



UNIVERSITAT DE
BARCELONA

Role of tRNA modifications in the synthesis of the extracellular matrix

Marta Rodríguez Escribà

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



UNIVERSITAT DE
BARCELONA



UNIVERSITAT DE BARCELONA
FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

Programa de doctorat en Biomedicina

INSTITUT DE RECERCA BIOMÈDICA DE BARCELONA

ROLE OF tRNA MODIFICATIONS IN THE SYNTHESIS OF THE EXTRACELLULAR MATRIX

Memòria presentada per Marta Rodríguez Escribà per a optar al grau de doctora
per la Universitat de Barcelona

Director

Lluís Ribas de Pouplana

Director

Adrián Gabriel Torres

Tutor

Antonio Zorzano Olarte

Doctoranda

Marta Rodríguez Escribà

2020

ACKNOWLEDGEMENTS

Voldria donar les gràcies a totes aquelles persones que, d'una manera o altra, han contribuït a l'elaboració d'aquesta tesi.

En primer lloc, vull agrair-te, Lluís, el fet d'haver-me donat l'oportunitat de fer el doctorat al teu laboratori. No sempre ha estat fàcil, hi ha hagut moments complicats! Però crec que m'emporto moltes coses bones del temps que hem treballat junts. De tu, em quedo sobretot amb la teva valentia per a començar projectes arriscats fora de la zona de confort.

Gracias Adrián por aceptar ser co-director de mi tesis, creo que ha sido toda una experiencia para los dos. Querría agradecerte el haber estado ahí, siempre disponible para echar una mano y resolver mis dudas, ya fuera en el terreno experimental o con tu conocimiento de la literatura.

Noelia! ¡Te mereces un párrafo entero! Arribar fins aquí hagués sigut molt més difícil sense tu. Gràcies per la teva contribució a molts dels experiments d'aquesta tesi, ja fos tenint cura de les meves cèl·lules, fent infinites extraccions de RNA, o participant en les maratonianes purificacions en tàndem. Estic segura que més d'una persona coincidirà amb mi quan dic que treballar amb tu dóna gust! Però gràcies també per fer-me riure, per tolerar la meva incontinència verbal i, és clar, pels nostres *lover's night* improshows!

Als companys i companyes del Ribas Lab, gràcies per les estones compartides! Als que ja hi eren des del principi, Daria, David, Fede i Àlbert, em vau acollir al laboratori tot ajudant-me a donar els primers passos d'aquesta tesi, un camí que em va portar al món de les modificacions dels tRNAs després d'una curta interacció amb SLIMP. A l'Adé, que tot i que ja estaves a Omnia quan jo vaig arribar, més d'una vegada em vas donar un cop de mà tant dins com fora del lab! Als que van venir després, Albert i Helena. Albert, moltes gràcies per tot, per les discussions sobre ADAT i SLIMP i, bé, sobre la vida en general, ja fos tard al laboratori, fent un cafè a en Paco, o dins de la càpsula que és cultius on sembla que el temps corri a un altre ritme. A l'Helena per la seva contribució al projecte del mutant d'ADAT3. I als que van venir més tard, Alba i Enric, gràcies per crear bon ambient al laboratori! A més a més, al llarg d'aquest temps, hi hagut també la presència més o menys esporàdica del col·lectiu Minion dins del Ribas lab. Gràcies especialment a la Núria Bosch (què fariem sense aquells "*Demoler*"?), a la Maria Carretero, a l'Aleksandra Fesiuk i a en Thomas Wulff.

Vull agrair també la contribució de totes aquelles persones amb qui he col·laborat durant el doctorat. Thanks to everyone in Sebastian Leidel's lab in Münster. To Juliane, Danny, Claudia and Benedikt for teaching me every step in the ribosome profiling ladder, from cell

lysis to data analysis. To Martin, Karin, Fiona, Puneet and Antonio for being such great lab mates. And, finally, to Sandra for sharing the newness of Münster inside and outside the lab with me. Gràcies a la Fàtima Gebauer per acollir-me al seu laboratori al CRG i a la Róisín-Ana per ensenyar-me a fer servir aquell fraccionador pre-històric que, de tant en tant, accedia a fer els *polysome profiles* com tocava. Treballant al CRG hi ha també l'Eva Novoa a qui vull agrair la seva ajuda en l'anàlisi de les dades de *ribosome profiling*, i el Gerard Cantero, que em va ajudar amb els experiments de secreció. Per últim, voldria donar les gràcies a l'Ana Pardo per la seva contribució als experiments amb les cèl·lules H292, tant per generar el KD d'ADAT2 en aquest model cel·lular com pels seus suggeriments i consells experimentals.

I no per ser "no-oficials" són menys importants les col·laboracions dins de l'IRB. Gracias a Elisa y Xavi por estar disponibles allí arriba, por dejarme reactivos cuando los míos aún no habían llegado, por explicarme técnicas que al principio me parecían magia oscura, y por todas las cenas en Gràcia. A les del Méndez lab, Berta, Clara, Rosa i Irene, pels moments i els protocols compartits. A les del Salvatella lab, Ela, Busra i Mirea, perquè els passadissos i la sala de centrífugues haguessin sigut molt menys interessants sense la vostra companyia. I finalment, als labs de l'Antonio Zorzano i del Manuel Palacín, perquè tots i totes m'heu donat un cop de mà quan feia falta. Però vull donar les gràcies especialment a l'Aida, la Sara i l'Alba, per les converses dins i fora del lab, pels esmorzars, dinars i sopars, pels viatges, per l'ajuda que sempre m'heu sabut donar quan us la demanava, en fi, gràcies per tantes i tantes coses!

Hi ha força gent fora del laboratori a qui vull agrair la seva presència constant durant els anys del doctorat. To Elena, Emil and Kyle, for all the fun times at conferences! To the Edinburgh family, Paul, Chas and Eliza, for keeping in touch through all that time, for hosting me during my short visits and for travelling south to Barcelona. A special thanks to Eliza, not only for having long skype sessions with me, but also for her excellent scientific art skills (www.lizawolfson.co.uk) and providing such a nice image for my thesis cover. A les de la uni, Natalia, Ari i Blanca, moltes gràcies per les converses interminables sobre tot i res, per aguantar-me ja sigui en persona, via àudio o vídeo, i, en definitiva, per ser-hi!

Finalment, voldria donar les gràcies als de casa, Pauli i Paula, per fer-me costat dia rere dia, per la vostra paciència, i per estimar-me. A la Carme, perquè tot i no ja ser-hi, estic convençuda que jo no seria on sóc avui si no fos per tu també. A very special thanks goes to Andrei, for your unconditional support and for believing in me every step of the way.

Gràcies!

SUMMARY

Transfer RNAs (tRNAs) are key adaptor molecules that mediate decoding of messenger RNAs (mRNAs) into proteins by complementary pairing of their anticodons with mRNA codons. tRNAs that undergo adenosine to inosine editing at the wobble base, or position 34, display expanded codon decoding capacity as inosine enables pairing not only with uridine, but also with cytosine and adenosine. The essential heterodimeric enzyme Adenosine Deaminase Acting on Transfer RNA (ADAT) catalyzes this post-transcriptional modification in eukaryotes and is comprised of subunits ADAT2 and ADAT3. Emergence of heterodimeric ADAT has been proposed to have shaped both tRNA gene content and codon composition of eukaryotic genomes in such a way that these two features became mirrored. Although the exact contribution of wobble inosine (I34) to translation elongation has not been established, previous reports have suggested that it might play a role in improving translational efficiency and accuracy of genes enriched in codons recognized by I34-modified tRNAs.

To further understand the role of the inosine modification in translation, we generated cell lines depleted in the catalytic subunit ADAT2. Silencing of ADAT2 lead to impaired cellular proliferation and had a variable impact on the expression of genes coding for extracellular matrix (ECM) proteins such as mucins. Notably, ADAT2 deficiency did not have major effects on the post-translational glycosylation of mucins, neither did it trigger the unfolded protein response. Supported by the absence of clear defects in decoding rates in ADAT2 depleted cells, as measured by ribosome profiling, our findings suggest that a reduced pool of I34-modified tRNAs might suffice to carry out cellular functions in steady-state conditions. However, we found that, under circumstances involving a high demand for these tRNAs such as airway remodeling, ADAT2 is required for the proper translation of an ECM gene enriched in stretches of codons read by I34-tRNAs. Taken together, our results suggest that the inosine modification is particularly relevant for the synthesis of ECM proteins during specialized processes including neural development and airway remodeling.

The importance of the inosine modification has been recently underscored by the identification of pathogenic mutations in the gene encoding ADAT3, all of which share common neurodevelopmental phenotypes. The most prevalent mutation identified to date is a valine to methionine (V144M) substitution that is linked to intellectual disability and strabismus. In the present study we characterized human ADAT in terms of activity and quaternary structure, and investigated the effect of the ADAT3 V144M mutation on the enzyme. We showed that the V144M substitution leads to decreased enzymatic activity of ADAT, which might result from alterations in the tertiary structure and subcellular localization of ADAT3 that were found to be associated to the mutation.

RESUM

Els ARNs de transferència (ARNt) són molècules que tenen un paper clau en el procés de traducció dels ARN missatgers (ARNm) en proteïnes mitjançant la interacció del seu anticodó amb codons d'ARNm. Els ARNt que passen per un procés d'edició d'adenosina a inosina a la base *wobble*, o posició 34, són capaços de llegir més d'un codó d'ARNm gràcies a la capacitat de la inosina de reconèixer els tres nucleòtids uridina, citidina i adenosina. L'enzim responsable d'aquesta modificació post-transcripcional en eucariotes s'anomena Adenosina Deaminasa específica per l'ARNt (ADAT), es tracta d'un complex heterodimèric format per les subunitats ADAT2 i ADAT3 que és essencial per a la viabilitat de l'organisme. Estudis previs han proposat que l'aparició d'ADAT va determinar el nombre de còpies gèniques de cada ARNt així com la composició de codons presents als genomes eucariòtics de tal manera que aquests dos factors estiguessin mútuament balancejats. Tot i que la contribució precisa de la inosina 34 (I34) a la traducció de proteïnes durant la fase d'elongació encara s'ha determinat experimentalment, algunes investigacions han suggerit que podria jugar un rol en l'eficiència i fidelitat de traducció de gens enriquits en codons reconeguts per ARNt modificats amb I34.

Amb l'objectiu d'investigar el rol de la inosina en la traducció, hem generat línies cel·lulars on el gen codificant per ADAT2 ha estat silenciada. La depleció d'ADAT2 comporta un retard en el creixement cel·lular i té un efecte variable en l'expressió gènica de proteïnes de la matriu extracel·lular. El patró de modificacions post-traduccional de glicosilació d'aquestes proteïnes no resulta alterat per la deficiència d'ADAT2, que tampoc activa la resposta a proteïnes desplegadas. Juntament amb l'absència de defectes en la velocitat d'elongació analitzada per *ribosome profiling*, aquestes observacions suggereixen que la cèl·lula és capaç de dur a terme les seves funcions amb un nombre reduït d'ARNt modificats amb inosina. Hem vist, però, que en condicions que requereixen majors quantitats d'ARNt inosinats, la depleció d'ADAT2 dona lloc a la traducció ineficient d'un gen de matriu extracel·lular altament enriquit en codons sensibles llegits per ARNt modificats. Així doncs, els nostres resultats indiquen que la inosina pot exercir un rol important en la síntesi de proteïnes de la matriu extracel·lular, particularment durant processos de desenvolupament neuronal i de remodelat de les vies respiratòries.

La rellevància de la modificació I34 s'ha vist reforçada recentment per la identificació de mutacions de caire patogènic localitzades al gen que codifica ADAT3. Totes elles tenen en comú la presència de fenotips relacionats amb el desenvolupament neurològic. La mutació d'ADAT3 més comuna consisteix en la substitució d'un residu valina per un metionina (V144M) i està associada a la manifestació de discapacitat intel·lectual i estrabisme. En el

present estudi hem caracteritzat l'activitat enzimàtica i l'estructura quaternària de l'ADAT humà, així com l'impacte de la mutació V144M d'ADAT3 en el complex heterodimèric. Els nostres resultats revelen que la substitució V144M dona lloc a una menor activitat enzimàtica d'ADAT. És possible que aquesta reducció es vegi influïda per les alteracions en l'estructura terciària i en la localització cel·lular d'ADAT3 que induïx la mutació.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	III
SUMMARY	VII
RESUM.....	IX
TABLE OF CONTENTS	XIII
LIST OF FIGURES.....	XVII
LIST OF TABLES.....	XIX
LIST OF ABBREVIATIONS.....	XXI
INTRODUCTION	1
1.1. Overview of gene expression	3
1.2. Eukaryotic translation.....	6
1.2.1. Components of the translation machinery.....	6
1.2.2. Translation initiation	7
1.2.3. Translation elongation	9
1.2.4. Translation termination and ribosome recycling.....	10
1.3. Modulators of translation: a focus on elongation rates.....	12
1.3.1. mRNA secondary structure	13
1.3.2. Amino acid signatures	14
1.3.3. Codon usage and tRNA pools	15
1.3.3.1. Translation efficiency.....	17
1.3.3.2. Translation accuracy and protein folding.....	18
1.3.3.3. Codon context	19
1.3.3.4. Genetic programs and tRNA dynamics.....	21
1.4. Transfer RNA	23
1.4.1. tRNA structure.....	23
1.4.2. tRNA biogenesis.....	24
1.4.3. tRNA modifications and editing.....	25
1.4.4. tRNA gene copy number, tRNA abundance, and codon usage bias.....	27
1.5. A-to-I editing at the tRNA wobble base.....	30
1.5.1. Functional implications of I34	31
1.5.2. I34 in human disease	33
1.5.3. I34-dependent codons cluster in extracellular matrix genes	34
1.6. Extracellular matrix.....	38

1.6.1. Mucins.....	38
1.6.2. Dystroglycan.....	41
1.6.3. Syndecan-3	43
MATERIALS AND METHODS	47
2.1. General laboratory material	49
2.2. General laboratory solutions.....	49
2.3. Oligonucleotides.....	49
2.4. Antibodies	53
2.5. Vectors.....	54
2.6. Cloning and site directed mutagenesis	55
2.7. Cell culture and <i>in vitro</i> assays	55
2.7.1. Cell lines and culture conditions.....	55
2.7.2. Generation of lentivirus and stable cell lines	56
2.7.3. Amphiregulin treatment.....	56
2.7.4. Thapsigargin treatment.....	57
2.7.5. Cell growth assay	57
2.8. Protein expression, detection and isolation.....	57
2.8.1. Recombinant protein expression and purification.....	57
2.8.2. Size-exclusion chromatography analysis	58
2.8.3. Mass Spectrometry analysis of proteins.....	59
2.8.4. Protein lysates and Western blotting.....	60
2.8.5. Dot blot analysis	61
2.8.6. Immunofluorescence.....	61
2.8.7. Live cell imaging	62
2.8.8. Flow cytometry.....	63
2.9. RNA extraction, detection and synthesis	63
2.9.1. RNA isolation and cDNA synthesis	63
2.9.2. Real-time qPCR.....	63
2.9.3. RT-PCR analysis of <i>XBP1</i> splicing.....	64
2.9.4. <i>In vitro</i> tRNA transcription and purification	64
2.9.5. Electrophoretic mobility gel shift assay (EMSA)	64
2.9.6. Deamination assays	65
2.10. Mucin isolation and glycosylation quantification.....	65
2.10.1. Mucin purification.....	65
2.10.2. Glycoprofiling of mucin-type O-glycans.....	67

2.11. Analysis of protein translation.....	68
2.11.1. Polysome profiling	68
2.11.2. Ribosome profiling.....	69
2.11.3. Ribosome profiling sequencing data analysis	71
2.12. Statistical analysis.....	72
OBJECTIVES.....	75
RESULTS	79
4.1. ACTIVITY AND STABILITY OF ADAT ARE COMPROMISED BY A MUTATION LINKED TO HUMAN INTELLECTUAL DISABILITY.....	81
4.1.1. Effect of an ADAT3 mutation on heterodimer structure and tRNA affinity	81
4.1.2. Characterization of the enzymatic activity of human ADAT	85
4.1.3. Subcellular localization of ADAT and impact of the V144M mutation	88
4.2. ROLE OF ADAT IN THE TRANSLATION OF ECM PROTEINS	96
4.2.1. ADAT2 silencing impairs cell proliferation.....	96
4.2.2. Depletion of ADAT2 does not activate the unfolded protein response	98
4.2.3. Assessing the impact of ADAT2 KD in translation by polysome profiling	99
4.2.4. Ribosome profiling does not show decoding defects in ADAT2 KD cells	104
4.2.5. ADAT2 KD has variable impact on the synthesis of extracellular matrix proteins.....	114
4.2.6. ADAT2 KD does not alter mucin glycosylation patterns	118
4.2.7. ADAT2 is required for MUC5AC protein synthesis in an <i>in vitro</i> cellular model of airway remodeling.....	122
DISCUSSION.....	129
CONCLUSIONS	143
REFERENCES	147

LIST OF FIGURES

Figure 1.1. Central dogma of molecular biology	3
Figure 1.2. Standard genetic code	4
Figure 1.3. Canonical and wobble base-pairing types between tRNA and mRNA	5
Figure 1.4. Major components of the eukaryotic translation machinery	7
Figure 1.5. Canonical cap-dependent eukaryotic translation initiation	8
Figure 1.6. Eukaryotic translation elongation	10
Figure 1.7. Translation termination and ribosome recycling in eukaryotes	11
Figure 1.8. <i>H. sapiens</i> codon usage.....	16
Figure 1.9. Ribosome profiling experimental workflow	18
Figure 1.10. tRNA structure	24
Figure 1.11. tRNA biogenesis	25
Figure 1.12. Distribution and frequency of human cytosolic tRNA modifications, and their link to disease.....	26
Figure 1.13. Human genomic tRNA copy number and codon usage.....	29
Figure 1.14. A-to-I conversion and human ADAT enzymes.....	31
Figure 1.15. ADAT2 silencing in HEK293T human cells.....	36
Figure 1.16. Gene ontology enrichment analysis of genes containing ADAT stretches	37
Figure 1.17. Mucin structure and oligomerization.....	39
Figure 1.18. Dystroglycan biosynthesis, structure and role in pathogenesis.....	42
Figure 1.19. Schematic representation of the four vertebrate syndecans.....	43
Figure 4.1. Analysis of wt ADAT quaternary structure	82
Figure 4.2. Analysis of mutant ADAT quaternary structure.....	84
Figure 4.3. Identification of the proteolytic cleavage site in ADAT3 V144M.....	84
Figure 4.4. ADAT binding affinity for tRNA.....	85
Figure 4.5. <i>In vitro</i> enzymatic activity of wt ADAT and ADAT* for tRNA ^{Ala} _{AGC} by RFLP	86
Figure 4.6. <i>In vitro</i> deamination assays to determine kinetic constants of wt and mutant ADAT for tRNA ^{Ala} _{AGC} by RFLP	87
Figure 4.7. Subcellular localization of endogenous ADAT2 and ADAT3.....	89
Figure 4.8. Subcellular localization of co-expressed wild type ADAT2 and ADAT3	90
Figure 4.9. Subcellular localization of wild type ADAT2 and ADAT3 using different fluorescent tags inserted at either end of the proteins.....	91
Figure 4.10. Impact of the V144M mutation upon the subcellular localization of ADAT3	92
Figure 4.11. Impact of the V144M mutation upon subcellular localization of ADAT3 expressed in neuroblastoma cells.....	93

