, N, Weis, K, Floor, S, Jones, B, Hernandez, G,

- Gross, J, Franks, T, Lykke-Andersen, J, Fromont-Racine, M, Mayes, A, Brunet-Simon, A, Rain, J, Colley, A, Dix, I, Decourty, L, Joly, N, Ricard, F, Beggs, J, Legrain, P, Gaudon, C, Chambon, P, Losson, R, Haas, G, Braun, J, Igreja, C, Tritschler, F, Nishihara, T, Izaurralde, E, Harigaya, Y, Jones, B, Muhrad, D, Gross, J, Parker, R, Ito, T, et al. 2012. The structural basis of Edc3- and Scd6-mediated activation of the Dcp1:Dcp2 mRNA decapping complex. *EMBO J.* 31: 279–90.
- Fukuda, T, Naiki, T, Saito, M, Irie, K. 2009. hnRNP K interacts with RNA binding motif protein 42 and functions in the maintenance of cellular ATP level during stress conditions. *Genes Cells* 14: 113–28.
- Garces, RG, Gillon, W, Pai, EF. 2007. Atomic model of human Rcd-1 reveals an armadillo -like-repeat protein with in vitro nucleic acid binding properties. *Protein Sci.* 16: 176–188.
- Garneau, NL, Wilusz, J, Wilusz, CJ. 2007. The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* 8: 113–126.
- Girard, C, Will, CL, Peng, J, Makarov, EM, Kastner, B, Lemm, I, Urlaub, H, Hartmuth, K, Lührmann, R. 2012. Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat. Commun.* 3: 994.
- Godwin, AR, Kojima, S, Green, CB, Wilusz, J. 2013. Kiss your tail goodbye: The role of PARN, Nocturnin, and Angel deadenylases in mRNA biology. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1829: 571–579.
- Goldstrohm, AC, Hook, BA, Seay, DJ, Wickens, M. 2006. PUF proteins bind Pop2p to regulate messenger RNAs. *Nat. Struct. Mol. Biol.* 13: 533–539.
- Goldstrohm, AC, Seay, DJ, Hook, BA, Wickens, M. 2007. PUF protein-mediated deadenylation is catalyzed by Ccr4p. *J. Biol. Chem.* 282: 109–14.
- Goldstrohm, AC, Wickens, M. 2008. Multifunctional deadenylase complexes diversify mRNA control. *Nat. Rev. Mol. Cell Biol.* 9: 337–344.
- Gouy, M, Guindon, S, Gascuel, O. 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27: 221–4.
- Graille, M, Chaillet, M, van Tilbeurgh, H. 2008. Structure of yeast Dom34: a protein related to translation termination factor Erf1 and involved in No-Go decay. *J. Biol. Chem.* 283: 7145–54.
- Guindon, S, Dufayard, JF, Lefort, V, Anisimova, M, Hordijk, W, Gascuel, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* 59: 307–321.
- Gulshan, K, Thommandru, B, Moye-Rowley, WS. 2012. Proteolytic degradation of the Yap1 transcription factor is regulated by subcellular localization and the E3 ubiquitin ligase Not4. *J. Biol. Chem.* 287: 26796–805.
- Gupta, K, Sari-Ak, D, Haffke, M, Trowitzsch, S, Berger, I. 2016. Zooming in on Transcription Preinitiation. *J. Mol. Biol.* 428: 2581–91.
- Hahn, MW. 2008. Toward a selection theory of molecular evolution. *Evolution (N. Y.)*. 62: 255–265.

- Halees, AS, El-Badrawi, R, Khabar, KSA. 2008. ARED Organism: expansion of ARED reveals AU-rich element cluster variations between human and mouse. *Nucleic Acids Res.* 36: D137-40.
- Hamid, FM, Makeyev, E V. 2016. Exaptive origins of regulated mRNA decay in eukaryotes. *Bioessays* 38: 830-8.
- Hammet, A, Pike, BL, Heierhorst, J. 2002. Posttranscriptional regulation of the RAD5 DNA repair gene by the Dun1 kinase and the Pan2-Pan3 poly(A)-nuclease complex contributes to survival of replication blocks. *J. Biol. Chem.* 277: 22469-74.
- Hautbergue, GM, Hung, M-L, Golovanov, AP, Lian, L-Y, Wilson, SA. 2008. Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP. *Proc. Natl. Acad. Sci. U. S. A.* 105: 5154-9.
- He, G-J, Liu, W-F, Yan, Y-B. 2011. Dissimilar roles of the four conserved acidic residues in the thermal stability of poly(A)-specific ribonuclease. *Int. J. Mol. Sci.* 12: 2901-16.
- Heath, CG, Viphakone, N, Wilson, SA. 2016. The role of TREX in gene expression and disease. *Biochem. J.* 473: 2911-35.
- van Hoof, A, Frischmeyer, PA, Dietz, HC, Parker, R. 2002. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* 295: 2262-4.
- Hook, BA, Goldstrohm, AC, Seay, DJ, Wickens, M. 2007. Two yeast PUF proteins negatively regulate a single mRNA. *J. Biol. Chem.* 282: 15430-8.
- Horikawa, W, Endo, K, Wada, M, Ito, K. 2016. Mutations in the G-domain of Ski7 cause specific dysfunction in non-stop decay. *Sci. Rep.* 6: 29295.
- Horiuchi, M, Takeuchi, K, Noda, N, Muroya, N, Suzuki, T, Nakamura, T, Kawamura-Tsuzuku, J, Takahasi, K, Yamamoto, T, Inagaki, F. 2009. Structural basis for the antiproliferative activity of the Tob-hCaf1 complex. *J. Biol. Chem.* 284: 13244-13255.
- Hosoda, N, Funakoshi, Y, Hirasawa, M, Yamagishi, R, Asano, Y, Miyagawa, R, Ogami, K, Tsujimoto, M, Hoshino, S. 2011. Anti-proliferative protein Tob negatively regulates CPEB3 target by recruiting Caf1 deadenylase. *EMBO J.* 30: 1311-1323.
- Hossain, MA, Johnson, TL. 2014. Using yeast genetics to study splicing mechanisms. *Methods Mol. Biol.* 1126: 285-98.
- Houseley, J, Tollervey, D. 2009. The Many Pathways of RNA Degradation. *Cell* 136: 763-776.
- Huang, K-L, Chadee, AB, Chen, C-YA, Zhang, Y, Shyu, A-B. 2013. Phosphorylation at intrinsically disordered regions of PAM2 motif-containing proteins modulates their interactions with PABPC1 and influences mRNA fate. *RNA* 19: 295-305.
- Huelsenbeck, JP, Crandall, KA. 2003. PHYLOGENY ESTIMATION AND HYPOTHESIS TESTING USING MAXIMUM LIKELIHOOD. <http://dx.doi.org/10.1146/annurev.ecolsys.28.1.437>.
- Huelsenbeck, JP, Rannala, B, Felsenstein, J, Ou, C-Y, Nichol, ST, Hjelle, B, Swofford, DL, Thorne, JL, Felsenstein, J, Wiegmann, BM, Hillis, DM, Bull, JJ, Archie, JW, Maddison, D, Felsenstein, J, Cavalli-Sforza, LL, Edwards, AWF, Felsenstein, J, Kuhner, MK, Felsenstein, J, Huelsenbeck, JP, Goldman, N, Felsenstein, J, Chang, JT, Yang, Z, King,