Figure 4.12. Subcellular localization of ADAT3* using different fluorescent tags inserted at either end of ADAT3*	94
Figure 4.13. Expression levels of eGFP fusion proteins and insoluble nature of ADAT3 ...	95
Figure 4.14. ADAT2 depletion in several cell lines	96
Figure 4.15. Effect of ADAT2 silencing on cellular growth	97
Figure 4.16. UPR activation in CTRL and ADAT2 KD HEK293T and HT-29 M6 cells.....	99
Figure 4.17. Polysome profiles and associated mRNAs in pre- and post-confluent cultured HT-29 M6 cells.....	101
Figure 4.18. Polysome profiles and associated mRNAs in HEK293T under basal conditions and puromycin run-off assays	103
Figure 4.19. Replicability of mRNA and ribosome footprint data.....	105
Figure 4.20. Quality control of ribosome profiling data.....	106
Figure 4.21. Global ribosomal codon occupancy.....	107
Figure 4.22. Global ribosomal codon occupancy in HEK293T cells.....	109
Figure 4.23. Global ribosomal codon occupancy in HT-29 M6 cells	110
Figure 4.24. Global ribosomal codon occupancy in Cal27 cells.....	111
Figure 4.25. Global ribosomal codon occupancy in FaDu cells	112
Figure 4.26. Ribosome codon occupancy in genes containing ADAT-stretches	113
Figure 4.27. Mucin expression in HT-29 M6 cells.....	115
Figure 4.28. Dystroglycan expression in HT-29 M6 cells.....	117
Figure 4.29. Syndecan-3 expression in HEK293T cells.....	117
Figure 4.30. Purification of mucins from HT-29 M6 cells by two-step CsCl isopycnic density-gradient centrifugation.....	120
Figure 4.31. Mass spectrometry analysis of glycan structures present in mucins from HT-29 M6 cells.....	121
Figure 4.32. ADAT2 silencing in NCI-H292 cells	123
Figure 4.33. Immunostaining detection of MUC5AC in AREG-treated NCI-H292 cells....	124
Figure 4.34. Flow cytometry analysis of MUC5AC expression in AREG-treated NCI-H292 cells	125
Figure 4.35. Expression of mucous differentiation markers and AREG activation pathway in NCI-H292 cells.....	126
Figure 5.1. Distribution of the human proteins according to intrinsic disorder score.....	139

LIST OF TABLES

Table 1.1. Human genes presenting the highest local relative enrichment in ADAT-sensitive codons.....	37
Table 2.1. Oligonucleotides (5' → 3') used in this study.	50
Table 2.2. Oligonucleotides (5' → 3') used in this study for RT-PCR analysis.	51
Table 2.3. Oligonucleotides (5' → 3') used in this study for <i>in vitro</i> transcription of tRNA ^{Ala} _{AGC}	52
Table 2.4. Oligonucleotides (5' → 3') used in this study for ribosome profiling.	52
Table 2.5. Antibodies used in this study.	53
Table 2.6. Constructs used in this study.	54
Table 4.1. Deamination kinetic parameters of wt and mutant ADAT for tRNA ^{Ala} _{AGC}	87

LIST OF ABBREVIATIONS

2D	Two-dimensional
3'	Three prime
3D	Three-dimensional
5'	Five prime
A	Adenosine
Å	Angstrom
A-site	Aminoacyl-site
A.U.	Arbitrary units
A245	Absorbance at 245 nm
A260	Absorbance at 260 nm
A280	Absorbance at 280 nm
aa	Amino acid
aa-tRNA	Aminoacyl-tRNA
aaRS	Aminoacyl-tRNA synthetase
ACN	Acetonitrile
ADAR	Adenosine deaminase acting on mRNA
ADAT	Adenosine deaminase acting on tRNA
ADAT2	Adenosine deaminase tRNA specific 2
ADAT3	Adenosine deaminase tRNA specific 3
ADP	Adenosine-5'-diphosphate
AREG	Amphiregulin
ATP	Adenosine-5'-triphosphate
αDG	Dystroglycan subunit alpha
βDG	Dystroglycan subunit beta
BLAST	basic local alignment search tool
bp	Base pair
BSA	Bovine serum albumin
C	Cytidine
c	Centi-
C-ter	Carboxy terminal
cDNA	Complementary DNA
CDS	Coding sequence
CO ₂	Carbon dioxide
COPD	Chronic obstructive pulmonary disease
COX19	Cytochrome c oxidase assembly factor COX19
CsCl	Cesium chloride
Da	Dalton
DAG1	Dystroglycan 1
DAPI	4',6-Diamidino-2-phenylindole dihydrochloride
ddC	Dideoxycytidine
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
dn/dc	Refractive index increment
DNA	Deoxyribonucleic acid

DNase	Deoxyribonuclease
DPPIV	Dipeptidyl peptidase 4
dsDNA	Double-stranded DNA
dT	Deoxythymine
DTT	Dithiothreitol
E-site	Exit-site
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
eEF	Eukaryotic elongation factor
eGFP	Enhanced green fluorescent protein
EGFR	Epidermal growth factor receptor
EGTA	Ethylene glycol-bis(β -aminoethyl ether)-N,N,N',N'-tetraacetic acid
eIF	Eukaryotic initiation factor
ELISA	Enzyme-linked immunosorbent assay
EMEM	Eagle's Minimum Essential Medium
EMSA	Electrophoretic mobility gel shift assay
ER	Endoplasmic reticulum
eRF	Eukaryotic recycling factor
ERK	Extracellular-signal-regulated kinase
ESI-MS	Electrospray ionization mass spectrometry
F	Fucose
FA	Formic acid
FBS	Fetal bovine serum
FDR	False discovery rate
Flow	Flow cytometry
FPLC	Fast protein liquid chromatography
G	Guanosine
g	Gram
<i>g</i>	G-force
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GCNT3	Glucosaminyl (N-acetyl) transferase 3, mucin type
GDP	Guanosine-5'-diphosphate
GTP	Guanosine-5'-triphosphate
GU	Glucose unit
GuHCl	Guanidinium chloride
h	Hour
H	Hexose
HCl	Hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HILIC-UHPLC	Hydrophilic interaction chromatography ultra-high-performance liquid chromatography
HRP	Horseradish peroxidase
I	Inosine
I34	Inosine at position 34 of tRNA
IAM	Iodoacetamide
IF	Immunofluorescence
Ig	Immunoglobulin

IP	Immunoprecipitation
IPTG	Isopropyl β -D-1-thiogalactopyranoside
k	Kilo-
k_{cat}	Catalytic constant
KCl	Potassium chloride
KD	Knockdown
K_m	Michaelis-Menten constant
KOH	Potassium hydroxide
L	Liter
LB	Luria-Bertrani
LC-MS/MS	Liquid chromatography-coupled tandem quadrupole mass-spectrometry
μ	Micro-
m	Milli-
M	Molar (mol/L)
MALS	Multi-angle light scattering
MCS	Multiple cloning site
METTL7A	Methyltransferase like 7A
$MgCl_2$	Magnesium chloride
min	Minute
mol	Mole
mRNA	Messenger RNA
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MUC1	Mucin 1
MUC5AC	Mucin 5AC
MUC5B	Mucin 5B
Mw	Molecular weight
n	Nano-
N	N-acetylhexosamine
N-ter	Amino-terminal
Na_3VO_4	Sodium orthovanadate
NaCl	Sodium chloride
NaF	Sodium fluoride
NCBI	National Center for Biotechnology Information
NEM	<i>N</i> -ethylmaleimide
NH_4Cl	Ammonium chloride
nt	Nucleotide
O-Ac	O-acetyl
$^{\circ}C$	Degrees Celsius
OD_{600}	Optical density at 600 nm
ORF	Open reading frame
p	Pico-
P-site	Peptidyl-site
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline-tween

PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PEI	Polyethylenimine
Phos	Phosphorylation
PIC	Pre-initiation complex
PMSF	Phenylmethylsulphonylfluoridate
PROC	Procainamide
psi	Pounds per square inch
PVDF	Polyvinylidene fluoride
rApp	5' adenylate
RBS	Ribosome binding site
RFLP	Restriction Fragment Length Polymorphism
RI	Refractive index
RIPA	Radioimmunoprecipitation assay
RNA	Ribonucleic acid
RNase	Ribonuclease
RPF	Ribosome protected fragment
rpm	Revolutions per minute
RPMI 1640	Roswell Park Memorial Institute 1640 medium
rRNA	Ribosomal RNA
RT	Retrotranscription
RT-qPCR	Real-time quantitative PCR
s	Second
S	Sialic acid Neu5Ac
SDC3	Syndecan 3
SDHA	Succinate dehydrogenase complex flavoprotein subunit A
SDS	Sodium dodecyl sulfate
SEC	Size-exclusion chromatography
shRNA	Short hairpin RNA
siRNA	Small interference RNA
SNP	Single-nucleotide polymorphism
SpC18	Hexa-ethyleneglycol spacer
SPDEF	SAM pointed domain containing ETS transcription factor
ssDNA	Single-stranded DNA
T	Thymine
<i>T. Thermophila</i>	<i>Tetrahymena thermophila</i>
TadA	tRNA adenosine deaminase A
TBE	Tris-borate-EDTA buffer
TBS	Tris buffered saline
TBS-T	Tris buffered saline-tween
TC	Ternary complex
TEMED	N,N,N',N'-Tetramethylethylenediamine
TFA	Trifluoroacetic acid
Tris	Tris-(hydroxymethyl) aminomethane
tRNA	Transfer RNA
tRNA ^{aa} _{NNN}	tRNA specific for aa bearing the codon NNN

U	Uridine
UTR	Untranslated region
UV	Ultraviolet
V	Volt
V _e	Elution volume
V _{max}	Maximum velocity
V _o	Void volume
w/v	weight per volume
WB	Western blot
wt	Wild type

Amino acids

A	Ala	Alanine
R	Arg	Arginine
N	Asn	Asparagine
D	Asp	Aspartic acid
C	Cys	Cysteine
E	Glu	Glutamic acid
Q	Gln	Glutamine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
L	Leu	Leucine
K	Lys	Lysine
M	Met	Methionine
F	Phe	Phenylalanine
P	Pro	Proline
S	Ser	Serine
T	Thr	Threonine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
V	Val	Valine

1. INTRODUCTION

1.1. Overview of gene expression

The central dogma of molecular biology defines the flow of genetic information within biological systems and states that once this sequential information has passed into protein it cannot get out again (**Figure 1.1**). This concept was first published as part of an article entitled “On protein synthesis” by Francis H. Crick in which key ideas about gene function were outlined (Crick, 1958). Crick proposed that deoxyribonucleic acid (DNA) was the molecule responsible for storing the genetic information that contained the instructions for protein synthesis. However, he speculated that the information stored in DNA was first transferred, or transcribed, to a mediator molecule later known as the ribonucleic acid (RNA) copy of that DNA or messenger RNA (mRNA) (Brenner et al., 1961), which served as a template to direct the correct order in which the 20 most commonly occurring amino acids were assembled into proteins. Three years earlier, on a privately circulated note entitled “On degenerate templates and the adaptor hypothesis” (Crick, 1955) to the RNA Tie Club, Crick had anticipated the discovery of transfer RNA (tRNA) and aminoacyl-tRNA synthetases (aaRS). It was suggested that amino acids were carried to the mRNA template in the ribosome by at least twenty different kinds of “adaptor” molecule (tRNA) –conceiving the possibility that there were more than one adaptor molecule per amino acid– and that twenty specific enzymes (aaRS) were required to join the amino acids to their corresponding adaptors. In his 1958 article, Crick envisioned that the aminoacyl-tRNA, charged with its cognate amino acid, may interact with the mRNA template through base-pairing for protein synthesis to take place (Crick, 1958), a process currently known as decoding.

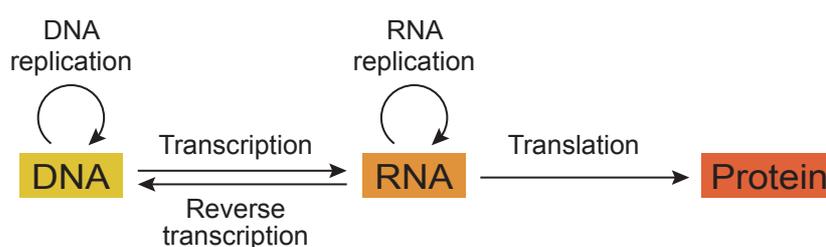


Figure 1.1. Central dogma of molecular biology. The central dogma provides a framework for how genetic information is transferred between DNA, RNA, and protein. Arrows represent the directional flow of sequence information. Adapted from (Crick, 1956, 1970).

Cracking how the genetic code defines the rules of translation from nucleotides to amino acids took another three years (Crick et al., 1961; Nirenberg and Matthaei, 1961). Today we know that it follows a non-overlapping three-nucleotide code system formed by combinations of four different bases –adenine (A), guanine (G), cytosine (C), and thymine (T) in DNA or

uracil (U) in RNA— that yield 64 possible triplets or codons (**Figure 1.2**). Of those, three act as translation stop signals while the other 61 encode a pool of 20 amino acids. The difference in number of codons and amino acids implies that the code is degenerate, that is, most amino acids are encoded by multiple codons. The codons that differ in sequence but are translated into the same amino acid are termed synonymous. With very few exceptions, the basic structural patterns of the genetic code are universal to all domains of life (Vetsigian et al., 2006). The genetic code table depicted in **Figure 1.2** is nonrandomly organized, possessing two types of order: synonymous order and relatedness order. The first refers to the general organization of the 64 codons in sets of two or four codons assigned to a specific amino acid, with the third base of each codon set being synonymous. The second alludes to the apparent trend to occupy contiguous areas of the table by related amino acids. Of particular interest is the role of the second position of a codon as a specificity determinant in terms of amino acid physico-chemical properties. For example, all codons with U at the second position specify hydrophobic amino acids. Thus, given that changes in the third and first position of the codon are likely to preserve the amino acid assignment or lead to incorporation of an amino acid with similar properties to the correct one, respectively, the probability of mutations or translation errors having a deleterious effect is minimized by the structure of the genetic code (Crick, 1968; Koonin and Novozhilov, 2017; Woese et al., 1966).

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G
						Third letter

Figure 1.2. Standard genetic code. Assignment of the 64 codons to their corresponding amino acid or translation stop signal.

As speculated by Crick in 1958, each of the mRNA sense codons are decoded through complementary base-pairing with the trinucleotide region of the tRNA molecule termed

anticodon. Such interaction can be achieved by means of canonical Watson-Crick hydrogen bonding, which generates A:U and G:C pairs (**Figure 1.3 A**). Additionally, nonstandard (non-Watson-Crick or wobble) pairing between the first base of the tRNA anticodon (position 34 or wobble base) and the third base of the mRNA codon is also possible, which gives rise to U:G, I:A, I:C, and I:U pairs (**Figure 1.3 B**), where “I” represents the modified base inosine (Crick, 1966). Correct interactions between the mRNA and the tRNA are necessary for protein synthesis to progress smoothly and accurately.

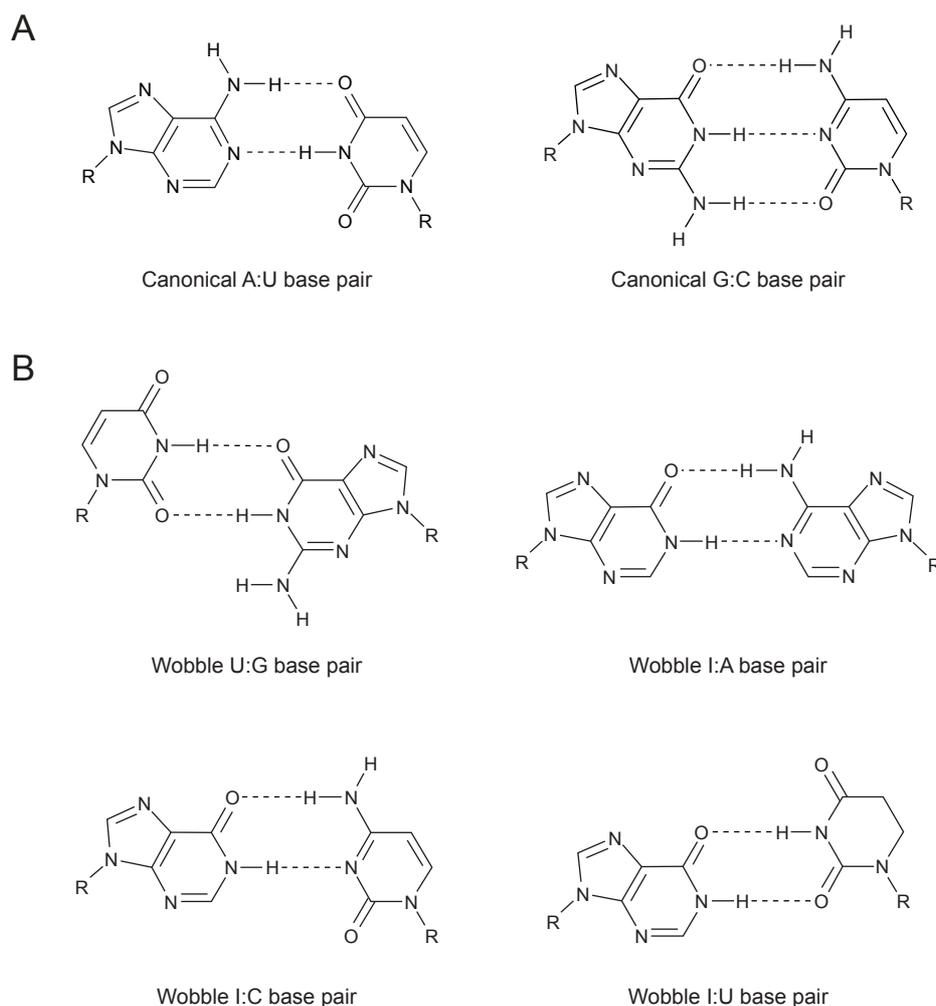


Figure 1.3. Canonical and wobble base-pairing types between tRNA and mRNA. (A) Canonical Watson-Crick A:U and G:C base pairs. (B) Wobble U:G, I:A, I:C, and I:U base pairs. Dotted lines indicate hydrogen bonds.

1.2. Eukaryotic translation

Protein synthesis is a complex process whereby amino acids are linked into peptide chains as dictated by a sequence encoded in the genome. A distinct feature of eukaryotic protein synthesis is its uncoupling from transcription, the physical compartmentalization of transcription in the nucleus and translation in the cytoplasm allows for further layers of translation regulation.

1.2.1. Components of the translation machinery

The major components of the eukaryotic translation machinery are transfer RNAs (tRNAs), aminoacyl-tRNA synthetases (aaRSs), the small (40S) and large (60S) ribosomal subunits, mRNA, numerous protein factors, and GTP and ATP as energy sources (**Figure 1.4**). As key adaptor molecules that mediate the decoding of mRNAs into proteins, tRNAs are first linked to their cognate amino acid in a aminoacylation reaction carried out by the enzymes aminoacyl-tRNA synthetases (**Figure 1.4 A**). The resulting aminoacyl-tRNAs (aa-tRNAs) will then be delivered to the ribosome to pair with their cognate mRNA codons. The 40S and 60S ribosomal subunits come together to form a fully functional 80S ribosome, a universally conserved ribonucleoprotein complex where protein synthesis takes place and that has, in its core, three specific sites tRNAs can bind, the aminoacyl (A), peptidyl (P) and exit (E) sites, each one of them carrying different roles during translation (**Figure 1.4 B**). The two ribosomal subunits differentiate from each other both in composition and in function. The human 40S subunit consists of a single 18S ribosomal RNA (rRNA) and 33 proteins, and it hosts the binding site for mRNA. Whereas the 60S subunit, with three ribosomal RNAs (28S, 5S and 5.8S) and 47 proteins, possesses the catalytic peptidyl transferase center and the GTPase-associated center (Anger et al., 2013; Khatter et al., 2015; Yusupova and Yusupov, 2014). In order to translate the nucleotide code of RNA into the amino acid alphabet of proteins, mRNAs are decoded by aa-tRNAs through base-pairing interactions. These occur in the ribosome, where mRNA codons are recognized by complimentary sequences of three nucleotides located in the anticodon region of the aa-tRNA, and the charged tRNA delivers the next residue of the nascent protein chain (**Figure 1.4 B**). The multistep process through which the translation machinery achieves efficient protein synthesis can be divided into four main phases: initiation, elongation, termination, and ribosome recycling.