- JL, Jukes, TH, Hasegawa, M, Kishino, H, Yano, T, Palumbi, SR, Yang, Z, Yang, Z, \_\_\_, Roberts, D, Hafner, MS, Page, RDM, Huelsenbeck, JP, Rannala, B, Yang, Z, Rzhetsky, A, Nei, M, Yang, Z, Goldman, N, Friday, A, Dykhuizen, DE, Green, L, Hibbett, DS, Henderson, WW, Brooks, DR, Hafner, MS, Nadler, SA, Huelsenbeck, JP, Bull, JJ, Page, RDM, Schöniger, M, Haeseler, A von, Huelsenbeck, JP, Hillis, DM, Nielsen, R. 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* 276: 227–32.
- Huntzinger, E, Braun, JE, Heimstädt, S, Zekri, L, Izaurralde, E. 2010. Two PABPC1-binding sites in GW182 proteins promote miRNA-mediated gene silencing. *EMBO J.* 29: 4146–60.
- Huntzinger, E, Izaurralde, E. 2011. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat. Rev. Genet.* 12: 99–110.
- Huntzinger, E, Kuzuoglu-Öztürk, D, Braun, JE, Eulalio, A, Wohlbold, L, Izaurralde, E, Kuzuo??lu-??zt??rk, D, Braun, JE, Eulalio, A, Wohlbold, L, Izaurralde, E. 2013. The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. *Nucleic Acids Res.* 41: 978–994.
- Hyrien, O. 2016. How MCM loading and spreading specify eukaryotic DNA replication initiation sites. *F1000Research* 5.
- Inada, T, Makino, S. 2014. Novel roles of the multi-functional CCR4-NOT complex in post-transcriptional regulation. *Front. Genet.* 5: 135.
- Ingelfinger, D, Arndt-Jovin, DJ, Lührmann, R, Achsel, T. 2002. The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrn1 in distinct cytoplasmic foci. *RNA* 8: 1489–501.
- Inoue, T, Morita, M, Hijikata, A, Fukuda-Yuzawa, Y, Adachi, S, Isono, K, Ikawa, T, Kawamoto, H, Koseki, H, Natsume, T, Fukao, T, Ohara, O, Yamamoto, T, Kurosaki, T. 2015. CNOT3 contributes to early B cell development by controlling Igh rearrangement and p53 mRNA stability. *J. Exp. Med.* 212: 1465–79.
- Isken, O, Maquat, LE. 2007. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev.* 21: 1833–56.
- Ito, K, Inoue, T, Yokoyama, K, Morita, M, Suzuki, T, Yamamoto, T. 2011a. CNOT2 depletion disrupts and inhibits the CCR4-NOT deadenylase complex and induces apoptotic cell death. *Genes to Cells* 16: 368–379.
- Ito, K, Takahashi, A, Morita, M, Suzuki, T, Yamamoto, T. 2011b. The role of the CNOT1 subunit of the CCR4-NOT complex in mRNA deadenylation and cell viability. *Protein Cell* 2: 755–763.
- Jaruzelska, J, Kotecki, M, Kusz, K, Spik, A, Firpo, M, Reijo Pera, RA. 2003. Conservation of a Pumilio-Nanos complex from Drosophila germ plasm to human germ cells. *Dev. Genes Evol.* 213: 120–6.
- Jinek, M, Fabian, MR, Coyle, SM, Sonenberg, N, Doudna, JA. 2010. Structural insights into the human GW182-PABC interaction in microRNA-mediated deadenylation. *Nat. Struct. Mol. Biol.* 17: 238–40.
- Johnson, JM, Castle, J, Garrett-Engele, P, Kan, Z, Loerch, PM, Armour, CD, Santos, R, Schadt, EE, Stoughton, R, Shoemaker, DD. 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon

- junction microarrays. *Science* 302: 2141–4.
- Johnson, TL, Vilardeell, J. 2012. Regulated pre-mRNA splicing: the ghostwriter of the eukaryotic genome. *Biochim. Biophys. Acta* 1819: 538–45.
- Joly, W, Chartier, A, Rojas-Rios, P, Busseau, I, Simonelig, M. 2013. The CCR4 deadenylase acts with Nanos and Pumilio in the fine-tuning of Mei-P26 expression to promote germline stem cell self-renewal. *Stem cell reports* 1: 411–24.
- Jonstrup, AT, Andersen, KR, Van, LB, Brodersen, DE. 2007. The 1.4-Å crystal structure of the *S. pombe* Pop2p deadenylase subunit unveils the configuration of an active enzyme. *Nucleic Acids Res.* 35: 3153–3164.
- Juven-Gershon, T, Hsu, J-Y, Theisen, JW, Kadonaga, JT. 2008. The RNA polymerase II core promoter - the gateway to transcription. *Curr. Opin. Cell Biol.* 20: 253–9.
- Kadonaga, JT. 2012. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip. Rev. Dev. Biol.* 1: 40–51.
- Karousis, ED, Nasif, S, Mühlemann, O. 2016. Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact. *Wiley Interdiscip. Rev. RNA* 7: 661–682.
- Katoh, K, Misawa, K, Kuma, K, Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30: 3059–3066.
- Katoh, K, Standley, DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30: 772–780.
- De Keersmaecker, K, Atak, ZK, Li, N, Vicente, C, Patchett, S, Girardi, T, Gianfelici, V, Geerdens, E, Clappier, E, Porcu, M, Lahortiga, I, Lucà, R, Yan, J, Hulselmans, G, Vranckx, H, Vandepoel, R, Sweron, B, Jacobs, K, Mentens, N, Wlodarska, I, Cauwelier, B, Cloos, J, Soulier, J, Uyttebroeck, A, Bagni, C, Hassan, BA, Vandenberghe, P, Johnson, AW, Aerts, S, Cools, J. 2013. Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. *Nat. Genet.* 45: 186–90.
- Kerr, SC, Azzouz, N, Fuchs, SM, Collart, MA, Strahl, BD, Corbett, AH, Laribee, RN. 2011. The Ccr4-Not complex interacts with the mRNA export machinery. *PLoS One* 6: e18302.
- Körner, CG, Wahle, E. 1997. Poly(A) tail shortening by a mammalian poly(A)-specific 3'-exoribonuclease. *J. Biol. Chem.* 272: 10448–56.
- Kostrewa, D, Zeller, ME, Armache, K-J, Seizl, M, Leike, K, Thomm, M, Cramer, P. 2009. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* 462: 323–30.
- Krissinel, E, Henrick, K. 2007. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* 372: 774–797.
- Kshirsagar, M, Parker, R. 2004. Identification of Edc3p as an Enhancer of mRNA Decapping in *Saccharomyces cerevisiae*. *Genetics* 166.
- Kumar, P, Henikoff, S, Ng, PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4: 1073–81.