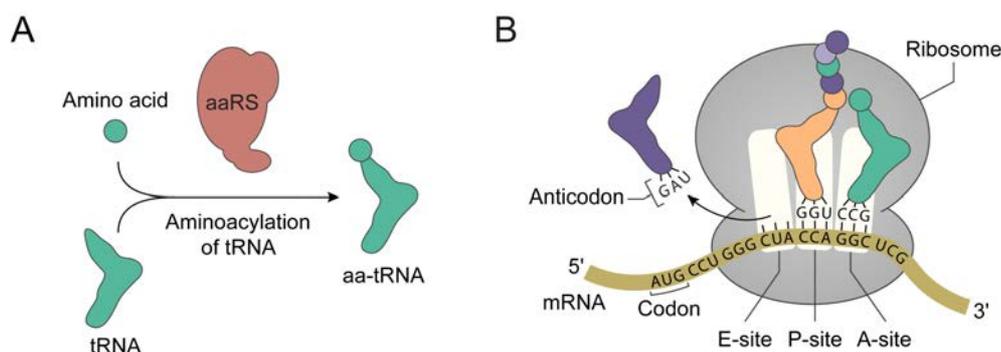


Figure 1.4. Major components of the eukaryotic translation machinery. (A) Depiction of the aminoacylation of tRNA, in which a tRNA is paired with its cognate amino acid by the enzyme aminoacyl-tRNA synthetase (aaRS). (B) Schematic representation of the RNA-binding sites in the ribosome. Each ribosome has one binding site for mRNA and three binding sites for tRNA: the acceptor site (A-site), where the aminoacyl-tRNA (aa-tRNA) is placed; the peptidyl site (P-site), which contains the tRNA with the growing polypeptide; and the exit site (E-site), where the empty tRNA exits the ribosome.

1.2.2. Translation initiation

Initiation of protein synthesis in eukaryotes is a tightly regulated process that, under most conditions, becomes the rate-limiting step of translation with about 1.3 to 3.6 initiation events per minute, compared to the average elongation rate of 3 to 12 amino acids per second (Ingolia et al., 2011; Morisaki et al., 2016; Sharma et al., 2019; Wu et al., 2016; Yan et al., 2016). Canonical translation initiation is shown schematically in **Figure 1.5**, it involves at least twelve eukaryotic initiation factors (eIFs) that participate in the assembly of elongation-competent 80S ribosomes on the mRNA, where the AUG start codon is base-paired with the anticodon of initiator tRNA ($\text{Met-tRNA}_i^{\text{Met}}$) in the ribosomal P-site (Hinnebusch, 2011; Jackson et al., 2010; Merrick and Pavitt, 2018; Pestova et al., 2007; Sonenberg and Hinnebusch, 2009).

The identification of the start codon takes place by a scanning mechanism and it begins with the association of $\text{Met-tRNA}_i^{\text{Met}}$, in a ternary complex (TC) with GTP-bound eIF2, with eIF5 and the 40S subunit bound to eIFs 1, 1A and 3, therefore forming a larger 43S preinitiation complex (PIC). mRNAs are activated by binding, at their 5' cap, to the eIF4F complex formed by the cap-binding protein eIF4E, eIF4G and the RNA helicase eIF4A. eIF4F is, in turn, associated to eIF4B and to the poly(A)-binding protein (PABP), which binds the tRNA at its 3' forming a "closed-loop" circularized structure (Park et al., 2011). Unwinding of secondary structures in the 5' untranslated region (UTR) of the mRNA is carried out by eIF4A and it facilitates recruitment of the 43S PIC to the cap via eIF3 and eIF4G interaction (Kumar et al., 2016; Pestova et al., 2007; Villa et al., 2013). Once bound, the 43S PIC scans the

mRNA 5' UTR in a 3' direction searching for the AUG initiation codon. Recognition of AUG through perfect base pairing with the anticodon of P-site-bound Met-tRNA^{Met} triggers the arrest of scanning and irreversible hydrolysis of GTP bound to eIF2, producing a stable 48S initiation complex (Algire et al., 2005). Release of GDP-bound eIF2 and other eIFs allows for the 60S subunit to join and form a 80S ribosome containing Met-tRNA^{Met} base paired to AUG in the P-site ready to begin the elongation phase.

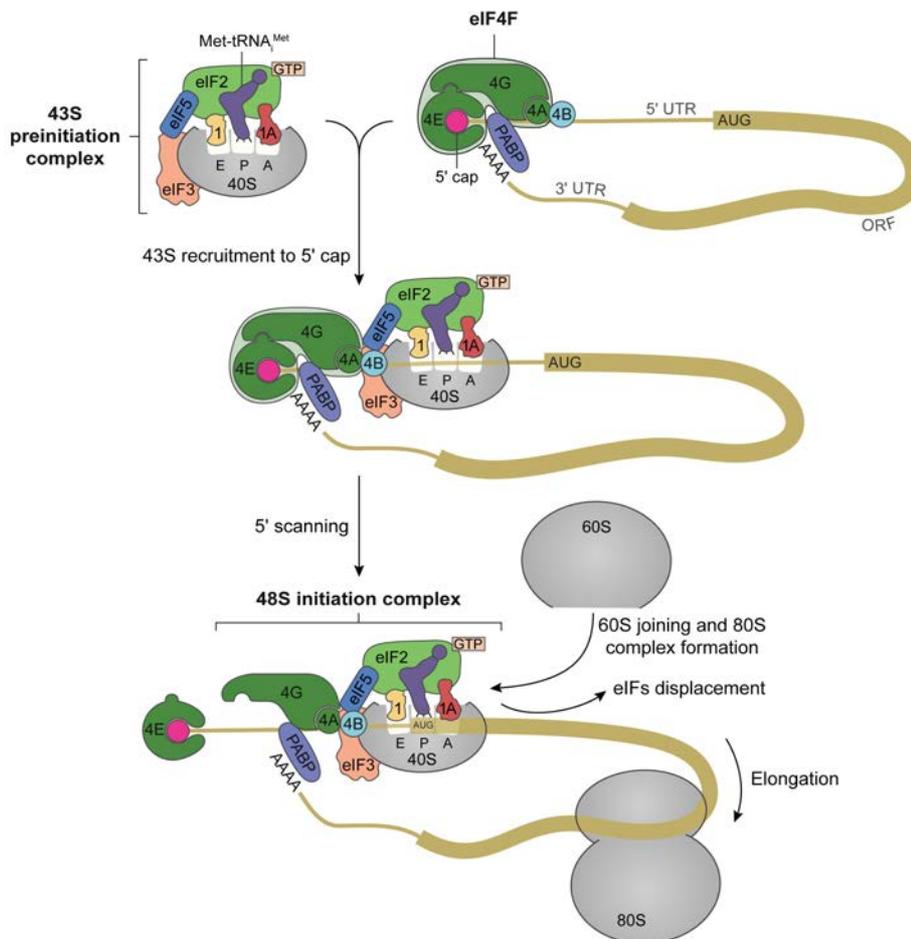


Figure 1.5. Canonical cap-dependent eukaryotic translation initiation. The first step of initiation is the assembly of the eIF4F complex (eIF4A, eIF4E, and eIF4G) on the mRNA 5' cap, wherein eIF4G interacts with Poly(A)-binding protein (PABP) on the 3', thereby promoting the circularization of the mRNA. The 43S preinitiation complex, which is formed by the association of the recycled 40S subunit and eukaryotic initiation factors eIF1, eIF1A, eIF3, and eIF5, to the ternary complex of GTP-bound eIF2 and initiator methionyl-tRNA, is then recruited to the mRNA 5' cap via eIF4G:eIF3 interaction. The association of the 43S to the mRNA leads the formation of the 48S initiation complex competent for scanning the 5' UTR towards the start codon (AUG). Binding of the 48S complex at the start codon triggers hydrolysis of the eIFs, and 60S large ribosomal recruitment to the 40S subunit to form the elongation-competent 80S ribosome. Abbreviations: 1, eIF1; 1A, eIF1A; 4B, eIF4B; 4E, eIF4E; 4G, eIF4G; ORF, open reading frame.

1.2.3. Translation elongation

Once docked onto the AUG start codon, and with a vacant A-site, the 80S ribosome starts moving along the mRNA one codon at a time while aminoacyl-tRNAs transition through the three ribosomal sites helped by several eukaryotic elongation factors (eEFs) (Dever et al., 2018; Rodnina, 2016; Schuller and Green, 2018). Translation elongation comprises three steps: tRNA selection or decoding, peptide-bond formation, and translocation (**Figure 1.6**).

During decoding, aminoacyl-tRNAs in a ternary complex with GTP-bound eEF1A are delivered to the A-site of the ribosome; however, only aminoacyl-tRNAs that are cognate to the mRNA codon will be selected. Correct codon–anticodon interaction triggers GTP hydrolysis by eEF1A, leading to GDP-bound eEF1A release and accommodation of the aminoacyl-tRNA into the A-site (Shao et al., 2016). Peptide-bond formation is then catalyzed by the peptidyl-transfer center within the 60S subunit, where the amino group of the amino acid bound to the A-site tRNA attacks the ester linkage on the peptidyl-tRNA in the P-site. This reaction results in the transfer of the nascent peptide chain to the A-site tRNA and the deacylation of the P-site tRNA (Beringer and Rodnina, 2007). Once peptide-bond is formed, translocation of the mRNA-tRNA complexes is required to resume elongation. Translocation involves the rotation of the ribosomal subunits with respect to each other, wherein tRNAs are repositioned into a “hybrid” state such that their anticodons remain in the P and A-sites, while their acceptor ends are located in the E and P-sites (P/E and A/P states, respectively) (Behrmann et al., 2015; Budkevich et al., 2011; Moazed and Noller, 1989). GTP-bound eEF2 then binds the A-site of the ribosome promoting ribosomal structural rearrangements that accompany tRNA movement into the classical E/E and P/P states (Ferguson et al., 2015; Flis et al., 2018; Spahn et al., 2004; Taylor et al., 2007). The elongation cycle continues until a stop codon is reached, signaling the end of translocation.

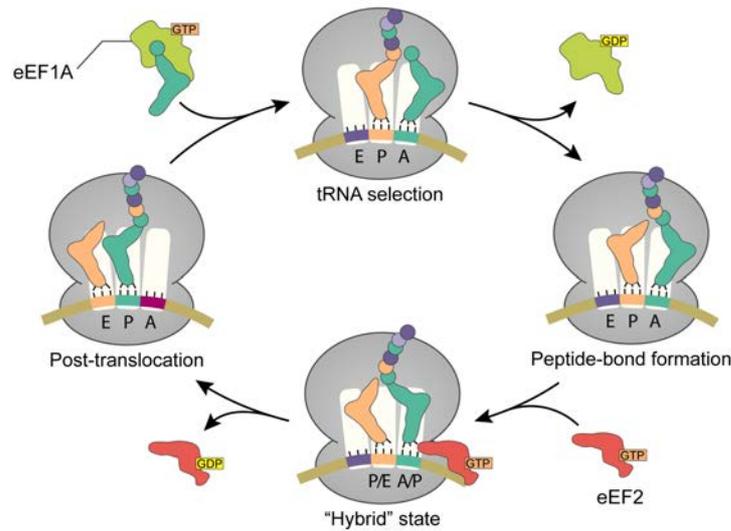


Figure 1.6. Eukaryotic translation elongation. During translation elongation, an aminoacyl-tRNA binds the vacant A-site on the ribosome. Then, the amino group of the incoming amino acid (green) attacks the ester linkage on the peptidyl-tRNA (orange) in the ribosomal P-site to transfer the growing peptide chain to the A-site tRNA. Following peptide bond formation, tRNAs are translocated, the empty tRNA is positioned in the E-site and the peptidyl-tRNA is now found in the P-site. The A-site becomes available for loading the next aminoacyl-tRNA and the cycle begins again.

1.2.4. Translation termination and ribosome recycling

Termination of translation is triggered by the entrance of a stop codon in the ribosomal A-site and it is mediated by the eukaryotic peptide chain release factors (eRFs) eRF1 and eRF3, which promote the dissociation of the nascent protein from the ribosome. The resulting post-termination complex comprising the 80S ribosome bound to deacylated tRNA in the P-site, mRNA and eRF1 is then recycled by the conserved ATP-binding cassette subfamily E member 1 (ABCE1) protein into separate subunits that will participate in new rounds of translation (**Figure 1.7**) (Hellen, 2018; Jackson et al., 2012; Schuller and Green, 2018).

In eukaryotes, only one release factor is needed for stop codon recognition. eRF1 binds all three stop codons (UAA, UAG, and UGA) in the A-site and is responsible for the hydrolysis of the peptide chain from the peptidyl-tRNA (Frolova et al., 1994). eRF1 performs this dual function through its three different structural domains: an amino-terminal domain that recognizes the three termination codons, a middle domain with a conserved catalytic GGQ motif that enters the peptidyl-transferase center in the 60S to induce polypeptide release, and a carboxy-terminal domain that binds to both eRF3 and ABCE1 (Bertram et al., 2000; Frolova et al., 1999; Mantsyzov et al., 2010; Song et al., 2000). In order to carry out its role, eRF1 forms a ternary complex with GTP-bound eRF3 (Alkalaeva et al., 2006; Frolova et al.,

1996), a member of the translational GTPases family that includes eEF1A, bacterial EF-Tu, and eukaryotic ribosome rescue factor Hbs1 (Atkinson et al., 2008; Shao et al., 2016). Once the eRF1-eRF3·GTP complex engages the ribosome through stop codon recognition by eRF1, hydrolysis of eRF3-bound GTP promotes eRF3 dissociation and induces conformational changes in eRF1 that allow its GGQ motif to accommodate in the peptidyl-transferase center, close to the CCA of the peptidyl-tRNA, where it coordinates a water molecule that hydrolyses the growing peptide chain from the tRNA (Brown et al., 2015; Preis et al., 2014; Shao et al., 2016). Release of the peptide chain from the ribosome signs the end of termination.

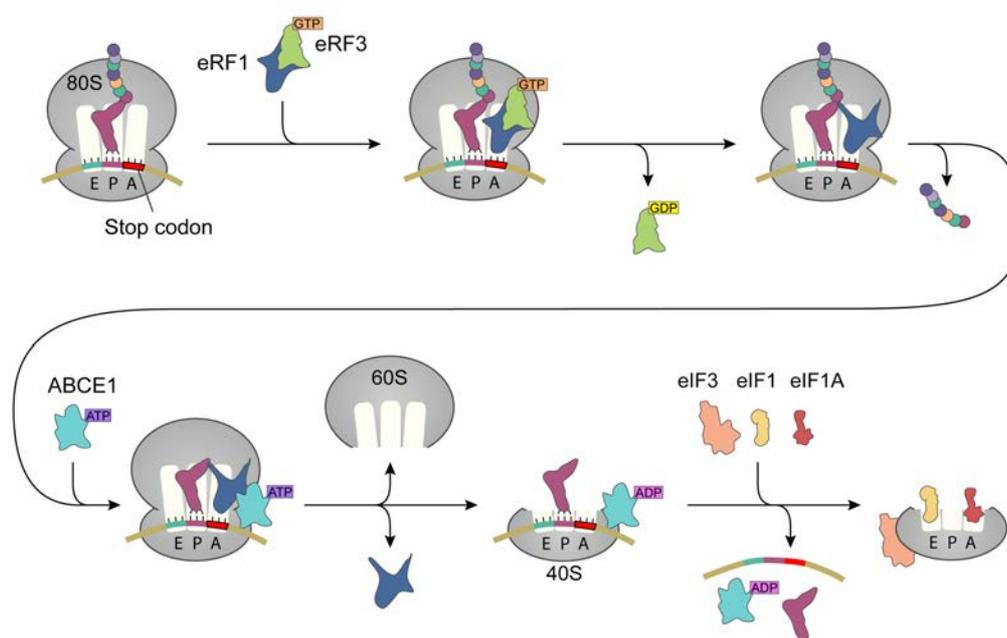


Figure 1.7. Translation termination and ribosome recycling in eukaryotes. When the elongating ribosome encounters a stop codon, release factor eRF1 enters the A-site as part of a pre-assembled ternary complex comprising eRF1 and GTP-bound eRF3. Hydrolysis of bound GTP induces eRF1-mediated cleavage of the peptide chain from the peptidyl-tRNA, forming a post-termination complex. ABCE1 ATPase then facilitates the dissociation of the ribosomal subunits. Finally, eukaryotic initiation factors eIF1, eIF1A, and eIF3 bind the released 40S subunit, displacing the mRNA and the peptidyl-tRNA.

Subsequent disassembly of post-termination complexes is accomplished by the essential protein ABCE1 in collaboration with eRF1, which remains in the ribosomal A-site after termination (Pisarev et al., 2010; Preis et al., 2014). ATP hydrolysis by ABCE1 promotes dissociation of the ribosome into free 60S and mRNA- and P-site deacylated tRNA-bound 40S subunits (Pisarev et al., 2010; Shoemaker and Green, 2011). Release of P-site tRNA and mRNA requires binding of eIF1, eIF1A and eIF3 to the 40S

subunit's solvent side and enhances mRNA release, eIF1 and eIF1A bind the subunit's interface in an adjacent position to the P-site and in the A-site, respectively (Fraser et al., 2007; Des Georges et al., 2015; Pisarev et al., 2007). eIF1 mediates the release of the tRNA and also discriminates against non-initiator tRNAs preventing their binding to the P-site (Lomakin et al., 2006). The binding of eIFs is also thought to prevent re-association with 60S subunits. An alternative mechanism for tRNA/mRNA release from the 40S involves the participation of ligatin (eIF2D) or malignant T-cell amplified sequence 1 (MCT-1) together with density regulated protein (DENR) (Skabkin et al., 2010).

1.3. Modulators of translation: a focus on elongation rates

The cellular levels of proteins are determined by a balance between transcription rate, mRNA processing and stability, translation speed, and protein degradation. The translation rate of a given gene is determined by the number of ribosomes that bind its mRNA and the speed of peptide-bond formation, which can be influenced by a range of factors including mRNA signatures, and availability of aminoacyl-tRNAs and translation factors (Gingold and Pilpel, 2011; Li et al., 2014). With protein synthesis being a complex multi-step choreographed process to which cells devote a considerable amount of resources (Rolfe and Brown, 1997), translational regulation is required in order to ensure energy conservation and, equally important, protein homeostasis.