- Kuzuoğlu-Öztürk, D, Bhandari, D, Huntzinger, E, Fauser, M, Helms, S, Izaurralde, E. 2016. miRISC and the CCR4-NOT complex silence mRNA targets independently of 43S ribosomal scanning. *EMBO J.* 35: 1186–203.
- Kuzuoğlu-Öztürk, D, Huntzinger, E, Schmidt, S, Izaurralde, E. 2012. The *Caenorhabditis elegans* GW182 protein AIN-1 interacts with PAB-1 and subunits of the PAN2-PAN3 and CCR4-NOT deadenylase complexes. *Nucleic Acids Res.* 40: 5651–5665.
- Łabno, A, Tomecki, R, Dziembowski, A. 2016. Cytoplasmic RNA decay pathways - enzymes and mechanisms. *Biochim. Biophys. Acta - Mol. Cell Res.*
- LaGrandeur, TE, Parker, R, Adams, B, Morgan, M, Muthukrishnan, S, Hecht, S, Shatkin, A, Bashkirov, V, Scherthan, H, Solinger, J, Buerstedde, J, Heyer, W, Beelman, C, Parker, R, Beelman, C, Stevens, A, Caponigro, G, LaGrandeur, T, Hatfield, L, Fortner, D, Parker, R, Blank, A, Sugiyama, R, Dekker, C, Caponigro, G, Parker, R, Cereghino, G, Scheffler, I, Couttet, P, Fromont-Racine, M, Steel, D, Pictet, R, Grange, T, Cui, Y, Hagan, K, Zhang, S, Peltz, S, Decker, C, Parker, R, Gonzalez, C, Martin, C, Hagan, K, Ruiz-Echevarria, M, Quan, Y, Peltz, S, Harris, M, Nolan, J, Malhotra, A, Brown, J, Harvey, S, Pace, N, Hatfield, L, Beelman, C, Stevens, A, Parker, R, He, F, Jacobson, A, Heaton, B, Decker, C, Muhlrad, D, Donahue, J, Jacobson, A, Parker, R, Heyer, W, Johnson, A, Reinhart, U, Kolodner, R, Higgs, D, Colbert, J, Hodel, A, Shi, X, Gershon, P, Konarska, M, Padgett, R, Sharp, P, Laemmli, U, LaGrandeur, T, Hüttenhofer, A, Noller, H, Pace, N, Leeds, P, Peltz, S, Jacobson, A, Culbertson, M, Leeds, P, Wood, J, Lee, B, Culbertson, M, Lim, S, Maquat, L, Marcotrigianol, J, Gingras, A, Sonenberg, N, Burley, S, McCubbin, W, Ederly, I, Altmann, M, Sonenberg, N, et al. 1998. Isolation and characterization of Dcp1p, the yeast mRNA decapping enzyme. *EMBO J.* 17: 1487–96.
- Lai, F, King, M Lou. 2013. Repressive translational control in germ cells. *Mol. Reprod. Dev.* 80: 665–76.
- Lanave, C, Preparata, G, Saccone, C, Serio, G. 1984. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* 20: 86–93.
- Landick, R. 2009. Functional divergence in the growing family of RNA polymerases. *Structure* 17: 323–5.
- Laribee, RN, Shibata, Y, Mersman, DP, Collins, SR, Kemmeren, P, Roguev, A, Weissman, JS, Briggs, SD, Krogan, NJ, Strahl, BD. 2007. CCR4/NOT complex associates with the proteasome and regulates histone methylation. *Proc. Natl. Acad. Sci. U. S. A.* 104: 5836–41.
- Lau, N-C, Kolkman, A, van Schaik, FM a, Mulder, KW, Pijnappel, WWMP, Heck, AJR, Timmers, HTM. 2009. Human Ccr4-Not complexes contain variable deadenylase subunits. *Biochem. J.* 422: 443–453.
- Lee, HH, Kim, Y-S, Kim, KH, Heo, I, Kim, SK, Kim, O, Kim, HK, Yoon, JY, Kim, HS, Kim, DJ, Lee, SJ, Yoon, HJ, Kim, SJ, Lee, BG, Song, HK, Kim, VN, Park, C-M, Suh, SW. 2007. Structural and functional insights into Dom34, a key component of no-go mRNA decay. *Mol. Cell* 27: 938–50.
- Lee, TI, Young, RA. 2013. Transcriptional Regulation and Its Misregulation in Disease. *Cell* 152: 1237–1251.
- Lefort, V, Hordijk, W, Gascuel, O. 2010-Phyml. 1–37.

- Lejeune, F, Li, X, Maquat, LE. 2003. Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. *Mol. Cell* 12: 675–87.
- Lejeune, F, Maquat, LE. 2005. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr. Opin. Cell Biol.* 17: 309–15.
- Lemey, P, Salemi, M, Vamdamme, A. 2009. The phylogenetic handbook. 723 p.
- Liu, HY, Badarinarayana, V, Audino, DC, Rappsilber, J, Mann, M, Denis, CL. 1998. The NOT proteins are part of the CCR4 transcriptional complex and affect gene expression both positively and negatively. *EMBO J.* 17: 1096–1106.
- Liu, HY, Toyn, JH, Chiang, YC, Draper, MP, Johnston, LH, Denis, CL. 1997. DBF2, a cell cycle-regulated protein kinase, is physically and functionally associated with the CCR4 transcriptional regulatory complex. *EMBO J.* 16: 5289–5298.
- Loh, B, Jonas, S, Izaurralde, E. 2013. The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev.* 27: 2125–38.
- Lowell, JE, Rudner, DZ, Sachs, AB. 1992. 3'-UTR-dependent deadenylation by the yeast poly(A) nuclease. *Genes Dev.* 6: 2088–99.
- Lykke-Andersen, J, Wagner, E. 2005. Recruitment and activation of mRNA decay enzymes by two ARE-mediated decay activation domains in the proteins TTP and BRF-1. *Genes Dev.* 19: 351–61.
- Ma, J, Ptashne, M. 1987. A new class of yeast transcriptional activators. *Cell* 51: 113–9.
- Maillet, L, Collart, MA. 2002. Interaction between Not1p, a component of the Ccr4-not complex, a global regulator of transcription, and Dhh1p, a putative RNA helicase. *J. Biol. Chem.* 277: 2835–42.
- Malvar, T, Biron, RW, Kaback, DB, Denis, CL. 1992. The CCR4 protein from *Saccharomyces cerevisiae* contains a leucine-rich repeat region which is required for its control of ADH2 gene expression. *Genetics* 132: 951–962.
- Mangus, DA, Evans, MC, Agrin, NS, Smith, M, Gongidi, P, Jacobson, A. 2004. Positive and negative regulation of poly(A) nuclease. *Mol. Cell. Biol.* 24: 5521–33.
- Maquat, LE. 2002. Molecular biology. Skiing toward nonstop mRNA decay. *Science* 295: 2221–2.
- Marshall, NF, Price, DH. 1995. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J. Biol. Chem.* 270: 12335–8.
- Martin, A, Ochagavia, ME, Rabasa, LC, Miranda, J, Fernandez-de-Cossio, J, Bringas, R. 2010. Bisogenet: a new tool for gene network building, visualization and analysis. *BMC Bioinformatics* 11: 91.
- Martin Haas, ‡, Michaela Siegert, ‡, André Schürmann, ‡, Beate Sodeik, § and Heiner Wolfes\*, ‡. 2004. c-Myb Protein Interacts with Rcd-1, a Component of the CCR4 Transcription Mediator Complex†.
- Martinez, J, Ren, YG, Nilsson, P, Ehrenberg, M, Virtanen, A. 2001. The mRNA cap structure stimulates rate of poly(A) removal and amplifies

- processivity of degradation. *J. Biol. Chem.* 276: 27923–9.
- Martinez, J, Ren, YG, Thuresson, AC, Hellman, U, Astrom, J, Virtanen, A. 2000. A 54-kDa fragment of the Poly(A)-specific ribonuclease is an oligomeric, processive, and cap-interacting Poly(A)-specific 3' exonuclease. *J. Biol. Chem.* 275: 24222–30.
- Maryati, M, Kaur, I, Jadhav, GP, Olotu-Umoren, L, Oveh, B, Hashmi, L, Fischer, PM, Winkler, GS. 2014. A fluorescence-based assay suitable for quantitative analysis of deadenylase enzyme activity. *Nucleic Acids Res.* 42.
- Mathys, H, Basquin, J, Ozgur, S, Czarnocki-Cieciura, M, Bonneau, F, Aartse, A, Dziembowski, A, Nowotny, M, Conti, E, Filipowicz, W. 2014. Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. *Mol. Cell* 54: 751–65.
- Matsuda, S, Rouault, J, Magaud, J, Berthet, C. 2001. In search of a function for the TIS21/PC3/BTG1/TOB family. *FEBS Lett.* 497: 67–72.
- Mauxion, F, Chen, CYA, Séraphin, B, Shyu, A Bin. 2009. BTG/TOB factors impact deadenylases. *Trends Biochem. Sci.* 34: 640–647.
- Mauxion, F, Faux, C, Séraphin, B. 2008. The BTG2 protein is a general activator of mRNA deadenylation. *EMBO J.* 27: 1039–1048.
- Mauxion, F, Prève, B, Séraphin, B. 2013. C2ORF29/CNOT11 and CNOT10 form a new module of the CCR4-NOT complex. *RNA Biol.* 10: 267–76.
- McManus, CJ, Graveley, BR. 2011. RNA structure and the mechanisms of alternative splicing. *Curr. Opin. Genet. Dev.* 21: 373–9.
- Mersman, DP, Du, H-N, Fingerman, IM, South, PF, Briggs, SD. 2009. Polyubiquitination of the demethylase Jhd2 controls histone methylation and gene expression. *Genes Dev.* 23: 951–62.
- Meyer, S, Temme, C, Wahle, E. 2004. Messenger RNA Turnover in Eukaryotes: Pathways and Enzymes. *Crit. Rev. Biochem. Mol. Biol.* 39: 197–216.
- Mitchell, SF, Walker, SE, Algire, MA, Park, E-H, Hinnebusch, AG, Lorsch, JR. 2010. The 5'-7-methylguanosine cap on eukaryotic mRNAs serves both to stimulate canonical translation initiation and to block an alternative pathway. *Mol. Cell* 39: 950–62.
- Mittal, S, Aslam, A, Doidge, R, Medica, R, Winkler, GS. 2011. The Ccr4a (CNOT6) and Ccr4b (CNOT6L) deadenylase subunits of the human Ccr4-Not complex contribute to the prevention of cell death and senescence. *Mol. Biol. Cell* 22: 748–58.
- Miyasaka, T, Morita, M, Ito, K, Suzuki, T, Fukuda, H, Takeda, S, Inoue, JI, Semba, K, Yamamoto, T. 2008. Interaction of antiproliferative protein Tob with the CCR4-NOT deadenylase complex. *Cancer Sci.* 99: 755–761.
- Mohammed, H, D'Santos, C, Serandour, AA, Ali, HR, Brown, GD, Atkins, A, Rueda, OM, Holmes, KA, Theodorou, V, Robinson, JLL, Zwart, W, Saadi, A, Ross-Innes, CS, Chin, S-F, Menon, S, Stingl, J, Palmieri, C, Caldas, C, Carroll, JS. 2013. Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Rep.* 3: 342–9.
- Monecke, T, Schell, S, Dickmanns, A, Ficner, R. 2008. Crystal structure of the RRM domain of poly(A)-specific ribonuclease reveals a novel

- m(7)G-cap-binding mode. *J. Mol. Biol.* 382: 827–34.
- Morin, RD, Mendez-Lago, M, Mungall, AJ, Goya, R, Mungall, KL, Corbett, RD, Johnson, NA, Severson, TM, Chiu, R, Field, M, Jackman, S, Krzywinski, M, Scott, DW, Trinh, DL, Tamura-Wells, J, Li, S, Firme, MR, Rogic, S, Griffith, M, Chan, S, Yakovenko, O, Meyer, IM, Zhao, EY, Smailus, D, Moksa, M, Chittaranjan, S, Rimsza, L, Brooks-Wilson, A, Spinelli, JJ, Ben-Neriah, S, Meissner, B, Woolcock, B, Boyle, M, McDonald, H, Tam, A, Zhao, Y, Delaney, A, Zeng, T, Tse, K, Butterfield, Y, Birol, I, Holt, R, Schein, J, Horsman, DE, Moore, R, Jones, SJM, Connors, JM, Hirst, M, Gascoyne, RD, Marra, MA. 2011. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* 476: 298–303.
- Morita, M, Oike, Y, Nagashima, T, Kadomatsu, T, Tabata, M, Suzuki, T, Nakamura, T, Yoshida, N, Okada, M, Yamamoto, T. 2011. Obesity resistance and increased hepatic expression of catabolism-related mRNAs in *Cnot3*<sup>+/-</sup> mice. *EMBO J.* 30: 4678–91.
- Morita, M, Suzuki, T, Nakamura, T, Yokoyama, K, Miyasaka, T, Yamamoto, T. 2007. Depletion of mammalian CCR4b deadenylase triggers elevation of the p27Kip1 mRNA level and impairs cell growth. *Mol. Cell. Biol.* 27: 4980–90.
- Morris, AR, Mukherjee, N, Keene, JD. 2010. Systematic analysis of posttranscriptional gene expression. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 2: 162–180.
- Mulder, KW, Brenkman, AB, Inagaki, A, van den Broek, NJF, Timmers, HTM. 2007. Regulation of histone H3K4 tri-methylation and PAF complex recruitment by the Ccr4-Not complex. *Nucleic Acids Res.* 35: 2428–39.
- Nagata, T, Suzuki, S, Endo, R, Shirouzu, M, Terada, T, Inoue, M, Kigawa, T, Kobayashi, N, Güntert, P, Tanaka, A, Hayashizaki, Y, Muto, Y, Yokoyama, S. 2008. The RRM domain of poly(A)-specific ribonuclease has a noncanonical binding site for mRNA cap analog recognition. *Nucleic Acids Res.* 36: 4754–67.
- Nasertorabi, F, Batisse, C, Diepholz, M, Suck, D, Böttcher, B. 2011. Insights into the structure of the CCR4-NOT complex by electron microscopy. *FEBS Lett.* 585: 2182–2186.
- Natalizio, BJ, Wenthe, SR. 2013. Postage for the messenger: designating routes for nuclear mRNA export. *Trends Cell Biol.* 23: 365–73.
- Neely, GG, Kuba, K, Cammarato, A, Isobe, K, Amann, S, Zhang, L, Murata, M, Elmén, L, Gupta, V, Arora, S, Sarangi, R, Dan, D, Fujisawa, S, Usami, T, Xia, C, Keene, AC, Alayari, NN, Yamakawa, H, Elling, U, Berger, C, Novatchkova, M, Kogelgruber, R, Fukuda, K, Nishina, H, Isobe, M, Pospisilik, JA, Imai, Y, Pfeufer, A, Hicks, AA, Pramstaller, PP, Subramaniam, S, Kimura, A, Ocorr, K, Bodmer, R, Penninger, JM. 2010. A global in vivo *Drosophila* RNAi screen identifies NOT3 as a conserved regulator of heart function. *Cell* 141: 142–53.
- Ng, PC, Henikoff, S. 2002. Accounting for human polymorphisms predicted to affect protein function. *Genome Res.* 12: 436–46.
- Ng, PC, Henikoff, S. 2001. Predicting deleterious amino acid substitutions. *Genome Res.* 11: 863–74.
- Ng, PC, Henikoff, S. 2006. Predicting the effects of amino acid substitutions on protein function. *Annu. Rev. Genomics Hum. Genet.* 7: 61–80.