While it is generally assumed that protein synthesis may be predominantly regulated at the initiation step –which is consistent with the idea that it is energetically more efficient to control a process at its outset rather than interrupting it midstream, considering the accumulation of intermediates that this would entail–, several studies have described cases in which regulation occurs at the elongation and termination stages (Rodnina, 2016; Schuller and Green, 2018). Of note, a substantial decrease in elongation rates, when occurring at the beginning of the mRNA, is also likely to affect initiation, given that the ribosome must move downstream the start codon in order for the next one to bind (Chu et al., 2014; Woolstenhulme et al., 2015).

Variations in elongation rates can shape the composition and quality of the cellular proteome. As a general idea, an increase in translational speed could allow for rapid ribosome turnover, therefore boosting protein production. However, it is known that the

process of elongation is not uniform along a transcript (Ingolia et al., 2009; Tuller et al., 2010a) and changes in local elongation speed can cause alterations in the quality of protein synthesis by affecting translation fidelity and protein folding, as well as gene expression by mediating mRNA decay and protein quality control pathways. Although it is challenging to assign such effects to individual mRNA signatures, given that most of them arise from combinations of different factors, further understanding of these features is currently growing and it is becoming more evident that local changes in elongation kinetics have a critical role in modulating the pathophysiology of human disease (Richter and Collier, 2015). Burst and pauses in elongation rates can originate from three main elements: mRNA secondary structure, amino acid signatures, and codon usage and tRNA pools.

1.3.1. mRNA secondary structure

mRNA secondary structures located around the start codon have been described to impede translation initiation and, therefore, serve as indicators of translation efficiency (Gray and Hentze, 1994; Kozak, 2005; Kudla et al., 2009; Pop et al., 2014; Tuller et al., 2010b). However, the function of mRNA structural elements within the CDS in translation elongation is less clear. The emergence of high-throughput techniques for RNA structure probing and determination during the last decade has yielded a genome-wide glimpse on the 'RNA structurome', which revealed regulatory effects of mRNA secondary structure in a large number of transcripts (Kwok et al., 2015; Mortimer et al., 2014; Strobel et al., 2018).

Two common features in all organisms studied to date are the presence of weak structure near the start codon (Beaudoin et al., 2018; Burkhardt et al., 2017; Del Campo et al., 2015; Deng et al., 2018; Ding et al., 2014; Li et al., 2014; Spitale et al., 2015) and the observation that mRNA is less structured *in vivo* than *in vitro* (Archer et al., 2013; Beaudoin et al., 2018; Burkhardt et al., 2017; Rouskin et al., 2014; Spitale et al., 2015). This decrease in structure *in vivo* is argued to be due to the helicase activity of the ribosome that unwinds the mRNA as it translocates (Beaudoin et al., 2018; Burkhardt et al., 2017; Takyar et al., 2005; Wen et al., 2008). Thus, when decoding is fast, the elongation rate of the ribosome may be contingent on structural elements present in the transcript. Another widespread, and poorly understood, characteristic is the presence of a triplet periodic structural signal in the CDS but not in UTRs that has been found both *in vitro* and *in vivo* (Del Campo et al., 2015; Deng et al., 2018; Ding et al., 2014; Incarnato et al., 2014; Kertesz et al., 2010; Spitale et al., 2015; Wan et al., 2014). The signal is consistent with previous genome-wide computational structural predictions for human and mouse mRNAs, in which it was shown that synonymous

codon usage bias contributed to the order and stability of mRNA secondary structure (Shabalina et al., 2006). It is possible that the triplet repeat provides a mechanism to guide the ribosome during translocation and prevent slipping.

The presence of thermodynamically stable structural elements in the mRNA is known to attenuate ribosome elongation. In some cases, RNA stem-loops or pseudoknots cause programmed ribosome frameshifting by stalling the ribosome on top of repetitive 'slippery' sequences located upstream of such structural elements. Here, this is considered an advantageous system to achieve the desired gene expression outcome (Atkins et al., 2016; Farabaugh, 1996). Ribosome pauses due to mRNA secondary structure have also been linked to protein co-translational folding (Thommen et al., 2017). A model whereby mRNA structural elements slow down elongation in order to facilitate native protein folding has been proposed, it suggests that such speed reduction would allow time for the nascent peptide chain to sample available configurations involving the residues of the already synthesized portion of the protein (Faure et al., 2016; Watts et al., 2009). This is compatible with evidence showing a correlation between RNA structure and protein compactness (Faure et al., 2016) and with observations that slowing down translation can minimize misfolding (Crombie et al., 1992; Komar et al., 1999; Purvis et al., 1987; Siller et al., 2010; Spencer et al., 2012; Zhang et al., 2009). Additionally, it could also fit with findings showing that acceleration in translation increases the probability of correct folding for proteins with long-distance contacts, which tend to be encoded by less structured mRNAs (Faure et al., 2017; O'Brien et al., 2014).

Finally, it has been reported that some single nucleotide variants (SNVs) that cause alterations in the RNA structure, known as 'riboSNitches', are associated with changes in gene expression and disease (Halvorsen et al., 2010; Wan et al., 2014). While the underlying mechanism of translation regulation by these riboSNitches is yet unclear for most cases, their discovery adds light to a growing body of evidence that suggests that an additional layer of genetic information is encoded in the RNA structure.

1.3.2. Amino acid signatures

Another factor that affects translation rates is inefficient peptide-bond formation. Despite the fact that the ribosome is able to carry out all possible peptide-bond combinations between the 20 amino acids, not all reactions are equally favorable. While this has little effect on overall translation for most amino acids, with decoding being the rate-limiting step, there are exceptions (Wohlgemuth et al., 2008). Proline (Pro) is unique in that its steric properties

make it a poor substrate for peptidyl transfer both as a donor in the P-site and as an acceptor in the A-site (Muto and Ito, 2008; Pavlov et al., 2009; Wohlgemuth et al., 2008). Poly-proline stretches are therefore challenging for the ribosome and cause stalling, hence it requires the activity of the specialized and essential eukaryotic translation factor eIF5A (EF-P in bacteria) to progress (Doerfel et al., 2013; Gutierrez et al., 2013; Ude et al., 2013).

Peptide-bond formation can also be indirectly inhibited downstream of amino acid sequences known to induce programmed stalling through electrostatic interactions with residues in the ribosomal peptide exit tunnel (Ito and Chiba, 2013; Lu and Deutsch, 2008; Wilson et al., 2016). Some of these sequences are rich in poly-basic amino acid stretches including lysine (Lys) and arginine (Arg), and they induce misalignment of reactive groups at the peptidyl transferase center of the ribosome, which slows down elongation (Charneski and Hurst, 2013; Weinberg et al., 2016).

An additional effect of Lys stretches on translation occurs when these are encoded by AAA codons. The slow peptide-bond formation kinetics between two consecutive Lys moieties triggers ribosome frameshifting on consecutive AAA codons, which can lead to premature termination events that will target the mRNA for nonsense-mediated decay (Arthur et al., 2015; Koutmou et al., 2015).

1.3.3. Codon usage and tRNA pools

The stochastic nature of decoding during translation elongation implies that non-cognate and near-cognate ternary complexes will compete with the cognate ones for binding to the ribosomal A-site. Therefore, the concentration of cognate aa-tRNA relative to the rest of aa-tRNAs in the ternary complex pool is a critical factor that can affect decoding rates. Despite the initial steps of binding to the ribosome having similar kinetics for all ternary complexes, cognate or not, cognate codon–anticodon pairs are selected and stabilized through interactions with three conserved residues (A1824, A1825 and G626) of the 18S rRNA in the decoding site (Demeshkina et al., 2012; Gromadski et al., 2006; Loveland et al., 2017; Ogle et al., 2001). These residues check for the correct geometry of the codon–anticodon complex by interacting with the first two pairs of the duplex, while enabling the third position of the codon and the first position of the tRNA anticodon (nucleotide 34 on the tRNA sequence, also called wobble position) to participate in non-Watson-Crick pairing related to the degeneracy of the genetic code (Loveland et al., 2017; Ogle et al., 2001). The process of decoding is further influenced by tRNA modifications (Agris et al., 2018) and the ability of

some base pairs, such as G:U, to maintain Watson-Crick geometry via rare tautomeric forms that the ribosome recognizes as cognate (Rozov et al., 2015; Satpati and Åqvist, 2014).

As a result of the stochastic recognition of mRNA codons, a given codon can be defined as optimal or non-optimal depending on how efficiently a cognate aa-tRNA is selected from the tRNA pool. Additionally, codon usage is known to be biased, with some synonymous codons being more frequently represented in the transcriptome than others that are considered rare (**Figure 1.8**), and to vary between organisms (Plotkin and Kudla, 2011). Whether codon usage bias is the consequence of a selective pressure for efficient and accurate protein translation or the effect of mutational processes occurred throughout evolution is still under debate, with the most likely is it being the result of the joint actions of mutation, drift, and selection (Hershberg and Petrov, 2008). It is becoming evident, however, that codon bias influences protein synthesis at different stages, including translation efficiency and fidelity, protein folding, and mRNA stability (Brule and Grayhack, 2017; Chaney and Clark, 2015; Hanson and Collier, 2017; Novoa and Ribas de Pouplana, 2012; Quax et al., 2015; Supek, 2016).

		Second letter				
		U	C	A	G	
U	UUU 1.76 } Phe	UCU 1.52 } Ser	UAU 1.22 } Tyr	UGU 1.06 } Cys	U	
	UUC 2.03 }	UCC 1.77 }	UAC 1.53 }	UGC 1.26 }	C	
	UUA 0.77 } Leu	UCA 1.22 }	UAA 0.10 Stop	UGA 0.16 Stop	A	
	UUG 1.29 }	UCG 0.44 }	UAG 0.08 Stop	UGG 1.32 Trp	G	
C	CUU 1.32 } Leu	CCU 1.75 } Pro	CAU 1.09 } His	CGU 0.45 }	U	
	CUC 1.96 }	CCC 1.98 }	CAC 1.51 }	CGC 1.04 }	C	
	CUA 0.72 }	CCA 1.69 }	CAA 1.23 } Gln	CGA 0.62 }	A	
	CUG 3.96 }	CCG 0.69 }	CAG 3.42 }	CGG 1.14 }	G	
A	AUU 1.60 } Ile	ACU 1.31 } Thr	AAU 1.70 } Asn	AGU 1.21 } Ser	U	
	AUC 2.08 }	ACC 1.89 }	AAC 1.91 }	AGC 1.95 }	C	
	AUA 0.75 }	ACA 1.51 }	AAA 2.44 } Lys	AGA 1.22 }	A	
	AUG 2.20 Met	ACG 0.61 }	AAG 3.19 }	AGG 1.20 }	G	
G	GUU 1.10 } Val	GCU 1.84 } Ala	GAU 2.18 } Asp	GGU 1.08 }	U	
	GUC 1.45 }	GCC 2.77 }	GAC 2.51 }	GGC 2.22 }	C	
	GUA 0.71 }	GCA 1.58 }	GAA 2.90 } Glu	GGA 1.65 }	A	
	GUG 2.81 }	GCG 0.74 }	GAG 3.96 }	GGG 1.65 }	G	

Figure 1.8. *H. sapiens* codon usage. Codon usage frequency in the human transcriptome, expressed in percentage units. Note the codon usage bias between synonymous codons. For example, the most frequently used alanine codon (GCC) appears in 2.77% of the transcriptome, while the least used one (GCG) is only present in 0.74% of it, a difference of roughly 3-fold. Data was obtained from the Codon Usage Database (Nakamura et al., 2000).

1.3.3.1. *Translation efficiency*

Initial observations of the impact of codon usage on heterologous protein expression led to the hypothesis that the average optimality of codons over an mRNA could affect its translation efficiency, defined as the amount of protein molecules synthesized per transcript in a given time (Ikemura, 1981). Evidence from several organisms that codon bias is more pronounced and evolutionally conserved in highly expressed genes, where it matches the most abundant cognate tRNAs, further reinforced this idea (Dong et al., 1996; Duret and Mouchiroud, 1999; Gouy and Gautier, 1982; Grosjean and Fiers, 1982; Ikemura, 1985; Kanaya et al., 1999; Moriyama and Powell, 1997; Sharp et al., 1986, 1995; Stenico et al., 1994; Urrutia and Hurst, 2001). Computational predictions and indirect measurements of translation rates using reporter genes suggested that ribosomes stall at rare codons (Komar et al., 1999; Sørensen et al., 1989; Tuller et al., 2010a; Varenne et al., 1984; Zhang et al., 2009). However, data in favor of a relationship between codon usage and translation efficiency in endogenous genes tends to be conflicting and remains a topic of intense discussion. The advent of ribosome profiling (**Figure 1.9**) (Ingolia et al., 2009, 2019) allowed experimental monitoring of the translation elongation rate across the transcriptome at single-codon resolution, providing an estimate of the relative density of ribosomes over specific codons within the A-site. While some ribosome profiling analyses failed to show a significant correlation between ribosome occupancy and codon bias (Charneski and Hurst, 2013; Lareau et al., 2014; Li et al., 2012; Pop et al., 2014; Qian et al., 2012), others have suggested that codon usage does indeed play a role in ribosome elongation rate, albeit with a small detectable effect of 3-fold at most (Dana and Tuller, 2012, 2014; Gardin et al., 2014; Stadler and Fire, 2011; Tuller et al., 2010b; Weinberg et al., 2016). Although the observed differences between studies could be due to the weakness of the pausing effects that may not always be properly identified, to technical biases inherent to the way the experiments were performed, or to biases in the downstream computational analysis (Artieri and Fraser, 2014; Hussmann et al., 2015; O'Connor et al., 2016; Santos et al., 2019), it is also possible that synonymous codons do in fact not cause ribosome stalling and that the observed outcome in protein synthesis and solubility is the product of concomitant variables. In support of this reasoning, a recent study found that the translation speed of a codon depends dramatically on its context within a transcript (Sharma et al., 2019). These contradictory results illustrate how difficult it can be to confidently measure subtle dependencies between codon usage bias and other biological factors.

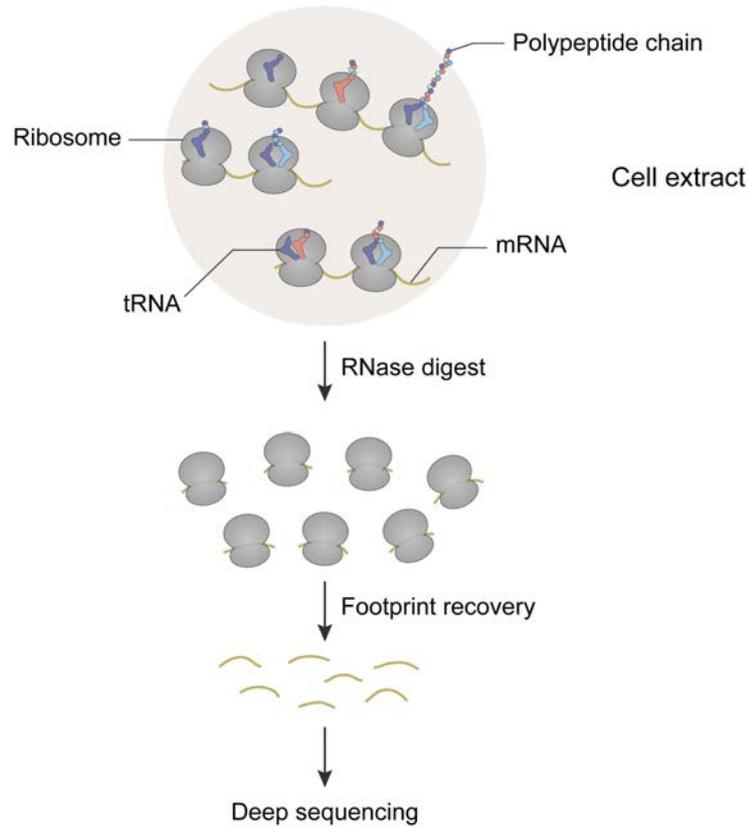


Figure 1.9. Ribosome profiling experimental workflow. The basic steps followed in a typical ribosome profiling experiment are shown: isolation of polysomes that are carrying out translation *in vivo*, RNase digestion that yields ribosome-protected mRNA fragments known as ribosome footprints, and recovery of these footprints followed by library preparation and deep sequencing.

1.3.3.2. Translation accuracy and protein folding

An alternative, non-exclusive, and well-established hypothesis on the selective utility of codon usage bias proposes that different codons have different probabilities of being mistranslated and, therefore, the preferential use of accurately translated codons is beneficial for the overall fitness of the cell, thus reducing the energy cost associated with non-functional and misfolded products (Akashi, 1994). Key residues of structural protein domains encoded in highly conserved mRNA regions have been found to display a bias towards optimal codons (Drummond and Wilke, 2008; Stoletzki and Eyre-Walker, 2006; Zhou et al., 2009). This, together with evidence that optimal codons are less prone to mistranslation than non-optimal ones (Dix and Thompson, 2006; Huang et al., 2009b; Kramer et al., 2010), would indicate that rapid translation rates at these positions have been selected for to prevent mistranslation events that would likely lead to the production of a dysfunctional protein (Drummond and Wilke, 2008).

Reinforcing the idea of codon bias acting as a mechanism for accurate translation, a study performed in *E. coli* showed a negative correlation between the dependency of correct protein folding on chaperone activity and the degree of codon optimality enrichment at structurally sensitive sites, suggesting a preventive role for codon usage in protein misfolding (Warnecke and Hurst, 2010). Folding is a process thought to occur co-translationally for many proteins both in prokaryotes and eukaryotes, which means that the nascent chain will begin to fold vectorially as it is translated, starting from the N-terminus and involving new residues as they emerge from the ribosome (Thommen et al., 2017). Varying local translation kinetics along a transcript have been associated to native folding of newly synthesized proteins (Purvis et al., 1987), with evidence suggesting that codon usage can be a critical factor regulating elongation rates and, in turn, proper folding (Nedialkova and Leidel, 2015; Pechmann and Frydman, 2013; Thanaraj and Argos, 1996). A correlation between well-structured protein regions and codon optimality has been drawn for several organisms, hinting at co-evolution of codon usage and protein structures (Thanaraj and Argos, 1996; Zhou et al., 2015). The fact that regions of non-optimal codons are often located downstream of structural domains also suggests that these sites may be used to promote co-translational folding of the upstream structures (Pechmann and Frydman, 2013; Weinberg et al., 2016; Zhang and Ignatova, 2009; Zhang et al., 2009). The functionality of codon usage bias in protein folding has also been observed for proteins that interact with the signal recognition particle (SRP) ribonucleoprotein for membrane targeting, where a non-optimal codon cluster of 35-40 codons downstream of the SRP-binding site, which is the length of the polypeptide that spans the ribosomal exit tunnel, may have a translational slowdown effect that enhances recognition by the SRP (Pechmann et al., 2014). Additionally, alterations of native elongation speed by inefficient decoding or codon optimization have been reported to result in protein misfolding, aggregation, decreased activity and instability (Buhr et al., 2016; Kimchi-Sarfaty et al., 2007; Nedialkova and Leidel, 2015; Yu et al., 2015). Thus, a mechanism whereby optimal codons are selected for at sites where mistranslation is detrimental and non-optimal codons are selected for downstream of structural domains where misfolding is most costly seems to be in place.

1.3.3.3. Codon context

As stated above, additional factors that may determine codon usage itself such as codon context, distribution or location can influence decoding rates of individual codons. It has been shown in several organisms that rare codons tend to cluster together within specific regions of a CDS (Clarke and Clark, 2008; Parmley and Huynen, 2009), and that these clusters are

often conserved in homologous genes of distantly related species (Chaney et al., 2017). One such example of codon clustering is the bias in codon pair usage across the transcriptome. The idea that codon pair usage is nonrandom and correlates with expression levels was first observed in *E. coli* (Gutman and Hatfield, 1989), with later studies reporting that bias is present in all three domains of life (Buchan et al., 2006; Tats et al., 2008). However, its role in translation was poorly understood until recent research, which used a systematic screen of a randomized GFP library containing random adjacent codons in *S. cerevisiae*, identified a set of 17 codon pairs that inhibit translation of the mRNA that encodes them (Gamble et al., 2016). Most pairs were formed by at least one codon that requires I:A or U:G wobble interactions for decoding, and in all cases where the inhibitory pair was only found in a single order of codons, such positioning was decisive to impair protein synthesis. A later study from the same group has shown that nine of the 17 inhibitory codon pairs are among the most conserved codon pairs at defined positions in genes, and that conservation is strongly correlated with slow translation of the pairs relative to their synonymous codons (Ghoneim et al., 2019). These results emphasize that not only codon identity but also codon order are parameters that can affect elongation rates, and suggest a potential translation regulatory mechanism that is conserved in yeast.

The order in which codons are distributed along the mRNA may be important as well. It has been shown that, in a given transcript, once a particular codon has been used, there is a strong tendency to use codons recognized by the same tRNA in subsequent occurrences of the amino acid it encodes (Cannarrozzi et al., 2010). The authors argued that this effect, termed “autocorrelation”, was a consequence of local tRNA recycling pressure. However, a subsequent study refuted the proposed explanation and suggested that codon autocorrelation is in fact a consequence of codon bias (Hussmann and Press, 2014). Some controversy can also be found regarding the so-called 5' “low-efficiency ramp” or “translational ramp”, that is, the region located ~100 nucleotides downstream of the start codon, which is generally enriched in rare codons and has been argued to control the rate of translation by forming an early traffic bottleneck in order to minimize ribosome collisions further down the mRNA (Tuller et al., 2010a). Recent ribosome profiling data provided evidence of a ribosome build-up at the 5' ramp, but noted that codon usage could only account for a third of it (Weinberg et al., 2016). The observation that mRNA secondary structure tends to be reduced near the start codon may partially explain it as well (Kudla et al., 2009; Pop et al., 2014). Thus, the overrepresentation of rare codons at the beginning of the coding sequence (CDS) as means to optimize translation rates may actually arise due to competing sequence constraints such as mRNA structural features.

1.3.3.4. Genetic programs and tRNA dynamics

Codon bias has also been shown to be relevant in the coordination of certain gene expression patterns. Functionally related genes tend to display similar codon usage and mRNA decay rates (Lithwick and Margalit, 2005; Najafabadi et al., 2009; Presnyak et al., 2015; Wang et al., 2002), hinting at codon bias as a strategy to optimize the expression of genes whose protein outputs work together in physiological systems and may be required to be present at comparable levels within the cell. Furthermore, gene expression demands, and with them codon usage patterns, change as tissues develop, differentiate, and respond to stimuli that require rapid and adequate cellular reprogramming. Given that the tRNA pool composition is known to be dynamic and, therefore, it can vary considerably between different conditions, it is thought that changes in tRNA pools may alter gene expression by mirroring the codon usage of stress- or tissue-specific genes, thus enhancing their translation (Dittmar et al., 2006; Plotkin et al., 2004; Schmitt et al., 2014).

The idea that functional adaptation of tRNA pools could impact translation efficiency in a dynamic manner was first proposed almost 50 years ago (Garel, 1974; Garel et al., 1974). It was shown that the tRNA composition of highly specialized cell systems such as the silk gland of *Bombyx mori* changed during the development of this organ in order to meet the translational demands of the extremely codon biased silk proteins fibroin and sericin, which are produced at high rates and account for 25% of the silkworm's bodyweight. Several additional scenarios wherein tRNA pools influence gene expression have been described since then. Amino acid starvation leads to changes in the relative concentration of aminoacyl-tRNA isoacceptors –i.e. tRNAs with different anticodon sequences that decode the same amino acid– that switch codon optimality both in *E. coli* and human cells, hence promoting the translation of transcripts enriched in previously non-optimal codons in order to adapt to cellular requirements to restore homeostasis. Different strategies are followed by the two species, while such remodeling drives the expression of genes that encode amino acid biosynthetic pathways in *E. coli* (Dittmar et al., 2005; Elf et al., 2003), in human cells it is the translation of mRNAs involved in protein recycling by the ubiquitin-proteasome system that gets favored (Saikia et al., 2016). Fluctuations in tRNA expression at different stages of the life of a cell may also enable differential expression of genes with codons that match the changing tRNA pools. For example, tRNA expression levels have been shown to oscillate during cell cycle, as do levels of aminoacyl-tRNA synthetases and ATP, and this has been suggested to shape the translational landscape of cell cycle-regulated genes, which display a strong preference for codons with low codon-anticodon binding affinity (Frenkel-Morgenstern et al., 2012). tRNA concentrations have also been found to vary between

proliferation and differentiation cell states. These two genetic programs present specific codon biases coupled to their corresponding tRNA pool composition, which is mostly determined by distinct histone modification patterns around tRNA genes, suggesting the existence of transcriptional programs coordinating tRNA supply and codon usage demand (Gingold et al., 2014). This mechanism seems to have been co-opted in breast cancer cells, where the selective upregulation of certain tRNAs is associated to increased proliferation and metastatic progression (Goodarzi et al., 2016; Pavon-Eternod et al., 2009).

Modulation of tRNA modification patterns (*see section 1.4.3*) renders yet an additional layer of regulation to adapt gene translation to different cellular requirements. Some tRNA modifications at the anticodon stem-loop influence decoding of particular codons by altering codon-anticodon binding affinities, which can bias the codon preferences of the tRNA pool in favor of mRNAs needed to respond appropriately to a variety of environmental factors (Chan et al., 2018). Different types of external stimuli convey selective reprogramming mechanisms of the tRNA epitranscriptome. For example, in yeast, exposure to DNA-damaging compounds leads to increased specific tRNA modifications that promote the translation of transcripts involved in DNA damage response and cell cycle control (Begley et al., 2007; Deng et al., 2015). While oxidative stress upregulates a different tRNA modification that has been linked with enhanced translation of 38 transcripts amongst which there is RPL22A, a key ribosomal protein that prevents reactive oxygen species sensitivity (Chan et al., 2012). A recent study found that resistance to therapeutic agents in human melanoma cells involves the upregulation of specific tRNA-modifying enzymes, which favor the translation of the stress-response transcription factor HIF1 α (Rapino et al., 2018). Thus, variations in the composition of aa-tRNA pools and in the modification status of tRNAs can regulate the expression of genes with adapted codon usage bias.

1.4. Transfer RNA

As the nexus molecules between mRNAs and proteins, transfer RNAs have a fundamental role in translation. Their canonical function is the delivery of amino acids to the ribosome in order to translate in a template-directed manner the genetic information encoded in mRNA into a polypeptide chain. tRNAs are therefore ubiquitous molecules present in all domains of life and represent the most abundant species of small non-coding RNAs in the cell. However, recent evidence has revealed that tRNAs are involved in other cellular processes beyond translation, such as drivers of adaptive protein synthesis to environmental cues, participation in biosynthetic pathways, and functions as signaling molecules in several regulatory networks.

1.4.1. tRNA structure

The most common cytosolic tRNA molecules that are involved in translation have a length of 76 nucleotides (nt), although this measure can oscillate between 75 to 95 nt (Rich and RajBhandary, 1976; Sprinzl et al., 1998). The cloverleaf-shaped secondary structure of tRNAs is highly conserved and can be divided into five functional parts: the acceptor stem, the dihydrouridine (D) stem-loop, the anticodon stem-loop, the variable loop and the thymidine–pseudouridine–cytidine (T Ψ C) stem-loop (Holley et al., 1965) (**Figure 1.10 A**). The acceptor stem contains the 5' and 3' ends of the tRNA, with the CCA trinucleotide sequence required for aminoacylation located at the 3' end. On the opposite end of the tRNA resides the anticodon, which is always numbered 34-36 and binds to cognate codons of the mRNA in the ribosome in an antiparallel complementary manner. The first base of the anticodon (position 34) is also called wobble base, due to its ability to form non-Watson-Crick base-pairing with the third position of the mRNA codon (Crick, 1966). Upon folding into its three-dimensional structure, the tRNA acquires an upside-down L-shaped conformation where the amino acid accepting site and the anticodon take the longest possible distance from each other (**Figure 1.10 B**). Such folding depends on an intricate network of tertiary interactions between conserved residues in the D and T Ψ C loops that stabilize the structure (Kim et al., 1974). All tRNAs need to fit properly in the ribosomal tRNA-binding sites in order to carry out their function efficiently, hence the architecture of the different tRNA species is rather homogeneous, mainly due to common identity rules and structural constraints, despite the high sequence variation they display (Giegé et al., 2012; Ramakrishnan, 2002; Saint-Léger et al., 2016). The extensive secondary and tertiary structures inherent to functional

tRNAs confer high stability, which renders tRNA half-lives of up to 3 days, roughly comparable with the half-life of rRNA (Phizicky and Hopper, 2010).

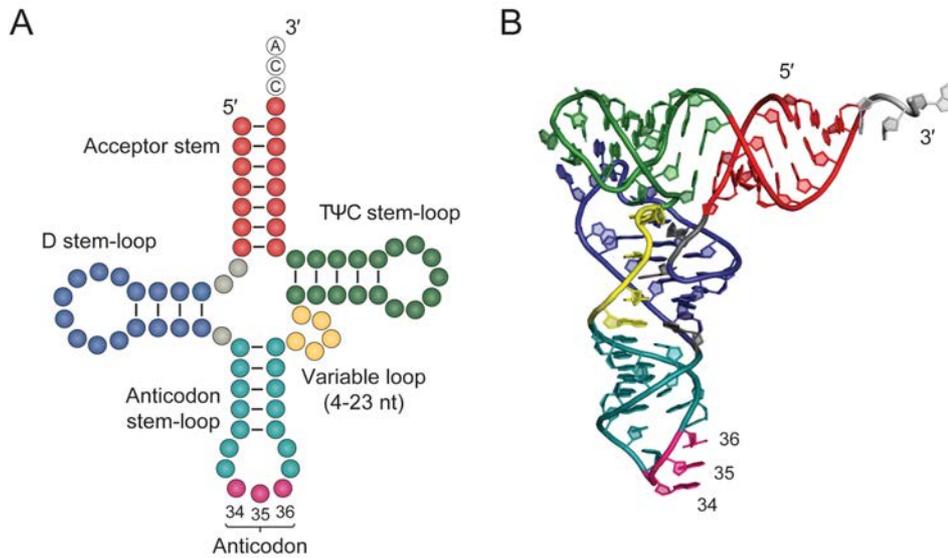


Figure 1.10. tRNA structure. (A) Cloverleaf diagram corresponding to the consensus secondary structure of tRNAs. (B) L-shaped structural 3D representation of a folded yeast tRNA^{Phe} showing extensive intramolecular base-pairing (PDB: 1EHZ) (Shi and Moore, 2000). Structural elements of the tRNA are color-coded equally in both representations: acceptor stem (red), dihydrouridine (D) stem-loop (blue), anticodon stem-loop (cyan), anticodon (bases 34, 35, 36 in pink), variable loop (yellow), TΨC stem-loop (green), and the conserved CCA-3' end (white).

1.4.2. tRNA biogenesis

The life of eukaryotic tRNAs starts with transcription in the nucleolus (**Figure 1.11**), a process mediated by the coordinated action of the transcription factors TFIIIB and TFIIIC, and RNA polymerase III (Pol III) that generates precursor or pre-tRNAs containing 5' leader and 3' trailer sequences. Some tRNA genes also contain introns of variable length. A series of maturation steps involving the removal of 5' leader and 3' trailer, as well as splicing of the introns, follow. The order and cellular location of these events can vary across species (Hopper, 2013; Yoshihisa, 2014). Finally, the CCA sequence required for nuclear export and tRNA aminoacylation is added to the 3' end of tRNAs. Further alterations in the tRNA sequence that can take place both before and after export to the cytoplasm include the enzymatic modification of some of its bases, which confer varying properties to the individual mature tRNA species. Two surveillance pathways, one within the nucleus and the other in the cytoplasm, monitor the integrity of tRNAs during and after processing as part of a quality control mechanism. Export of nuclear-encoded tRNAs to the cytoplasm or specific cellular organelles is done in a controlled manner so that only mature tRNAs that have undergone

proper processing can leave the nucleus (Arts et al., 1998). Mature tRNAs act as substrates for aminoacylation by aaRS and are subsequently used for protein synthesis in the ribosome (Phizicky and Hopper, 2010).

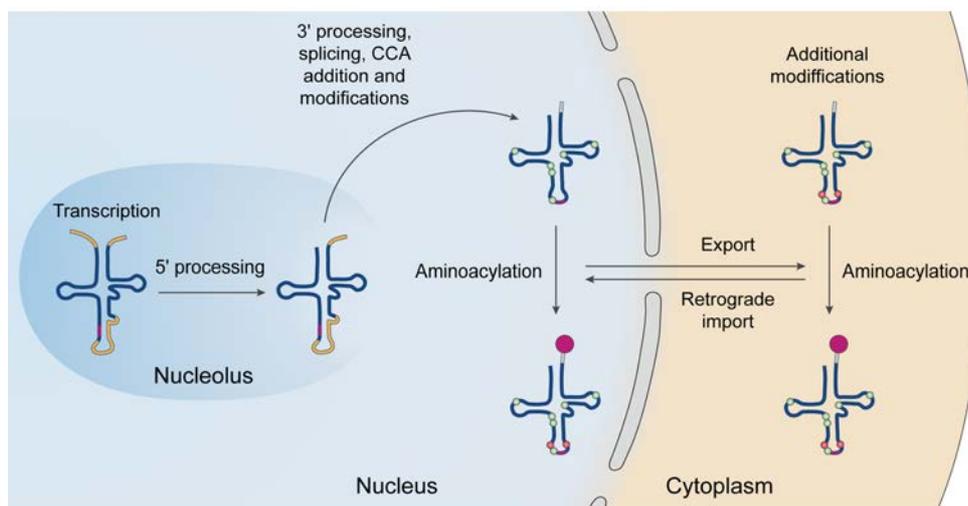


Figure 1.11. tRNA biogenesis. Transcription of tRNA genes occurs in the nucleolus, it yields immature transcripts containing 5' leader, 3' trailer, and intronic sequences (yellow). The anticodon is shown in magenta and parts of the tRNA that are maintained in the mature sequence, in dark blue. The tRNA is then processed by 5' leader and 3' trailer removal, intron splicing, addition of the CCA (gray), and several base modification steps (green circles). Once tRNAs are exported to the cytoplasm, they may become further modified (red circles), aminoacylated with their cognate amino acid (magenta circle) and participate in translation. Aminoacylation has also been reported to occur in the nucleus (Lund and Dahlberg, 1998). Variations of this pathway have been described for some organisms such as yeast, where intron splicing happens on the surface of the outer mitochondrial membrane (Chatterjee et al., 2018).

1.4.3. tRNA modifications and editing

In his 1966 review, Carl Woese mentioned that “perhaps the most bizarre feature of tRNA is its high content of unusual nucleotides –which can amount to almost 20% of the total nucleotides in the molecule” (Woese et al., 1966). This statement reflects the notion that tRNAs are the molecules with the highest density of post-transcriptional modifications among all RNA types. Today we know that the degree of modification varies between organisms and individual tRNAs, as it can go from 1.7% (a single modified residue) up to 23.7% (18 modified residues) of all tRNA bases (Machnicka et al., 2014). Nuclear-encoded human tRNAs contain an average of 13 modifications per molecule (Pan, 2018), which are sampled from the more than 50 structures of tRNA modified bases described up to date in eukaryotes (**Figure 1.12**) (Cantara et al., 2011; Machnicka et al., 2014). While some modifications are found in most nuclear-encoded tRNAs, such as dihydruridine in the D-loop or pseudouridine

in the TΨC loop, present in 72% and 94% of the available eukaryotic sequences, respectively; others are limited to specific tRNAs (Machnicka et al., 2014). The extent of specific modifications at particular bases can be dynamically regulated, thus expanding the complexity of the tRNA pool and exerting regulatory roles in cellular adaptation to environmental changes.

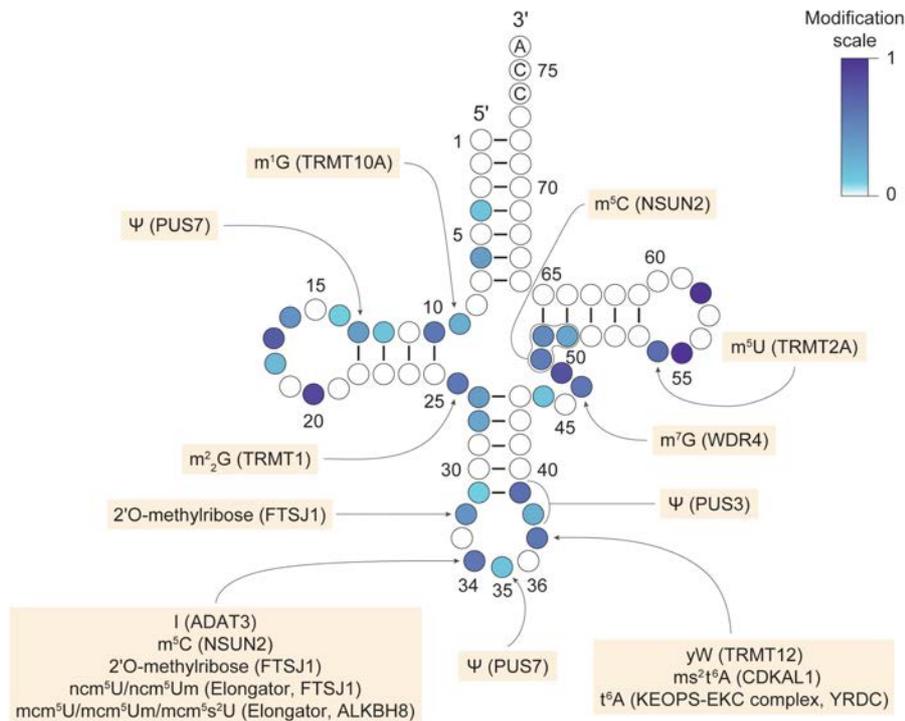


Figure 1.12. Distribution and frequency of human cytosolic tRNA modifications, and their link to disease. Schematic representation of the distribution of modified residues in human cytosolic tRNAs. The frequency of modification within each position (Machnicka et al., 2014), (Machnicka et al., 2014), is mapped on the consensus tRNA cloverleaf structure and color coded as indicated. Modifications that have been associated with human pathologies are shown together with their corresponding enzyme (in parenthesis). Abbreviations: m¹G, 1-methylguanosine; TRMT10A, tRNA methyltransferase 10 homologue A; Ψ, pseudouridine; PUS7, pseudourine synthase-7; m²₂G, N²,2'-O-dimethylguanosine; TRMT1, tRNA (guanine(26)-N(2))- dimethyltransferase; FTSJ1, FtsJ RNA 2'-O-methyltransferase 1; I, inosine; ADAT3, adenosine deaminase acting on tRNA 3; m⁵C: 5-methylcytosine; NSUN2, NOP2/Sun RNA methyltransferase family member 2; ncm⁵U, 5-carbamoylmethyluridine; ncm⁵Um, 5-carbamoylmethyl-2'-O-methyluridine; mcm⁵U, 5-methoxycarbonylmethyluridine; mcm⁵Um, 5-methoxycarbonylmethyl-2'-O-methyluridine; mcm⁵s²U, 5-methoxycarbonylmethyl-2-thiouridine; ALKBH8, alkB homolog 8, tRNA methyltransferase; yW, wybutosine; TRMT12, tRNA methyltransferase 12 homolog; ms²t⁶A, 2-methylthio-N⁶-threonylcarbamoyladenine; CDKAL1, CDK5 regulatory subunit-associated protein 1-like 1; t⁶A, N⁶-threonylcarbamoyladenine; KEOPS, kinase, endopeptidase and other proteins of small size; EKC, endopeptidase-like and kinase associated to transcribed chromatin; YRDC, yrdC N(6)-threonylcarbamoyltransferase domain containing; PUS3, pseudouridine synthase-3; m⁷G, 7-methylguanosine; WDR4, WD repeat domain 4; m⁵U, 5-methyluridine; TRMT2A, tRNA methyltransferase 2 homologue A. Adapted from (Hawer et al., 2018; Monies et al., 2019a; Tahmasebi et al., 2018; Torres et al., 2014a).