- Ng, PC, Henikoff, S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31: 3812–4.
- Nielsen, R, Yang, Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148: 929–936.
- Nishi, K, Takahashi, T, Suzawa, M, Miyakawa, T, Nagasawa, T, Ming, Y, Tanokura, M, Ui-Tei, K. 2015. Control of the localization and function of a miRNA silencing component TNRC6A by Argonaute protein. *Nucleic Acids Res.* 43: 9856–73.
- Nishimura, T, Padamsi, Z, Gingras, A-C, Correspondence, MRF, Fakim, H, Milette, S, Dunham, WH, Fabian, MR. 2015. The eIF4E-Binding Protein 4E-T Is a Component of the mRNA Decay Machinery that Bridges the 5 Termini of Target mRNAs Termini of Target mRNAs. *Cell Rep.* 11: 1425–1436.
- Nissan, T, Rajyaguru, P, She, M, Song, H, Parker, R. 2010. Decapping activators in *Saccharomyces cerevisiae* act by multiple mechanisms. *Mol. Cell* 39: 773–83.
- Oberholzer, U, Collart, MA. 1998. Characterization of NOT5 that encodes a new component of the Not protein complex. *Gene* 207: 61–9.
- Ozgun, S, Basquin, J, Kamenska, A, Filipowicz, W, Standart, N, Conti, E. 2015a. Structure of a Human 4E-T/DDX6/CNOT1 Complex Reveals the Different Interplay of DDX6-Binding Proteins with the CCR4-NOT Complex. *Cell Rep.* 13: 703–711.
- Ozgun, S, Basquin, J, Kamenska, A, Filipowicz, W, Standart, N, Conti, E, Adams, PD, Afonine, PV, Bunkóczi, G, Chen, VB, Davis, IW, Echols, N, Headd, JJ, Hung, L-W, Kapral, GJ, Grosse-Kunstleve, RW, al., et, Anderson, P, Kederasha, N, Ayache, J, Bénard, M, Ernoult-Lange, M, Minshall, N, Standart, N, Kress, M, Weil, D, Barbee, SA, Estes, PS, Cziko, A-M, Hillebrand, J, Luedeman, RA, Collier, JM, Johnson, N, Howlett, IC, Geng, C, Ueda, R, al., et, Basquin, J, Roudko, VV, Rode, M, Basquin, C, Séraphin, B, Conti, E, Beck, M, Schmidt, A, Malmstroem, J, Claassen, M, Ori, A, Szyborska, A, Herzog, F, Rinner, O, Ellenberg, J, Aebersold, R, Boag, PR, Atalay, A, Robida, S, Reinke, V, Blackwell, TK, Braun, JE, Huntzinger, E, Fauser, M, Izaurralde, E, Carroll, JS, Munchel, SE, Weis, K, Chapat, C, Corbo, L, Chekulaeva, M, Mathys, H, Zipprich, JT, Attig, J, Colic, M, Parker, R, Filipowicz, W, Chen, CYA, Shyu, AB, Chen, VB, Arendall, WB, Headd, JJ, Keedy, DA, Immormino, RM, Kapral, GJ, Murray, LW, Richardson, JS, Richardson, DC, Chen, Y, Boland, A, Kuzuoğlu-Öztürk, D, Bawankar, P, Loh, B, Chang, C-T, Weichenrieder, O, Izaurralde, E, Cheng, Z, Collier, JM, Parker, R, Song, H, Chu, C-Y, et al. 2015b. Structure of a Human 4E-T/DDX6/CNOT1 Complex Reveals the Different Interplay of DDX6-Binding Proteins with the CCR4-NOT Complex. *Cell Rep.* 13: 703–11.
- Ozgun, S, Chekulaeva, M, Stoecklin, G. 2010. Human Pat1b connects deadenylation with mRNA decapping and controls the assembly of processing bodies. *Mol. Cell. Biol.* 30: 4308–23.
- Parker, R, Song, H. 2004. The enzymes and control of eukaryotic mRNA turnover. *TL - 11. Nat. Struct. Mol. Biol.* 11 VN-r: 121–127.
- Patton, JG, Mayer, SA, Tempst, P, Nadal-Ginard, B. 1991. Characterization

- and molecular cloning of polypyrimidine tract-binding protein: a component of a complex necessary for pre-mRNA splicing. *Genes Dev.* 5: 1237–51.
- Peng, W, Togawa, C, Zhang, K, Kurdistani, SK. 2008. Regulators of cellular levels of histone acetylation in *Saccharomyces cerevisiae*. *Genetics* 179: 277–89.
- Peterlin, BM, Price, DH. 2006. Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* 23: 297–305.
- Petit, A-PP, Wohlbold, L, Bawankar, P, Huntzinger, E, Schmidt, S, Izaurralde, E, Weichenrieder, O. 2012. The structural basis for the interaction between the CAF1 nuclease and the NOT1 scaffold of the human CCR4-NOT deadenylase complex. *Nucleic Acids Res.* 40: 11058–11072.
- Pettersen, EF, Goddard, TD, Huang, CC, Couch, GS, Greenblatt, DM, Meng, EC, Ferrin, TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25: 1605–12.
- Pfaff, J, Meister, G. 2013. Argonaute and GW182 proteins: an effective alliance in gene silencing. *Biochem. Soc. Trans.* 41: 855–60.
- Piccirillo, C, Khanna, R, Kiledjian, M. 2003. Functional characterization of the mammalian mRNA decapping enzyme hDcp2. *RNA* 9: 1138–47.
- Pisarev, A V, Hellen, CUT, Pestova, T V. 2007. Recycling of eukaryotic posttermination ribosomal complexes. *Cell* 131: 286–99.
- Posada, D. 2008. jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* 25: 1253–1256.
- Posada, D. 2003. Using MODELTEST and PAUP\* to select a model of nucleotide substitution. *Curr. Protoc. Bioinformatics* Chapter 6: Unit 6.5.
- Prévôt, D, Morel, AP, Voeltzel, T, Rostan, MC, Rimokh, R, Magaud, JP, Corbo, L. 2001. Relationships of the antiproliferative proteins BTG1 and BTG2 with CAF1, the human homolog of a component of the yeast CCR4 transcriptional complex: involvement in estrogen receptor alpha signaling pathway. *J. Biol. Chem.* 276: 9640–8.
- Proudfoot, NJ. 2011. Ending the message: poly(A) signals then and now. *Genes Dev.* 25: 1770–82.
- Quesada, V, Díaz-Perales, A, Gutiérrez-Fernández, A, Garabaya, C, Cal, S, López-Otín, C. 2004. Cloning and enzymatic analysis of 22 novel human ubiquitin-specific proteases. *Biochem. Biophys. Res. Commun.* 314: 54–62.
- Raisch, T, Bhandari, D, Sabath, K, Helms, S, Valkov, E, Weichenrieder, O, Izaurralde, E. 2016. Distinct modes of recruitment of the CCR4-NOT complex by *Drosophila* and vertebrate Nanos. *EMBO J.* 35: 974–90.
- Rehwinkel, J, Behm-Ansmant, I, Gatfield, D, Izaurralde, E. 2005. A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *RNA* 11: 1640–7.
- Ren, Y-G, Kirsebom, LA, Virtanen, A. 2004. Coordination of divalent metal ions in the active site of poly(A)-specific ribonuclease. *J. Biol. Chem.* 279: 48702–6.
- Ren, Y-G, Martínez, J, Kirsebom, LA, Virtanen, A. 2002a. Inhibition of Klenow DNA polymerase and poly(A)-specific ribonuclease by

- aminoglycosides. *RNA* 8: 1393–400.
- Ren, Y-G, Martínez, J, Virtanen, A. 2002b. Identification of the active site of poly(A)-specific ribonuclease by site-directed mutagenesis and Fe(2+)-mediated cleavage. *J. Biol. Chem.* 277: 5982–7.
- Rios, D, McLaren, WM, Chen, Y, Birney, E, Stabenau, A, Flicek, P, Cunningham, F. 2010. A database and API for variation, dense genotyping and resequencing data. *BMC Bioinformatics* 11: 238.
- Rodríguez, F, Oliver, JL, Marín, A, Medina, JR. 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* 142: 485–501.
- Rose, AM, Shah, AZ, Venturini, G, Rivolta, C, Rose, GE, Bhattacharya, SS. 2014. Dominant PRPF31 mutations are hypostatic to a recessive CNOT3 polymorphism in retinitis pigmentosa: a novel phenomenon of "linked trans-acting epistasis". *Ann. Hum. Genet.* 78: 62–71.
- Rouault, JP, Falette, N, Guéhenneux, F, Guillot, C, Rimokh, R, Wang, Q, Berthet, C, Moyret-Lalle, C, Savatier, P, Pain, B, Shaw, P, Berger, R, Samarut, J, Magaud, JP, Ozturk, M, Samarut, C, Puisieux, A. 1996. Identification of BTG2, an antiproliferative p53-dependent component of the DNA damage cellular response pathway. *Nat. Genet.* 14: 482–6.
- Roundtree, IA, He, C. 2016. Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Trends Genet.* 32: 320–1.
- Sachs, AB, Deardorff, JA. 1992. Translation initiation requires the PAB-dependent poly(A) ribonuclease in yeast. *Cell* 70: 961–73.
- Saito, S, Hosoda, N, Hoshino, S. 2013. The Hbs1-Dom34 protein complex functions in non-stop mRNA decay in mammalian cells. *J. Biol. Chem.* 288: 17832–43.
- Saitou N, NM. 1987. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees'. *Mol. Biol. Evol.* 4: 406–425.
- Sakai, A, Chibazakura, T, Shimizu, Y, Hishinuma, F. 1992. Molecular analysis of POP2 gene, a gene required for glucose-derepression of gene expression in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 20: 6227–6233.
- Sandler, H, Kreth, J, Timmers, HTM, Stoecklin, G. 2011. Not1 mediates recruitment of the deadenylase Caf1 to mRNAs targeted for degradation by tristetraprolin. *Nucleic Acids Res.* 39: 4373–4386.
- Sandler, H, Stoecklin, G. 2008. Control of mRNA decay by phosphorylation of tristetraprolin. *Biochem. Soc. Trans.* 36: 491–6.
- Schäfer, IB, Rode, M, Bonneau, F, Schüssler, S, Conti, E. 2014. The structure of the Pan2–Pan3 core complex reveals cross-talk between deadenylase and pseudokinase. *Nat. Publ. Gr.* 21.
- Schwartz, D, Decker, CJ, Parker, R. 2003. The enhancer of decapping proteins, Edc1p and Edc2p, bind RNA and stimulate the activity of the decapping enzyme. *RNA* 9: 239–51.
- Shannon, P, Markiel, A, Ozier, O, Baliga, NS, Wang, JT, Ramage, D, Amin, N, Schwikowski, B, Ideker, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13: 2498–504.
- Sharma, N. 2016. Regulation of RNA polymerase II-mediated