Modifications are introduced post-transcriptionally by specialized enzymes throughout the tRNA life cycle and they play major roles in tRNA folding and stability, tRNA interactions with the ribosome, and aminoacylation efficiency and accuracy (El Yacoubi et al., 2012). Furthermore, hypomodification of tRNAs has been linked to rapid tRNA decay (Alexandrov et al., 2006; Phizicky and Hopper, 2010), ribosomal pausing during translation elongation (Morscher et al., 2018; Nedialkova and Leidel, 2015; Zinshteyn and Gilbert, 2013), and mistranslation events (Patil et al., 2012; Taniguchi et al., 2018); thus, modifications also provide protection from degradation and ensure proper protein synthesis through modulation of codon-anticodon interactions. In this regard, modifications in the anticodon loop are of particular interest. Bases 34 and 37 are the most frequent locations for modifications critical for correct decoding. Modifications at base 37 adjacent to the anticodon stabilize codon-anticodon interactions and prevent translational frame-shifting (El Yacoubi et al., 2012). Modifications at position 34 (wobble base) can directly influence decoding by altering the pattern of hydrogen bond donors and acceptors between the codon and the anticodon. Such changes can either increase the wobbling capacity of a given tRNA or restrict it, which renders a broadened decoding potential that allows translation of additional codons to the one that is a perfect match, or limits it to a single one (Agris, 1991; Agris et al., 2018). In fully mature tRNAs, wobble adenosine (A34) is almost always edited to inosine (I) in a deamination process that leads to a complete base substitution, U34 is generally decorated by a number of modifications in all organisms, and C34 and G34 display the lowest level of modification (Machnicka et al., 2014). Because of the relevant roles of tRNA modifications in protein synthesis (Chou et al., 2017), it is not surprising that mutations in tRNAs or defects in their processing or modifying enzymes have been increasingly linked to several pathologies (**Figure 1.12**) (Hawer et al., 2018; Tahmasebi et al., 2018; Torres et al., 2014a).

1.4.4. tRNA gene copy number, tRNA abundance, and codon usage bias

tRNA gene copy number has traditionally been used to characterize tRNA pools in a given organism. Despite not being a direct measurement of tRNA abundance and thus presenting certain limitations, it offers interesting insights into some of the features of the tRNA pool that are expressed below.

The genetic code contains 61 sense codons and a stop codon that, under certain circumstances, specifies the amino acid selenocysteine (Sec). However, most genomes analyzed to date contain tRNA sets of 30 to 50 different anticodon types with at least one

molecule for each of the 20 amino acids. Of the 62 possible isoacceptors, eight tend to be spared in all domains of life. Seven of these eight isoacceptors correspond to codons in the two-codon boxes of the genetic table and their absence indicates that they could potentially lead to mistranslation events during elongation. Other anticodon sets are only avoided in certain kingdoms (Rak et al., 2018). For example, the human nuclear genome contains 47 tRNA isoacceptors, meaning that there is an additional set of seven anticodons that are missing, shown in blue in **Figure 1.13** (Chan and Lowe, 2016). Thus, while codons lacking a fully matched tRNA rely on wobble base-pairing to be decoded, most amino acids can be incorporated using different tRNA isoacceptors. Each tRNA isoacceptor can be encoded by multiple gene copies in a given genome, a number which is known to vary between and within species and that can amount to several hundred (Iben and Maraia, 2014; Marck and Grosjean, 2002; Parisien et al., 2013). In higher eukaryotes these copies are divided into gene families of tRNA isodecoders that bear the same anticodon but differ in their body sequence. As stated above, the frequency with which the different isoacceptors are present across the genome is uneven, with some tRNAs having up to 29 copies and others only 3 in the human genome (**Figure 1.13**) (Chan and Lowe, 2016). While a logic consequence of such tRNA gene copy multiplicity would be that tRNAs with higher number of copies were more abundant in the cell than those encoded by fewer copy numbers, it has also been shown that different copies of a tRNA can be differentially expressed under specific conditions or tissues (Ishimura et al., 2014, 2016; Sagi et al., 2016; Torres et al., 2019) and that roughly half of human tRNA genes are silent (Thornlow et al., 2018; Torres, 2019).

An interesting question pertains to the interaction between tRNA gene copy number, tRNA abundance and codon usage bias. In unicellular organisms such as bacteria and fungi, the genomic tRNA copy number correlates with the cellular tRNA concentration and with the codon usage of highly expressed genes (Dong et al., 1996; Ikemura, 1985; Kanaya et al., 1999; Percudani et al., 1997; Tuller et al., 2010a). However, in higher eukaryotes, said correlations differ. On the one hand, in these organisms, tRNA pools are dynamically regulated so that translation can be adapted to environmental cues, developmental stages and tissue specific genetic programs (*see section 1.3.3.4*), therefore making tRNA gene copy number an unsuitable proxy for tRNA abundance. On the other hand, although the correlation between tRNA copy number and codon usage is apparently poor—see **Figure 1.13** for amino acids such as alanine (Ala), whose most frequently used codon (GCC) has no matching anticodon (GGC)—, it is significantly improved when considering a species-specific tRNA modification at the wobble position of the anticodon that widens the codon recognition ability of the modified tRNAs (Novoa et al., 2012). Such modification is the conversion of adenine at position 34 of the anticodon (A34) to inosine (I34).

Four-box codon sets

Ala	Anticodon	AGC	GGC	CGC	UGC
	copy number	22	0	4	8
	Codon	GCU	GCC	GCG	GCA
	%	1.84	2.77	0.74	1.58
Gly	Anticodon	ACC	GCC	CCC	UCC
	copy number	0	14	5	9
	Codon	GGU	GGC	GGG	GGA
	%	1.08	2.22	1.65	1.65
Pro	Anticodon	AGG	GGG	CGG	UGG
	copy number	9	0	4	7
	Codon	CCU	CCC	CCG	CCA
	%	1.75	1.98	0.69	1.69
Thr	Anticodon	AGU	GGU	CGU	UGU
	copy number	9	0	5	6
	Codon	ACU	ACC	ACG	ACA
	%	1.31	1.89	0.61	1.51
Val	Anticodon	ACC	GAC	CAC	UAC
	copy number	9	0	11	5
	Codon	GUU	GUC	GUG	GUA
	%	1.1	1.45	2.81	0.71

Two-box codon sets

Phe	Anticodon	AAA	GAA	
	copy number	0	10	
	Codon	UUU	UUC	
	%	1.76	2.03	
Asn	Anticodon	AUU	GUU	
	copy number	0	20	
	Codon	AAU	AAC	
	%	1.7	1.91	
Lys	Anticodon		CUU	UUU
	copy number		15	12
	Codon		AAG	AAA
	%		3.19	2.44
Asp	Anticodon	AUC	GUC	
	copy number	0	13	
	Codon	GAU	GAC	
	%	2.18	2.51	
Glu	Anticodon		CUC	UUC
	copy number		8	7
	Codon		GAG	GAA
	%		3.96	2.9
His	Anticodon	AUG	GUG	
	copy number	0	10	
	Codon	CAU	CAC	
	%	1.09	1.51	
Gln	Anticodon		CUG	UUG
	copy number		13	6
	Codon		CAG	CAA
	%		3.42	1.23
Tyr	Anticodon	AUA	GUA	
	copy number	0	13	
	Codon	UAU	UAC	
	%	1.22	1.53	
Cys	Anticodon	ACA	GCA	
	copy number	0	29	
	Codon	UGU	UGC	
	%	1.06	1.26	

Odd codon sets

Ile	Anticodon	AAU	GAU	UAU
	copy number	14	3	5
	Codon	AUU	AUC	AUA
	%	1.6	2.08	0.75
Met	Anticodon		CAU	
	copy number		19	
	Codon		AUG	
	%		2.2	
Trp	Anticodon		CCA	
	copy number		7	
	Codon		UGG	
	%		1.32	
Sec	Anticodon			UCA
	copy number			2
	Codon			
	%			

Six-box codon sets

Ser	Anticodon	AGA	GGA	CGA	UGA	ACU	GCU	
	copy number	9	0	4	4	0	8	
	Codon	UCU	UCC	UCG	UCA	AGU	AGC	
	%	1.52	1.77	0.44	1.22	1.21	1.95	
Arg	Anticodon	ACG	GCG	CCG	UCG		CCU	UCU
	copy number	7	0	4	6		5	6
	Codon	CGU	CGC	CGG	CGA		AGG	AGA
	%	0.45	1.04	1.14	0.62		1.2	1.22
Leu	Anticodon	AAG	GAG	CAG	UAG		CAA	UAA
	copy number	9	0	9	3		6	4
	Codon	CUU	CUC	CUG	CUA		UUG	UUA
	%	1.32	1.96	3.96	0.72		1.29	0.77

Figure 1.13. Human genomic tRNA copy number and codon usage. Shown are the nuclear-encoded tRNA gene numbers for each anticodon together with the usage frequency of their matching codon (gray-shade rectangles) in the human transcriptome. Red-shaded rectangles correspond to the set of eight ANN isoacceptors that tend to be spared in all domains of life and blue-colored rectangles indicate the absence of that GNN tRNA gene in humans. tRNA gene copy number prediction data for the *H. sapiens* build GRCh37/hg19 Feb 2009 was obtained from (Chan and Lowe, 2016). Human codon usage data was taken from (Nakamura et al., 2000).

1.5. A-to-I editing at the tRNA wobble base

The non-canonical base inosine is ubiquitously found in all forms of RNA across the tree of life (Gott and Emeson, 2000). It is the product of the hydrolytic deamination of adenosine (**Figure 1.14 A**) (Auxilien et al., 1996), a process known as A-to-I editing that constitutes the most common form of RNA editing (Zinshteyn and Nishikura, 2009). Inosine can be found in a naked form at position 34 of eukaryotic and bacterial tRNAs, as well as in a methylated state at positions 37 in eukaryotes, and 57 or 58 in archaea (Grosjean et al., 1996; Machnicka et al., 2014). Inosine at position 34 (I34) is one of the few essential tRNA post-translational modifications whose formation is catalyzed by adenosine deaminases acting on tRNA (ADATs). The absence or loss of function of the genes encoding these proteins has been reported to be lethal, underscoring the biological importance of I34. The bacterial enzyme responsible for I34 biosynthesis TadA is homodimeric, predominantly targets a single tRNA (tRNA^{Arg}_{AGC}) –with exceptions reported (Rafels-Ybern et al., 2019)–, and relies on very specific tRNA anticodon loop features such as the presence of a particular sequence (UACG) and proper loop size and structure (Elias and Huang, 2005; Wolf et al., 2002). In contrast, the eukaryotic ADAT is a heterodimer formed by subunits ADAT2 and ADAT3 (**Figure 1.14 B**) and is able to deaminate all tRNAs bearing a genomically encoded A34 in their anticodon, which can amount to up to eight different isoacceptors (tRNA^{Thr}_{IGU}, tRNA^{Ala}_{IGC}, tRNA^{Pro}_{IGG}, tRNA^{Ser}_{IGA}, tRNA^{Leu}_{IAG}, tRNA^{Ile}_{IAU}, tRNA^{Val}_{IAC}, tRNA^{Arg}_{ICG}) (Gerber and Keller, 1999; Rubio et al., 2007; Torres et al., 2015; Zhou et al., 2014). Furthermore, deamination by heterodimeric ADAT is dependent on the correct entire tRNA tertiary structure in addition to the local conformation of the anticodon loop, the presence of A at position 34, and a preference for a purine base at position 35 (Achsel and Gross, 1993; Auxilien et al., 1996; Haumont et al., 1984; Saint-Léger et al., 2016).

ADATs are closely related to the cytidine deaminase (CDA) protein family and contain the same active site sequence motif (H/C)XEX_nPCXXC (where X represents any amino acid, and *n* is any number of residues) in their deaminase domain (**Figure 1.14 B**). The conserved residues include three amino acids involved in the coordination of Zn²⁺ (one histidine and two cysteines), a proline that acts as the ammonium group binding site, and a glutamate residue that is responsible for proton shuttling and, therefore, essential for catalysis (Gerber and Keller, 1999). The eukaryotic ADAT heterodimer consists of two paralogous subunits wherein ADAT2 harbors the critical active site glutamate while in ADAT3 it has been substituted for a non-catalytic residue (valine in *H. sapiens*). This would indicate that, as it is the case for many eukaryotic tRNA modification enzymes, ADAT2 is the catalytic subunit and ADAT3 plays an auxiliary structural role in tRNA substrate recognition (Gerber and

Keller, 1999; Guy and Phizicky, 2014). However, one should keep in mind that both proteins are required for tRNA binding and deamination (Gerber and Keller, 1999). Moreover, studies in *Trypanosoma brucei* have shown that ADAT3 might also be essential for catalysis by contributing to zinc binding (Spears et al., 2011) and that ADAT2 could be participating in tRNA binding via a lysine- and arginine-rich domain (KR-domain) located in its C-terminal region (Ragone et al., 2011). Altogether, this suggests that both substrate recognition and catalytic activity of heterodimeric ADATs may work as a function of the coordinated action of active-site residues and distal domains located in the two subunits.

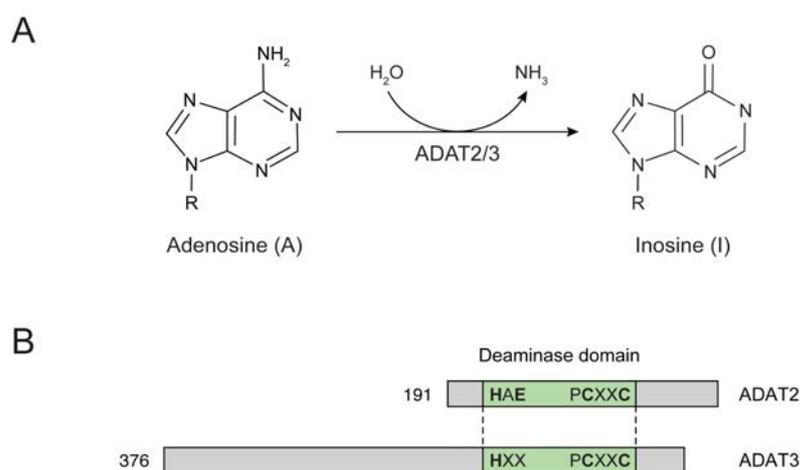


Figure 1.14. A-to-I conversion and human ADAT enzymes. (A) Hydrolytic deamination of adenosine leads to the formation of inosine and is catalyzed by the ADAT2/3 heterodimer. (B) Schematic representation of the protein domain structure of human ADAT2 and ADAT3. The deaminase domain (green box) contains an active site comprised of three evolutionarily conserved Zn^{2+} -binding residues (H, C, C) and a critical proton-shuttling glutamate (E) that is found in ADAT2 but not in ADAT3. Numbers indicate the amino acid length of the proteins.

1.5.1. Functional implications of I34

From an evolutionary point of view, the emergence of the I34 modification has been associated with a kingdom-specific increase in tRNA pool size and composition, and a cognate shift in codon usage (Novoa et al., 2012). As mentioned above, while I34 is used to very limited extent in bacterial tRNAs, its presence in eukaryotes has been expanded to an apparent near-saturation point (Saint-Léger et al., 2016) and it is linked to the loss of synonymous G34-containing tRNA isoacceptors (Grosjean et al., 2010) (Figure 1.13). The role of I34 in shaping the codon bias of different organisms is granted by the broadened decoding capacity it confers to tRNAs that bear it, which alters the tRNA pool available to certain codons whose usage may have been favored over time. According to Crick's wobble hypothesis, while A34 can only pair with U, I34 is able to wobble with C, U, and A (Crick,

1966). Therefore, a single INN tRNA anticodon can decode up to three different mRNA codons, thus solving the apparent inconsistency between the abundance of C-ended codons for the six- and four-box amino acids –except glycine– and the complete absence of the corresponding GNN anticodons (**Figure 1.13**).

Although wobble codon-anticodon interactions involving I34 can be influenced by additional tRNA modifications within the anticodon loop and display different decoding efficiencies of the NNC, NNU, and NNA codons (Agris et al., 2018), it is clear that the reading of NNC codons should be affected by the presence or absence of the edited base given the lack of a complementary G34-containing anticodon. Such dependence on I34 by NNC codons may have evolved to preserve the ability to regulate protein synthesis through a system of secondary information derived from the redundancy of the genetic code (Maraia and Iben, 2014). Consistent with this hypothesis, gene ontology (GO) analysis of mRNAs enriched or depleted in C-ended codons sensitive to the I34 modification across different organisms revealed a significant grouping of mRNAs enriched in I34-sensitive NNC codons into GO categories that were species-specific and whose number increased with the complexity of the organism (Torres et al., 2014b). Interestingly, no correlation with biological complexity was observed for GOs containing mRNAs depleted in I34-sensitive NNC codons. These results are indicative of a trend towards codon usage adaptation to I34 decoding among functionally related mRNAs.

Studies regarding the biological consequences of I34 modification deficiency are scarce and the essentiality of the enzyme can make this a challenging task. However, previous research has yielded some relevant findings. A viable strain of *Schizosaccharomyces pombe* encoding a mutant *tad3* (the gene encoding a homolog of *ADAT3*) that results in a S257N substitution in its protein sequence displays reduced I34 levels in tRNA^{Ala}_{AGC}, increased cycloheximide intolerance, and strong temperature-sensitive growth phenotypes due to irreversible cell cycle arrest in both mitotic G₁ and G₂ phases, suggesting a critical role of I34 in cell cycle progression in fission yeast (Tsutsumi et al., 2007). An additional *S. pombe* Tad3 mutant has been recently associated with impaired growth at high temperature and increased mating efficiency, which the authors linked to defects in TORC1 (target of rapamycin complex 1) phosphorylation activity (Otsubo et al., 2018). Depletion of ADAT2 in *Trypanosoma brucei* leads to a decrease in A-to-I editing at position 34 and, surprisingly, C-to-U editing at position 32 of tRNA^{Thr}_{AGU}, as well as a inhibition of cellular growth (Rubio et al., 2007). In *Arabidopsis thaliana*, lack of I34 in chloroplast tRNA^{Arg}_{ACG} due to disruption or silencing of the *TADA* gene, responsible for the modification biosynthesis in this organelle, is not lethal but causes severe phenotypes including delayed growth, very poor fertility, and profound defects in chloroplast translation efficiency that lead to impaired photosynthetic function (Delannoy et al., 2009;

Karcher and Bock, 2009). In contrast, and as previously shown for other organisms, knockout of the *A. thaliana* *ADAT2* and *ADAT3* genes is not compatible with plant survival and induces arrested embryo development. Depletion of *ADAT2* and *ADAT3*, on the other hand, is viable in spite of leading to a mild growth retardation and decreased I34 formation that does not affect the overall cellular levels or the aminoacylation efficiency of at least two substrate tRNAs (tRNA^{Leu}_{AAG} and tRNA^{Ser}_{AGA}) (Zhou et al., 2014).