- transcriptional elongation: Implications in human disease. *IUBMB Life* 68: 709–16.
- She, M, Decker, CJ, Svergun, DI, Round, A, Chen, N, Muhrad, D, Parker, R, Song, H. 2008. Structural basis of dcp2 recognition and activation by dcp1. *Mol. Cell* 29: 337–49.
- Sheets, MD, Wickens, M. 1989. Two phases in the addition of a poly(A) tail. *Genes Dev.* 3: 1401–12.
- Shi, Y. 2012. Alternative polyadenylation: new insights from global analyses. *RNA* 18: 2105–17.
- Shi, Y, Di Giammartino, DC, Taylor, D, Sarkeshik, A, Rice, WJ, Yates, JR, Frank, J, Manley, JL. 2009. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell* 33: 365–76.
- Shi, Y, Manley, JL. 2015. The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev.* 29: 889–97.
- Shimodaira, H, Hasegawa, M. 1999. Letter to the Editor Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Mol. Biol. Evol.* 16: 1114–1116.
- Siddiqui, N, Mangus, DA, Chang, T-CC, Palermino, J-MM, Shyu, A-B Bin, Gehring, K. 2007. Poly(A) nuclease interacts with the C-terminal domain of polyadenylate-binding protein domain from poly(A)-binding protein. *J. Biol. Chem.* 282: 25067–25075.
- Sims, RJ, Belotserkovskaya, R, Reinberg, D. 2004. Elongation by RNA polymerase II: the short and long of it. *Genes Dev.* 18: 2437–68.
- Singh, G, Rebbapragada, I, Lykke-Andersen, J. 2008. A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol.* 6: e111.
- Smith, JE, Baker, KE. 2015. Nonsense-mediated RNA decay - a switch and dial for regulating gene expression. *BioEssays* 37: 612–623.
- Sperling, R. 2016. The nuts and bolts of the endogenous spliceosome. *Wiley Interdiscip. Rev. RNA.*
- Steiger, M, Carr-Schmid, A, Schwartz, DC, Kiledjian, M, Parker, R. 2003. Analysis of recombinant yeast decapping enzyme. *RNA* 9: 231–8.
- Stupfler, B, Birck, C, Séraphin, B, Mauxion, F, Matsuda, S, Rouault, J, Magaud, J, Berthet, C, Lim, NS, Okochi, K, Suzuki, T, Inoue, J, Matsuda, S, Yamamoto, T, Funakoshi, Y, Ezzeddine, N, Winkler, GS, Iacopetti, P, Barsacchi, G, Tirone, F, Maffei, L, Cremisi, F, Iacopetti, P, Attardo, A, Canzoniere, D, Fei, JF, Haffner, C, Huttner, WB, Farioli-Vecchioli, S, Park, S, Duriez, C, Boiko, AD, Kawakubo, H, Struckmann, K, Kawakubo, H, Mollerstrom, E, Xu, K, Bai, Y, Zhang, A, Zhang, Q, Bartlam, MG, Basquin, J, Tucker, M, Daugeron, MC, Mauxion, F, Seraphin, B, Yamashita, A, Lau, NC, Fabian, MR, Bhandari, D, Raisch, T, Weichenrieder, O, Jonas, S, Izaurralde, E, Mathys, H, Chen, Y, Guardavaccaro, D, Prevot, D, Passeri, D, Mauxion, F, Faux, C, Seraphin, B, Ezzeddine, N, Chen, CY, Shyu, AB, Sasajima, H, Nakagawa, K, Yokosawa, H, Hong, JW, Ryu, MS, Lim, IK, Xie, J, Kozlov, G, Gehring, K, Kuhn, U, Pieler, T, Melo, EO, Dhaliya, R, Sa, CM de, Standart, N, Neto, OP de M, Lin, J, Fabian, M, Sonenberg, N, Meller, A, Yang, X, Berthet, C, Horiuchi, M, Doidge, R, Mittal, S, Aslam, A, Winkler, GS, Deo, RC,

- Bonanno, JB, Sonenberg, N, Burley, SK, Kerekatte, V, Alvarez, E, et al. 2016. BTG2 bridges PABPC1 RNA-binding domains and CAF1 deadenylase to control cell proliferation. *Nat. Commun.* 7: 10811.
- Subtelny, AO, Eichhorn, SW, Chen, GR, Sive, H, Bartel, DP. 2014. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* 508: 66–71.
- Sun, H-Y, Kim, N, Hwang, C-S, Yoo, J-Y. 2015. Protein Degradation of RNA Polymerase II-Association Factor 1 (PAF1) Is Controlled by CNOT4 and 26S Proteasome. *PLoS One* 10: e0125599.
- Suyama, M, Torrents, D, Bork, P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34: W609-12.
- Suzuki, A, Niimi, Y, Shinmyozu, K, Zhou, Z, Kiso, M, Saga, Y. 2015a. Dead end 1 is an essential partner of NANOS 2 for selective binding of target RNAs in male germ cell development. *EMBO Rep.* 16: 1–10.
- Suzuki, T, K-Tsuzuku, J, Ajima, R, Nakamura, T, Yoshida, Y, Yamamoto, T. 2002. Phosphorylation of three regulatory serines of Tob by Erk1 and Erk2 is required for Ras-mediated cell proliferation and transformation. *Genes Dev.* 16: 1356–70.
- Suzuki, T, Kikuguchi, C, Sharma, S, Sasaki, T, Tokumasu, M, Adachi, S, Natsume, T, Kanegae, Y, Yamamoto, T. 2015b. CNOT3 suppression promotes necroptosis by stabilizing mRNAs for cell death-inducing proteins. *Sci. Rep.* 5: 14779.
- Takahashi, A, Kikuguchi, C, Morita, M, Shimodaira, T, Tokai-Nishizumi, N, Yokoyama, K, Ohsugi, M, Suzuki, T, Yamamoto, T. 2012. Involvement of CNOT3 in mitotic progression through inhibition of MAD1 expression. 268-273 p.
- Tamura, K, Nei, M. 1993. Estimation of the number of base nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10: 512–526.
- Tamura, K, Stecher, G, Peterson, D, Filipski, A, Kumar, S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30: 2725–2729.
- Taniguchi, I, Ohno, M. 2008. ATP-dependent recruitment of export factor Aly/REF onto intronless mRNAs by RNA helicase UAP56. *Mol. Cell Biol.* 28: 601–8.
- Temme, C, Zaessinger, S, Meyer, S, Simonelig, M, Wahle, E. 2004. A complex containing the CCR4 and CAF1 proteins is involved in mRNA deadenylation in *Drosophila*. *EMBO J.* 23: 2862–2871.
- Terry, LJ, Wentz, SR. 2009. Flexible gates: dynamic topologies and functions for FG nucleoporins in nucleocytoplasmic transport. *Eukaryot. Cell* 8: 1814–27.
- Tharun, S, He, W, Mayes, AE, Lennertz, P, Beggs, JD, Parker, R. 2000. Yeast Sm-like proteins function in mRNA decapping and decay. *Nature* 404: 515–518.
- Tharun, S, Parker, R. 2001. Targeting an mRNA for decapping: displacement of translation factors and association of the Lsm1p-7p complex on deadenylated yeast mRNAs. *Mol. Cell* 8: 1075–83.
- Thomas, MC, Chiang, C-M. 2006. The general transcription machinery and

- general cofactors. *Crit. Rev. Biochem. Mol. Biol.* 41: 105–78.
- Thore, S, Mauxion, F, Séraphin, B, Suck, D. 2003. X-ray structure and activity of the yeast Pop2 protein: a nuclease subunit of the mRNA deadenylase complex. *EMBO Rep.* 4: 1150–1155.
- Tian, B, Graber, JH. 2012. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip. Rev. RNA* 3: 385–96.
- Tirone, F. 2001. The gene PC3(TIS21/BTG2), prototype member of the PC3/BTG/TOB family: regulator in control of cell growth, differentiation, and DNA repair? *J. Cell. Physiol.* 187: 155–65.
- Topham, CM, Srinivasan, N, Blundell, TL. 1997. Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. *Protein Eng.* 10: 7–21.
- Topisirovic, I, Svitkin, Y V., Sonenberg, N, Shatkin, AJ. 2011. Cap and cap-binding proteins in the control of gene expression. *Wiley Interdiscip. Rev. RNA* 2: 277–298.
- Totaro, A, Renzi, F, La Fata, G, Mattioli, C, Raabe, M, Urlaub, H, Achsel, T. 2011. The human Pat1b protein: a novel mRNA deadenylation factor identified by a new immunoprecipitation technique. *Nucleic Acids Res.* 39: 635–47.
- Trinkle-Mulcahy, L, Boulon, S, Lam, YW, Urcia, R, Boisvert, F-M, Vandermoere, F, Morrice, NA, Swift, S, Rothbauer, U, Leonhardt, H, Lamond, A. 2008. Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J. Cell Biol.* 183: 223–39.
- Tucker, M, Valencia-Sanchez, MA, Staples, RR, Chen, J, Denis, CL, Parker, R. 2001. The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in *Saccharomyces cerevisiae*. *Cell* 104: 377–386.
- Uchida, N, Hoshino, S-I, Katada, T. 2004. Identification of a human cytoplasmic poly(A) nuclease complex stimulated by poly(A)-binding protein. *J. Biol. Chem.* 279: 1383–91.
- Ukleja, M, Cuellar, J, Siwaszek, A, Kasprzak, JM, Czarnocki-Cieciura, M, Bujnicki, JM, Dziembowski, A, Valpuesta, JM. 2016. The architecture of the *Schizosaccharomyces pombe* CCR4-NOT complex. *Nat. Commun.* 7: 10433.
- Venters, BJ, Pugh, BF. 2009. How eukaryotic genes are transcribed. *Crit. Rev. Biochem. Mol. Biol.* 44: 117–41.
- Venturini, G, Rose, AM, Shah, AZ, Bhattacharya, SS, Rivolta, C. 2012. CNOT3 is a modifier of PRPF31 mutations in retinitis pigmentosa with incomplete penetrance. *PLoS Genet.* 8: e1003040.
- Villanyi, Z, Collart, MA. 2015. Ccr4-Not is at the core of the eukaryotic gene expression circuitry. *Biochem. Soc. Trans.* 43: 1253–8.
- Villanyi, Z, Ribaud, V, Kassem, S, Panasenko, OO, Pahi, Z, Gupta, I, Steinmetz, L, Boros, I, Collart, MA, Hsin, J, Manley, J, Aguilera, A, Almeida, S de, Carmo-Fonseca, M, Harel-Sharvit, L, Eldad, N, Haimovich, G, Barkai, O, Duek, L, Goler-Baron, V, Selitrennik, M, Barkai, O, Haimovich, G, Lotan, R, Collart, M, Panasenko, O, Collart, M, Albert, T, Lemaire, M, Berkum, N van, Gentz, R, Collart, M, Temme, C, Zhang, L, Kremmer, E, Ihling, C, Chartier, A, Lau, N, Kolkman, A, Schaik,

- F van, Mulder, K, Pijnappel, W, Bawankar, P, Loh, B, Wohlbold, L, Schmidt, S, Izaurrealde, E, Mauxion, F, Preve, B, Seraphin, B, Farber, V, Erben, E, Sharma, S, Stoecklin, G, Clayton, C, Deluen, C, James, N, Maillet, L, Molinete, M, Theiler, G, Swanson, M, Qiu, H, Sumibcay, L, Krueger, A, Kim, S, Venters, B, Wachi, S, Mavrigh, T, Andersen, B, Jena, P, Kruk, J, Dutta, A, Fu, J, Gilmour, D, Reese, J, Lenssen, E, James, N, Pedruzzi, I, Dubouloz, F, Cameroni, E, Tucker, M, Valencia-Sanchez, M, Staples, R, Chen, J, Denis, C, Wahle, E, Winkler, G, Panasencko, O, Collart, M, Dimitrova, L, Kuroha, K, Tatematsu, T, Inada, T, Halter, D, Collart, M, Panasencko, O, Panasencko, O, Collart, M, et al. 2014. The Not5 Subunit of the Ccr4-Not Complex Connects Transcription and Translation. *PLoS Genet.* 10: e1004569.
- Viswanathan, P, Chen, J, Chiang, Y-C, Denis, CL. 2003. Identification of multiple RNA features that influence CCR4 deadenylation activity. *J. Biol. Chem.* 278: 14949–55.
- Waanders, E, Scheijen, B, van der Meer, LT, van Reijmersdal, S V, van Emst, L, Kroeze, Y, Sonneveld, E, Hoogerbrugge, PM, van Kessel, AG, van Leeuwen, FN, Kuiper, RP. 2012. The origin and nature of tightly clustered BTG1 deletions in precursor B-cell acute lymphoblastic leukemia support a model of multiclonal evolution. *PLoS Genet.* 8: e1002533.
- Waghray, S, Williams, C, Coon, JJ, Wickens, M. 2015. *Xenopus* CAF1 requires NOT1-mediated interaction with 4E-T to repress translation in vivo. *RNA* 21: 1335–45.
- Wagner, E, Clement, SL, Lykke-Andersen, J. 2007. An unconventional human Ccr4-Caf1 deadenylase complex in nuclear cajal bodies. *Mol. Cell. Biol.* 27: 1686–1695.
- Wahle, E, Winkler, GS. 2013. RNA decay machines: Deadenylation by the Ccr4-Not and Pan2-Pan3 complexes. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1829: 561–570.
- Wang, H, Morita, M, Yang, X, Suzuki, T, Yang, W, Wang, J, Ito, K, Wang, Q, Zhao, C, Bartlam, M, Yamamoto, T, Rao, Z. 2010. Crystal structure of the human CNOT6L nuclease domain reveals strict poly(A) substrate specificity. *EMBO J.* 29: 2566–76.
- Wang, S, Zhai, Y, Pang, X, Niu, T, Ding, Y-H, Dong, M-Q, Hsu, VW, Sun, Z, Sun, F. 2016. Structural characterization of coatomer in its cytosolic state. *Protein Cell* 7: 586–600.
- Wang, X, Lu, Z, Gomez, A, Hon, GC, Yue, Y, Han, D, Fu, Y, Parisien, M, Dai, Q, Jia, G, Ren, B, Pan, T, He, C. 2014. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505: 117–20.
- Washio-Oikawa, K, Nakamura, T, Usui, M, Yoneda, M, Ezura, Y, Ishikawa, I, Nakashima, K, Noda, T, Yamamoto, T, Noda, M. 2007. Cnot7-null mice exhibit high bone mass phenotype and modulation of BMP actions. *J. Bone Miner. Res.* 22: 1217–23.
- Watanabe, C, Morita, M, Hayata, T, Nakamoto, T, Kikuguchi, C, Li, X, Kobayashi, Y, Takahashi, N, Notomi, T, Moriyama, K, Yamamoto, T, Ezura, Y, Noda, M. 2014. Stability of mRNA influences osteoporotic bone mass via CNOT3. *Proc. Natl. Acad. Sci. U. S. A.* 111: 2692–7.
- Wente, SR, Rout, MP. 2010. The nuclear pore complex and nuclear