1.5.2. I34 in human disease

Although tRNA A-to-I editing activity has been known to occur in human cells for years (French and Trewyn, 1990; Kretz et al., 1990), the nature of the human enzyme responsible for I34 formation and the physiological relevance of the modification have been completely unexplored. Recent evidence hinting at the importance of I34 in humans has been provided by genetic screenings of individuals affected by neurological disorders.

An ancient single missense mutation of a highly conserved valine (V) residue in human *ADAT3* was first identified to be present in homozygosis in patients exhibiting intellectual disability and strabismus back in 2013 in Saudi Arabia (Alazami et al., 2013). Additional symptoms linked to the mutation included microcephaly, nonspecific brain abnormalities, hypotonia, epilepsy, growth delay, and behavioral problems (Alazami et al., 2013; El-Hattab et al., 2016; Sharkia et al., 2018). Due to the existence of two annotated *ADAT3* isoforms in the NCBI RefSeq database (NM_138422.3 and NM_001329533.1), the authors initially described the mutation as V128M (Alazami et al., 2013; El-Hattab et al., 2016), which was later updated to V144M (Monies et al., 2017, 2019b). Of note, individuals carrying the V144M mutation in heterozygosis were asymptomatic, indicating that the pathology has an autosomal-recessive genetic inheritance (Alazami et al., 2013). In the absence of a crystal structure for the *ADAT2/3* heterodimer, *in silico* modelling of *ADAT3* was performed and revealed that the substituted valine is located in a hook which protrudes from the surface of the protein, and that the Val-to-Met mutation induces a kink that obliterates this protrusion (Alazami et al., 2013). More recent experimental data obtained by hydrogen-deuterium exchange mass spectrometry of the *T. brucei* *ADAT* heterodimer mapped the conserved Val to a region with low hydrogen exchange rate, suggesting that it is either buried within the protein or at the subunit interface with *ADAT2* (McKenney, 2018). Considering these observations and the fact that I34 has been shown to be an essential tRNA modification in other organisms, the V144M mutation is likely to impact the enzymatic activity of *ADAT2/3* in such a way that partial function is retained.

Recently, another novel *ADAT3* mutation consisting in a 8-bp duplication, which introduces a frameshift that might give rise to a truncated protein, was described to be associated to mild intellectual disability, microcephaly, and hyperactivity (Salehi Chaleshtori et al., 2018). This second *ADAT3* mutation also has an autosomal-recessive mode of inheritance and shares many clinical features with the V144M mutation.

Because the levels of tRNA A-to-I editing were not analyzed in the patients carrying these mutations, it is difficult to establish a direct relation of causation between the mutations and the pathological symptoms at this point. Additionally, the lack of a three-dimensional structure of the ADAT heterodimer hinders the understanding of the molecular basis for the effects observed in the affected patients. However, the fact that the resulting phenotypes are reminiscent of those associated to mutations in other tRNA modification enzymes that catalyze wobble modifications (Ramos and Fu, 2019; Torres et al., 2014a) is suggestive of the crucial role of wobble modifications in translation, particularly in the nervous system.

1.5.3. I34-dependent codons cluster in extracellular matrix genes

Previous results from our laboratory indicated that *ADAT2* is an essential gene in humans. This was inferred from a screen of more than a hundred clones generated using CRISPR/Cas9 technology in human HEK293T cells with the aim of obtaining an *ADAT2* knockout (KO) cell line. While no full KO clones were identified, about 30% of the clones exhibited variable degrees of *ADAT2* knockdown (KD) and conserved a wild type allele. The cell line displaying the lowest *ADAT2* protein levels (“*ADAT2* KD” hereinafter) (**Figure 1.15 A**) presented a frameshift-causing mutation in one of the *ADAT2* alleles (data not shown) and showed a reduction of I34 levels in all tRNA substrates (**Figure 1.15 B**), suggesting that modulations of I34 levels are compatible with viability of human cells in culture. The silencing of *ADAT2* was also associated to a globally delayed cell cycle progression (**Figure 1.15 C**). Strikingly, no defects in protein synthesis for *ADAT2* KD cells were observed by either pulse-labelling assay (**Figure 1.15 D**) or by flow cytometry analysis of cells expressing eGFP proteins enriched or depleted in codons that are translated by I34-containing tRNAs (hereinafter “*ADAT*-sensitive codons”) and that were broadly distributed along the transcript (**Figure 1.15 E**).

Alternatively, the effect of I34 in translation could potentially be specific to a set of proteins whose mRNA codon composition is locally biased towards the use of codons dependent on I34 for decoding. In this regard, the results of a transcriptome-wide analysis of local codon

usage in human genes showed that CDSs biased towards the usage of codons specifying amino acids susceptible to be decoded by I34-containing tRNAs (i.e. T, A, P, S, L, I, V, and R; hereinafter called “ADAT amino acids”) are overrepresented in the human genome (Rafels-Ybern et al., 2015). Interestingly, this tendency to encode ADAT amino acids is accompanied by a specific enrichment in four (T, A, P, and S) of the eight amino acids and by a significant preference for ADAT-sensitive codons (Rafels-Ybern et al., 2015). A more recent analysis used the same approach to compare the transcriptomes of eukaryotic and bacterial organisms and found that the enrichment in ADAT-sensitive codons is pervasive among eukaryotes, but absent in bacteria (Rafels-Ybern et al., 2018). These observations reinforce the notion that I34-containing tRNAs are favored in eukaryotic translation for the synthesis of transcript regions enriched in ADAT amino acids. The reasons for this selection are possibly to improve fidelity and efficiency of translation (Novoa et al., 2012; Schaub and Keller, 2002), and may result from an evolutionary adaptation to the synthesis of highly repetitive sequences with skewed amino acid compositions that lead to an increase in abundance and length of such transcripts (Rafels-Ybern et al., 2018, 2019).

The results obtained from the computational analysis identified genes whose translation might potentially be affected by the lack of I34. Due to their enrichment in ADAT-sensitive codons, these transcript regions will be referred to as “ADAT stretches”. An excerpt of these findings is shown in **Table 1.1** with the top 15 human genes most enriched in ADAT stretches. Two relevant features of the list of 171 genes identified to contain ADAT stretches are the noticeable overrepresentation of the family of mucins and the enrichment in genes belonging the “extracellular matrix constituent” molecular function category of gene ontologies (**Figure 1.16**). Thus, it is tempting to speculate a possible connection between ADAT and the synthesis of extracellular matrix proteins such as mucins.

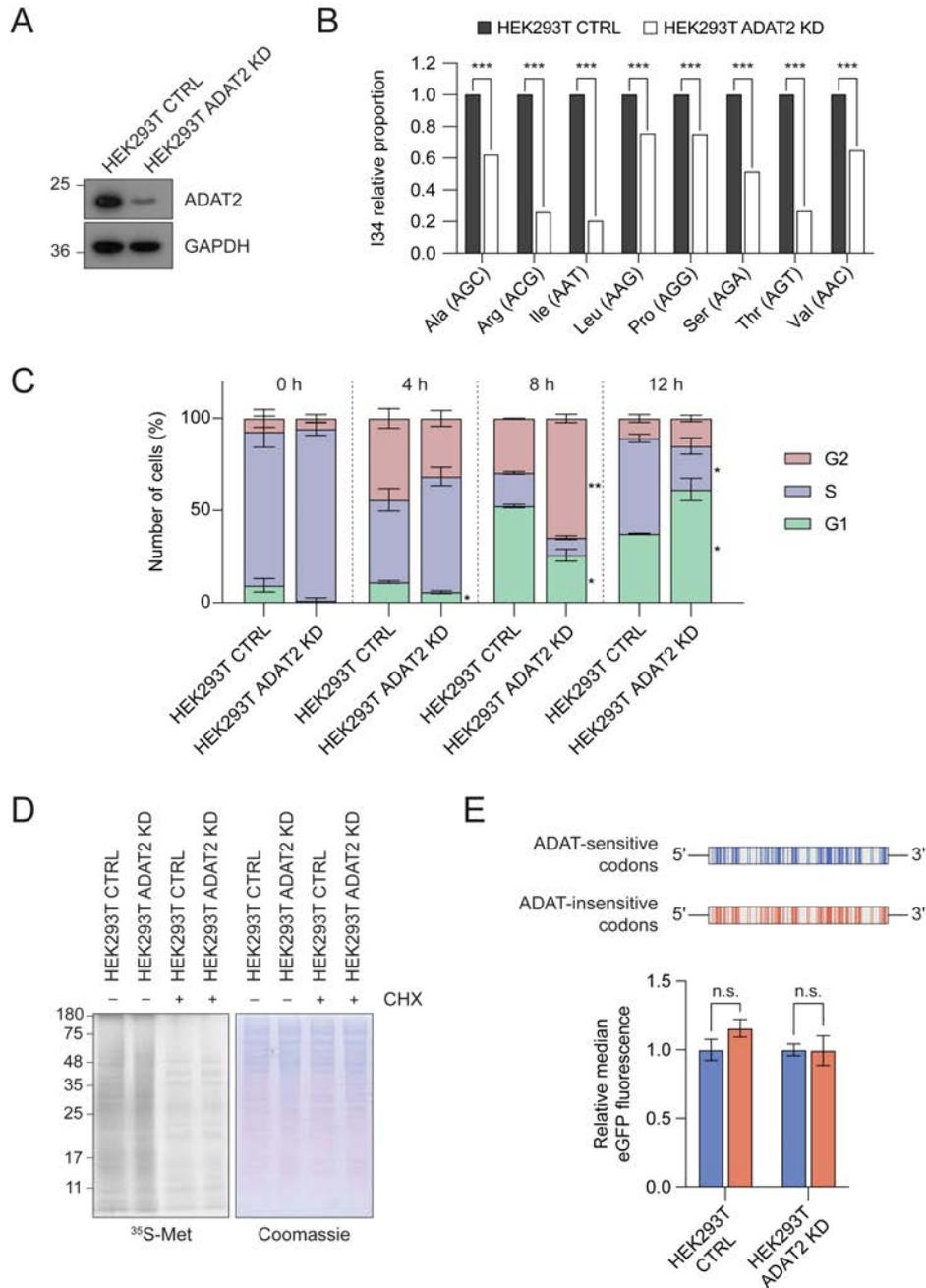


Figure 1.15. ADAT2 silencing in HEK293T human cells. (A) Protein levels of ADAT2 relative to GAPDH in CTRL and ADAT2 KD cells by Western blot. (B) Relative proportion of I34 in human ANN tRNAs for CTRL and ADAT2 KD cells ($***P \leq 0.001$). (C) Cell cycle progression analysis for CTRL and ADAT2 KD cells synchronized in S phase at time 0 h and analyzed at the indicated time points. The mean percentage and associated SD of cells in G1 (green), S (purple), and G2 (red) phases are shown for two independent replicates. Statistical analysis was performed with a *t*-test comparing each phase between CTRL and ADAT2 KD cells at each time point. ($*P \leq 0.05$, $**P \leq 0.01$). (D) Pulse-labelling assay. CTRL and ADAT2 KD cells were labeled with ^{35}S -Met. The translation inhibitor cycloheximide (CHX) was added as a control. (E) Relative eGFP fluorescence levels analyzed by flow cytometry for CTRL and ADAT2 cells expressing a codon engineered eGFP gene enriched in either ADAT-sensitive codons (blue) or ADAT-insensitive codons (red). Shown are the median and associated SD. (Source: Dr Adrián G Torres).

Table 1.1. Human genes presenting the highest local relative enrichment in ADAT-sensitive codons. The top fifteen genes obtained from a transcriptome-wide analysis of the ADAT amino acids codon usage in *H. sapiens* are shown. A sequence is considered an ADAT stretch when it encodes a proportion of ADAT amino acids larger than 0.837 within a window of 80 nucleotides. Data corresponding to the longest stretch identified for each gene is shown. The S:A ratio corresponds to the relative amount of ADAT-sensitive codons over the number of ADAT amino acids for a given stretch. Gene IDs have been ranked according to the length of their ADAT stretch and, at a secondary level, by their S:A ratio. (Source: Dr Àlbert Rafels Ybern).

Gene ID	ADAT stretch length (nt)	No. of ADAT amino acids (A)	No. of ADAT-sensitive codons (S)	S:A ratio	Transcript length (nt)
<i>MUC5AC</i>	1007	943	784	0.831	5655
<i>MUC5B</i>	719	621	492	0.792	5763
<i>MUC17</i>	571	485	366	0.755	4494
<i>SRCAP</i>	560	490	379	0.773	3231
<i>PRR36</i>	542	469	364	0.776	1347
<i>MUC4</i>	366	309	288	0.932	5413
<i>MUC3A</i>	363	309	265	0.858	3324
<i>SRRM2</i>	325	300	209	0.697	2753
<i>MUC6</i>	270	232	188	0.810	2440
<i>BCORL1</i>	251	221	178	0.805	1712
<i>KIAA0754</i>	248	209	184	0.880	1428
<i>PRRC2C</i>	247	223	210	0.942	2818
<i>MGAM2</i>	243	207	177	0.855	2516
<i>MUC7</i>	236	207	200	0.966	378
<i>AMOT</i>	216	189	167	0.884	676

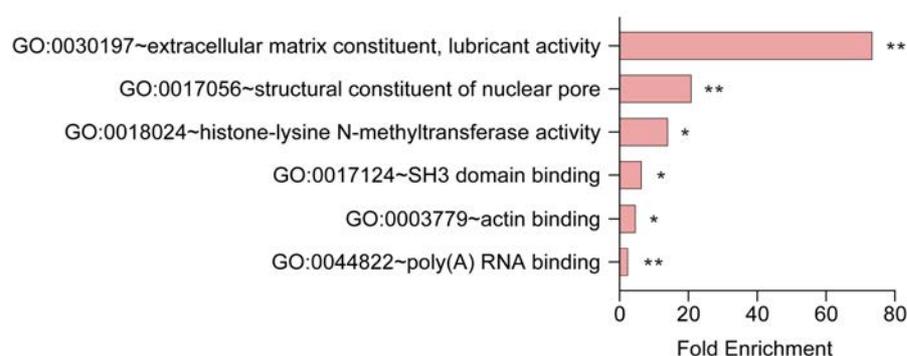


Figure 1.16. Gene ontology enrichment analysis of genes containing ADAT stretches. Shown are the significant GO categories for molecular functions of the identified 171 genes containing ADAT stretches obtained using DAVID Functional Annotation tool (Huang et al., 2009a). Statistical significance is defined by a Benjamini-corrected *P*-value ** $P < 0.01$, * $P < 0.05$.

1.6. Extracellular matrix

The extracellular matrix (ECM) is a highly specialized and dynamic three-dimensional network that bridges between cells providing tissues with mechanical support and signaling cues involved in cell survival, proliferation, differentiation and migration (Frantz et al., 2010; Theocharis et al., 2019). The ECM is, consequently, a central mediator of multicellularity. However, the composition of the ECM varies between different multicellular lineages due to their independently evolved mechanisms of multicellularity (Abedin and King, 2010). The core constituents of the metazoan ECM are large, secreted, multidomain, and often oligomeric proteins that are able to form extracellular fibrils and supramolecular structures. The internal architecture of the ECM proteins typically involves the presence of repeats of a particular set of protein domains that, despite not being exclusive to ECM proteins, are arranged in such a way that is highly characteristic of these type of proteins. Interestingly, each of these domains tends to be encoded by just one or a few exons in the genome, which is suggestive of multiple exon (domain) shuffling events during evolution (Hynes and Naba, 2012).

Although the precise composition of ECMs is tissue-specific, it generally includes collagens, elastin, fibronectin, laminins, glycoproteins, proteoglycans, and glycosaminoglycans (Theocharis et al., 2019). Matrix components interact with each other as well as with cell adhesion receptors found in the membranes of neighbor cells. Under normal conditions, the ECM is constantly remodeled through a finely regulated process whereby old matrix components are replaced by newly formed ones. The importance of appropriate maintenance of ECM composition and structure in tissue integrity and functionality (Karamanos et al., 2019) is underlined by observations that have identified abnormal matrix remodeling as a critical factor in the development and progression of a wide range of pathologies (Iozzo and Gubbiotti, 2018).

Of the multiple components of the human ECM, here we focus on a set of proteins that were found to be enriched in ADAT stretches. These are mucins, dystroglycan, and syndecan-3.

1.6.1. Mucins

Every mucosal membrane of the body is covered by mucus, a biological hydrogel whose primary component is the family of glycoproteins known as mucins (Bansil and Turner, 2006).

Mucins are large and heavily glycosylated extracellular proteins with roles in signal transduction as well as epithelial lubrication, hydration, and protection against pathogen invasion. Mucins also participate in mucociliary clearance and provide a selective permeable barrier for the diffusion, exchange and absorption of gases and nutrients (Bansil and Turner, 2006, 2018).

Due to their extensive glycosylation, the protein backbone of mucins accounts for only 20% (w/w) of the mass of these molecules, whereas their bound glycans constitute up to an 80% (w/w) of it (Corfield, 2015). The mucin protein architecture is generally characterized by the presence of a variable number of tandem repeats rich in proline, threonine, and serine (PTS domains) that are modified with O-linked glycosylation of serine and threonine moieties (**Figure 1.17**) (Bansil and Turner, 2006; Dekker et al., 2002). It is possible that the synthesis of these domains rich in ADAT amino acids require the presence of I34-modified tRNAs for prodder decoding.

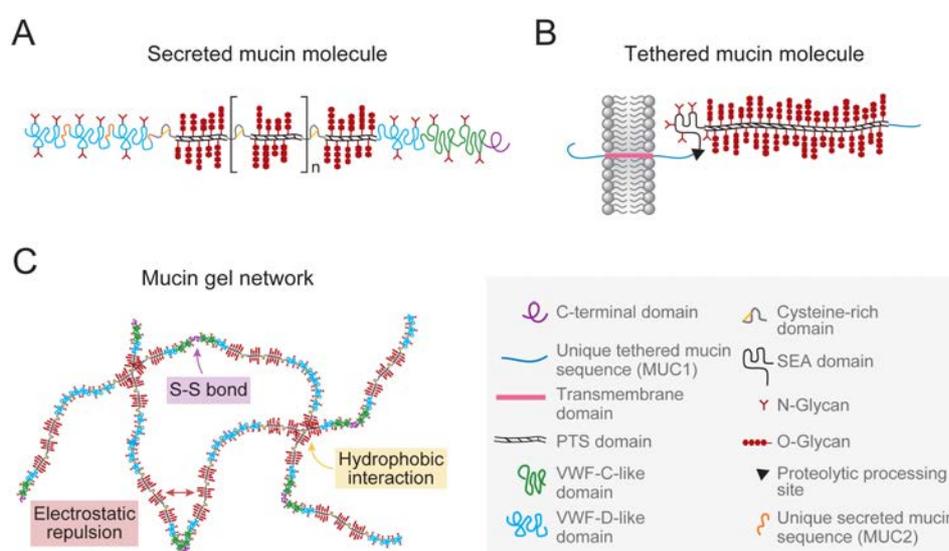


Figure 1.17. Mucin structure and oligomerization. Schematic representation of the general domain architecture of secreted (**A**) and cell-surface associated (**B**) mucin proteins. (**C**) Polymerization of gel-forming secreted mucins. The N- and C-terminal cysteine-rich domains create disulfide (S-S) bonds that lead to multimerization and formation of a macromolecular network that is stabilized by electrostatic repulsion between the charged glycan chains and hydrophobic interactions. Adapted from (Wagner et al., 2018).