- transport. *Cold Spring Harb. Perspect. Biol.* 2: a000562.
- West, S, Proudfoot, NJ, Dye, MJ. 2008. Molecular dissection of mammalian RNA polymerase II transcriptional termination. *Mol. Cell* 29: 600–10.
- Wilson, CJ, Chao, DM, Imbalzano, AN, Schnitzler, GR, Kingston, RE, Young, RA. 1996. RNA Polymerase II Holoenzyme Contains SWI/SNF Regulators Involved in Chromatin Remodeling. *Cell* 84: 235–244.
- Winkler, GS. 2010. The mammalian anti-proliferative BTG/Tob protein family. *J. Cell. Physiol.* 222: 66–72.
- Winkler, GS, Balacco, DL. 2013. Heterogeneity and complexity within the nuclease module of the Ccr4-Not complex. *Front. Genet.* 4: 296.
- Winkler, GS, Mulder, KW, Bardwell, VJ, Kalkhoven, E, Timmers, HTM. 2006. Human Ccr4-Not complex is a ligand-dependent repressor of nuclear receptor-mediated transcription. *EMBO J.* 25: 3089–99.
- Wolf, J, Passmore, LA. 2014. mRNA deadenylation by Pan2-Pan3. *Biochem. Soc. Trans.* 42: 184–7.
- Wolf, J, Valkov, E, Allen, MD, Meineke, B, Gordiyenko, Y, Mclaughlin, SH, Olsen, TM, Robinson, C V, Bycroft, M, Stewart, M, Passmore, LA. 2014. Structural basis for Pan3 binding to Pan2 and its function in mRNA recruitment and deadenylation. *EMBO J.* 33: 1514–1526.
- Worth, CL, Preissner, R, Blundell, TL. 2011. SDM--a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res.* 39: W215-22.
- Wu, L, Liu, Y, Kong, D. 2014. Mechanism of chromosomal DNA replication initiation and replication fork stabilization in eukaryotes. *Sci. China. Life Sci.* 57: 482–7.
- Wu, M, Reuter, M, Lilie, H, Liu, Y, Wahle, E, Song, H. 2005. Structural insight into poly(A) binding and catalytic mechanism of human PARN. *EMBO J.* 24: 4082–93.
- Wu, X, Brewer, G. 2012. The regulation of mRNA stability in mammalian cells: 2.0. *Gene* 500: 10–21.
- Xiao, Z, Guifang, J. 2016. RNA epigenetic modification: N6-methyladenosine. *Yi chuan = Hered.* 38: 275–88.
- Xu, K, Bai, Y, Zhang, A, Zhang, Q, Bartlam, MG. 2014. Insights into the structure and architecture of the CCR4-NOT complex. *Front. Genet.* 5: 137.
- Yamashita, A, Chang, T-C, Yamashita, Y, Zhu, W, Zhong, Z, Chen, C-YA, Shyu, A-B. 2005. Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nat. Struct. Mol. Biol.* 12: 1054–1063.
- Yanagie, H, Tanabe, T, Sumimoto, H, Sugiyama, H, Matsuda, S, Nonaka, Y, Ogiwara, N, Sasaki, K, Tani, K, Takamoto, S, Takahashi, H, Eriguchi, M. 2009. Tumor growth suppression by adenovirus-mediated introduction of a cell-growth-suppressing gene tob in a pancreatic cancer model. *Biomed. Pharmacother.* 63: 275–86.
- Yang, E, van Nimwegen, E, Zavolan, M, Rajewsky, N, Schroeder, M, Magnasco, M, Darnell, JE. 2003. Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome Res.* 13: 1863–72.
- Yang, X, Morita, M, Wang, H, Suzuki, T, Yang, W, Luo, Y, Zhao, C, Yu, Y,

- Bartlam, M, Yamamoto, T, Rao, Z. 2008. Crystal structures of human BTG2 and mouse TIS21 involved in suppression of CAF1 deadenylase activity. *Nucleic Acids Res.* 36: 6872–6881.
- Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24: 1586–1591.
- Yang, Z, Nielsen, R, Hasegawa, M. 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol. Biol. Evol.* 15: 1600–11.
- Yang, Z, Wong, WSW, Nielsen, R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22: 1107–1118.
- Yates, CM, Filippis, I, Kelley, LA, Sternberg, MJE. 2014. SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features. *J. Mol. Biol.* 426: 2692–701.
- Yeoh, E-J, Ross, ME, Shurtleff, SA, Williams, WK, Patel, D, Mahfouz, R, Behm, FG, Raimondi, SC, Relling, M V, Patel, A, Cheng, C, Campana, D, Wilkins, D, Zhou, X, Li, J, Liu, H, Pui, C-H, Evans, WE, Naeve, C, Wong, L, Downing, JR. 2002. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 1: 133–43.
- Yoneda, M, Suzuki, T, Nakamura, T, Ajima, R, Yoshida, Y, Kakuta, S, Katsuko, S, Iwakura, Y, Shibutani, M, Mitsumori, K, Yokota, J, Yamamoto, T. 2009. Deficiency of antiproliferative family protein Ana correlates with development of lung adenocarcinoma. *Cancer Sci.* 100: 225–32.
- Yoshida, Y, Nakamura, T, Komoda, M, Satoh, H, Suzuki, T, Tsuzuku, JK, Miyasaka, T, Yoshida, EH, Umemori, H, Kunisaki, RK, Tani, K, Ishii, S, Mori, S, Suganuma, M, Noda, T, Yamamoto, T. 2003. Mice lacking a transcriptional corepressor Tob are predisposed to cancer. *Genes Dev.* 17: 1201–6.
- Zekri, L, Huntzinger, E, Heimstädt, S, Izaurralde, E. 2009. The silencing domain of GW182 interacts with PABPC1 to promote translational repression and degradation of microRNA targets and is required for target release. *Mol. Cell. Biol.* 29: 6220–31.
- Zhang, J, Nielsen, R, Yang, Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22: 2472–9.
- Zhang, X, Dai, D, Ou-yang, L, Yan, H. 2014. Detecting overlapping protein complexes b1. Zhang, X., Dai, D., Ou-yang, L. & Yan, H. Detecting overlapping protein complexes based on a generative model with functional and topological properties. 9–16 (2014).ased on a generative model with functional an. 9–16.
- Zhang, X, Virtanen, A, Kleiman, FE. 2010. To polyadenylate or to deadenylate: that is the question. *Cell Cycle* 9: 4437–49.
- Zhang, Z, English, BP, Grimm, JB, Kazane, SA, Hu, W, Tsai, A, Inouye, C, You, C, Piehler, J, Schultz, PG, Lavis, LD, Revyakin, A, Tjian, R. 2016. Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis. *Genes Dev.* 30: 2106–2118.

- Zheng, D, Ezzeddine, N, Chen, CYA, Zhu, W, He, X, Shyu, A Bin. 2008. Deadenylation is prerequisite for P-body formation and mRNA decay in mammalian cells. *J. Cell Biol.* 182: 89–101.
- Zhong, J, Chaerkady, R, Kandasamy, K, Gucek, M, Cole, RN, Pandey, A. 2011. The interactome of a PTB domain-containing adapter protein, Odin, revealed by SILAC. *J. Proteomics* 74: 294–303.
- Zhou, R, Shanas, R, Nelson, MA, Bhattacharyya, A, Shi, J. 2010. Increased expression of the heterogeneous nuclear ribonucleoprotein K in pancreatic cancer and its association with the mutant p53. *Int. J. cancer* 126: 395–404.
- Zuo, Y, Deutscher, MP. 2001. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res.* 29: 1017–26.