Two different groups can be identified within the mucin family: secreted mucins and cell-surface associated mucins. Most secreted mucins form complex structural networks through polymerization via disulfide bonds between their N- and C-terminal cysteine-rich domains (**Figure 1.17 A, B**). On the other hand, cell-surface associated mucins present large

extracellular domains and short C-terminal domains that anchor them to the plasma membrane (**Figure 1.17 C**) (Hatrup and Gendler, 2008; Thornton et al., 2008). While transmembrane mucins are restricted to vertebrates, secreted mucins that form extracellular gels are also widely found among invertebrates and have been identified in deeply branching metazoans such as Cnidaria, Porifera and Ctenophora, suggesting a much earlier origin of these proteins during eukaryotic evolution (Lang et al., 2007, 2016). Interestingly, metazoan ADAT stretches are the longest and the most enriched in ADAT-sensitive codons, two parameters that have been shown to positively correlate in this lineage (Rafels-Ybern et al., 2018). It is possible that the emergence of heterodimeric ADAT facilitated the synthesis of proteins containing ADAT stretches in the animal kingdom.

Of the mucin glycoproteins identified to date in humans, eight belong to the secreted subgroup (MUC2, MUC5AC, MUC5B, MUC6, MUC7, MUC8, MUC9, and MUC19), and twelve are membrane-associated (MUC1, MUC3A, MUC3B, MUC4, MUC11, MUC12, MUC13, MUC15, MUC16, MUC17, MUC20, and MUC21) (Bansil and Turner, 2018). The synthesis of mucins starts in the endoplasmic reticulum, where they undergo initial glycosylation, and continues through the Golgi to get further glycosylated and packaged in mucin granules that will fuse with the plasma membrane of secretory cells (Bansil and Turner, 2018). Mucins are expressed by goblet or mucous cells in the epithelial surfaces of organs and glands that are exposed to the external environment such as those found in the respiratory tract, the gastrointestinal tract, the reproductive tract or the ocular surface. However, each mucin appears to be expressed at specific sets of tissues (Corfield, 2015).

Alterations in the expression and glycosylation patterns of mucins have been involved in the pathogenesis of several diseases. In the airway, the gel-forming mucins MUC5AC and MUC5B are predominantly expressed. Overexpression of MUC5AC is associated to several respiratory disorders such as cystic fibrosis (CF), asthma and chronic obstructive pulmonary disease (COPD) (Ma et al., 2018; Rose and Voynow, 2006). Additionally, MUC5B has been shown to be required for mucociliary clearance in a KO mouse model (Roy et al., 2014). In the colon, MUC2 is the most abundant mucin and it is critical for mucus barrier formation and protection against pathogenic bacteria. Mice genetically deficient in the *Muc2* gene present no intestinal mucus, which results in inflammation and cancer owing to closer contact of bacteria with the epithelium (Van der Sluis et al., 2006; Velcich et al., 2002). Opposed to the apparent protective role of MUC2, numerous studies on human tumors have shown that overexpression of mucins in a variety of cancers is linked to invasive proliferation and poor outcome (Hollingsworth and Swanson, 2004; Kufe, 2009). Moreover, aberrantly glycosylated mucins are observed during cancer, with the formation of specific glycan structures that have been associated to metastatic phenotypes (Chugh et al., 2015).

1.6.2. Dystroglycan

Dystroglycan is a cell membrane protein that binds ECM-associated factors in a variety of mammalian tissues (Durbeej et al., 1998). It is composed of two non-covalently bound protein subunits derived from a proteolytically cleaved precursor polypeptide that is translated from a single mRNA encoded in the *DAG1* gene (**Figure 1.18 A**) (Akhavan et al., 2007; Holt et al., 2000). The α -subunit (α DG) is an extracellular protein that undergoes extensive O-linked glycosylation at its mucin-like domain rich in serine and threonine residues (reminiscent of the PTS domain found in mucins), and functions as a receptor for ligands such as laminin. The β -subunit (β DG) is a membrane-spanning protein whose cytoplasmic domain interacts with the dystrophin complex, which is bound to the intracellular actin network (Huang et al., 2000; Ibraghimov-Beskrovnaya et al., 1992). Therefore, dystroglycan plays an important role in connecting cytoskeletal elements to the ECM.

Dystroglycan is expressed by cells of a broad range of tissues including skeletal muscle, nervous system, digestive tract, kidney, skin and reproductive organs (Durbeej et al., 1998). It has been implicated in early mouse development, structure and function of the central nervous system, myelination and nodal architecture of peripheral nerves, epithelial morphogenesis, cell adhesion and signaling (Barresi and Campbell, 2006).

Hypoglycosylation of human α DG is the major cause of a spectrum of human recessive genetic disorders called α -dystroglycanopathies, which are associated to progressive muscular dystrophy, brain malformations, and intellectual disability (Nickolls and Bönnemann, 2018). This is due to the critical role the attached glycans play in mediating interactions between α DG and its extracellular ligands (**Figure 1.18 B**) (Briggs et al., 2016). As mentioned above, α -dystroglycanopathies are characterized by degenerative muscular dystrophy, but they also exhibit a series of symptoms of variable severity ranging from profound brain and eye abnormalities accompanied by hydrocephaly that eventually lead to infant death (Walker-Warburg syndrome) (Cormand et al., 2001), to less severe brain structural defects frequently associated with hypotonia, developmental delay, intellectual disability, and retinal disorders (Godfrey et al., 2011). Several studies conducted in animal models have shown that the absence of dystroglycan results in disrupted extracellular matrix and aberrant neuronal migration in the brain, retina and spinal cord, suggesting an important role for dystroglycan in the maintenance and integrity of the ECM. Additionally, a connection between synaptic defects and dysfunctional dystroglycan has also been drawn (Nickolls and Bönnemann, 2018).

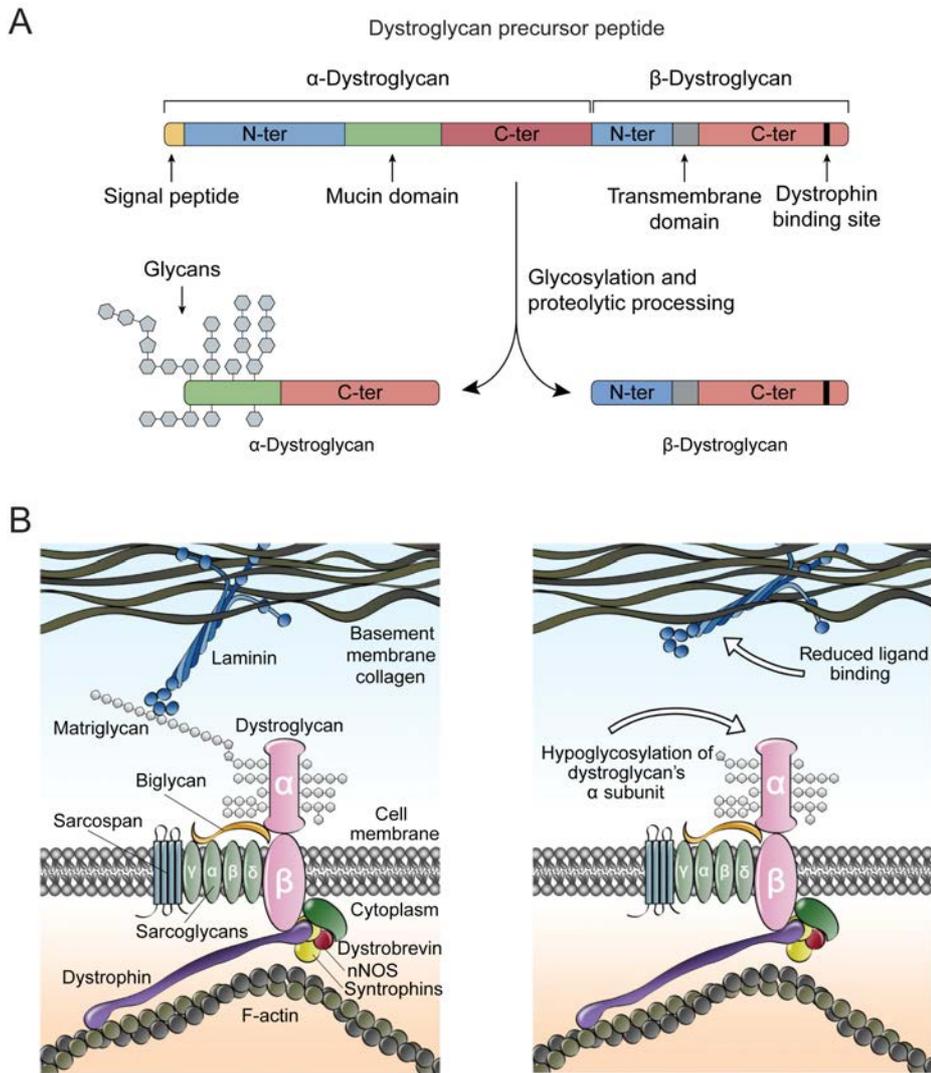


Figure 1.18. Dystroglycan biosynthesis, structure and role in pathogenesis. (A) Schematic representation of dystroglycan biosynthesis and domain architecture. The mRNA encoded by the *DAG1* gene contains a single open reading frame encoding both α DG and β DG as a precursor protein that undergoes proteolytical cleavage to generate the separate subunits. α DG contains a mucin-like domain flanked by two globular domains. While the N-terminal (N-ter) domain is critical for post-translational modification at the endoplasmic reticulum (ER) and Golgi apparatus, where it is ultimately removed, the C-terminal (C-ter) domain anchors α DG to its β DG partner by non-covalent bonding. The heavily O-linked glycosylated mucin-like domain is required for α DG interaction with its ligands. β DG consists of a transmembrane domain and a cytoplasmic C-terminus. (B) Molecular mechanisms of α -dystroglycanopathies. Left panel shows dystroglycan intracellular interactions with the dystrophin complex and extracellular binding to laminin under normal conditions. Right panel represents the pathological state in which hypoglycosylation of α DG results in loss of laminin binding and disruption of cell-ECM interactions. Adapted from (Nickolls and Bönnemann, 2018).

1.6.3. Syndecan-3

Syndecans are type I transmembrane cell surface heparan sulfate proteoglycans. They constitute an ancient family of proteins that is conserved across metazoan species (Ozbek et al., 2010), with reported presence in organisms belonging to the early-diverging metazoan phyla Cnidaria, Porifera and Ctenophora (Chakravarti and Adams, 2006; Draper et al., 2019). However, whereas mammals encode four family members (syndecan-1, -2, -3, and -4), invertebrate organisms contain a single one (Couchman et al., 2015). Syndecans have a common structural architecture consisting of an N-terminal ectodomain to which one or more glycosaminoglycan (GAGs) chains are covalently attached, a single transmembrane domain, and a C-terminal cytoplasmic domain (**Figure 1.19**). Each domain carries out different functions. While the ectodomain mediates direct cell-cell adhesions and cell-ECM interactions via its GAGs chains, the transmembrane domain is critical for the dimerization of the protein in homodimers, and the cytosolic domain is involved in the regulation of cell adhesion and migration through interactions with several intracellular kinases and the actin cytoskeleton (Gondelaud and Ricard-Blum, 2019). Interestingly, syndecan-3 is the only family member containing a mucin-like domain located between the glycosaminoglycans attachment sites that, in this case, does not undergo glycosylation (**Figure 1.19**) (Gould et al., 1992; UniProt Consortium, 2019). Information regarding the tertiary structure of syndecans is extremely limited, probably due to their high enrichment in intrinsically disordered regions (Leonova and Galzitskaya, 2015; Peysselson et al., 2011).

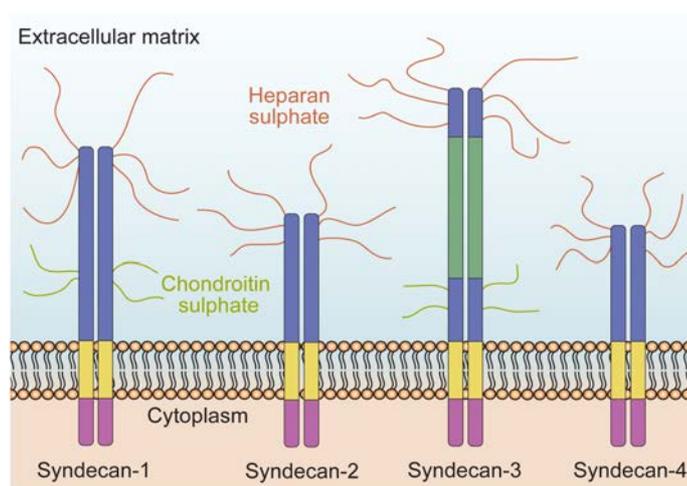


Figure 1.19. Schematic representation of the four vertebrate syndecans. Shown are the domain architecture and the pattern of glycosaminoglycan modification for the four syndecans. Protein domains are color coded as follows: cytoplasmic domain, magenta; transmembrane domain, yellow; ectodomain, blue; mucin-like domain, dark green. Heparan sulfate and chondroitin sulfate glycosaminoglycan chains are shown in brown and green, respectively. Adapted from (Manon-Jensen et al., 2010).

The four syndecan family members typically display tissue-specific distributions, with most cell types expressing at least one of them in vertebrates. Syndecan-1 is widely present in epithelial and plasma cells, syndecan-2 is mainly expressed in cells of mesenchymal origin including fibroblasts and smooth muscle cells, syndecan-3 is primarily found in neural and developing musculoskeletal tissues, and syndecan-4 exhibits a much broader distribution (Afratis et al., 2017). Despite the multiple roles attributed to syndecans in development, inflammation, angiogenesis, tissue regeneration and tumor progression (Afratis et al., 2017; Gondelaud and Ricard-Blum, 2019), almost no developmental defects have been observed in syndecan-1 or -4 null mice (Couchman et al., 2015). Interestingly, different reports have shown that syndecan-3 knockout mice present brain neurodevelopmental defects in the form of a disorganized structure of the cerebral cortex, which results from impaired neural migration along the glial fibers (Bespalov et al., 2011; Hienola et al., 2006). This phenotype is likely due to the absence of syndecan-3 as a receptor for ECM-associated factors that bind to its heparan sulfate and guide neurite outgrowth such as pleiotrophin (heparin-binding growth-associated molecule) or Glial cell-derived neurotrophic factor (Kinnunen et al., 1996, 1998). The fact that the invertebrate single syndecan is also widely present in the developing nervous system and its knockout in *C. elegans* and *D. melanogaster* exhibits a neurodevelopmental phenotype is perhaps indicative of a close relationship between mammalian syndecan-3 and the ancestral gene (Gysi et al., 2013; Steigemann et al., 2004).

In summary, a common feature of these extracellular matrix proteins, as well as of other proteins that contain ADAT stretches, is the presence of structural domains similar to the mucin PTS domains rich in proline, threonine, and serine. In some cases, these domains are important for the correct attachment of glycan structures, which have been shown to play important roles in development and cancer. The fact that many of the symptoms associated with an altered expression or glycosylation of dystroglycan and syndecan-3 resemble those linked to the ADAT3 V144M phenotype could indicate a connection between the translation of these proteins and a modulation in the levels of I34.

2. MATERIALS AND METHODS

2.1. General laboratory material

All chemicals were obtained from Sigma-Aldrich or Panreac, unless otherwise indicated. Glassware was obtained from VWR International and plastic items were purchased from Labclinics S.A.

2.2. General laboratory solutions

Water was used after purification by Milli-Q reagent Grade Water Ultrafiltration System (Millipore). Nuclease-free water was purchased from ThermoFisher (AM9937) and absolute ethanol (141086.1214), methanol (131091.1214) and isopropanol (131090.1214) were supplied by Panreac.

1× PBS	137 mM NaCl, 2.7 mM KCl, 10 mM Na ₂ HPO ₄ , 2 mM KH ₂ PO ₄ , pH 7.4
1× PBS-T	1× PBS + 0.05% (v/v) Tween-20
1× TBS	25 mM Tris base, 200 mM NaCl, 2 mM KCl, pH 7.4
1× TBS-T	1× TBS + 0.05% (v/v) Tween-20
0.5× TBE	44.5 mM Tris-HCl pH 8.3, 44.5 mM boric acid, 1 mM EDTA (pH 8), pH 8.3

2.3. Oligonucleotides

The oligonucleotides used in this study are listed in the following tables.

Table 2.1. Oligonucleotides (5' → 3') used in this study.

Gene/target	Purpose	Forward (5' → 3')	Reverse (5' → 3')
<i>ADAT2</i>	Amplification from human cDNA for cloning into pEGFP-N1 vector	TTATAAGCTTATGGAGGCGAAGGCCG CA	AACTGGTACCCAAGATTTCTGACATT CCTTTTTCCG
<i>ADAT2</i>	Amplification from human cDNA for cloning into pEGFP-C1 vector	TTATAAGCTTTGGAGGCGAAGGCCGC AC	TATTGGTACCTCAAGATTTCTGACAT TCCTTTTTCC
<i>ADAT3</i>	Amplification from human cDNA for cloning into pEGFP-N1 vector	ATATAAGCTTATGATCCTCTGCTCCCG TCT	ATATAAGCTTTGATCCTCTGCTCCC GTCTC
<i>ADAT3</i>	Amplification from human cDNA for cloning into pEGFP-C1 vector	TATAGGTACCGTGTCTCGGGTCCAGCC	TATAGGTACCTACGTGTCTCGGGTCC AG
<i>ADAT2</i>	Amplification from human cDNA for InFusion cloning into pPEU17 vector	AAGTTCTGTTTCAGGGCCCCGAGGCCG AAGGCGGCACCCA	ATGGTCTAGAAAGCTTTAAGATTTCT GACATTCCTTTTTCCGAAC
<i>ADAT2</i>	Amplification from human cDNA for InFusion cloning into pPEU19vector	AGGAGATATACCATGGAGGCGAAGGC GGCACCCAA	CTTCCAGACCGCTTGAAGATTTCTG ACATTCCTTTTTCCGAAC
<i>ADAT3</i>	Amplification from human cDNA for InFusion cloning into pPEU17 vector	AAGTTCTGTTTCAGGGCCCCGATCCTCT GCTCCCGTCTCTGTCT	ATGGTCTAGAAAGCTTTACGTGTCTG GGGTCCAGCCAGCG
<i>ADAT3</i>	Amplification from human cDNA for InFusion cloning into pPEU19 vector	AGGAGATATACCATGATCCTCTGCTCC CGTCTCTGTCT	CTTCCAGACCGCTTGACGTGTCTGGG GTCCAGCCAGCG
<i>ADAT3 V144M</i>	Quikchange site-directed mutagenesis primers for ADAT3 V144M (GTG → ATG) in pEGFP-N1/C1 and pPEU17/19 vectors	CCTTCCTGGTGCCCATGCCCGCCCGG CCG	CGGCCGGGCGGGCATGGGCACCA GGAAGG
pEGFP-N1	pEGFP-N1 MCS sequencing	CGCAATGGGCGGTAGGCGTG	TCAGGGGGAGGTGTGGGAGG
pEGFP-C1	pEGFP-C1 MCS sequencing	CATGGTCTGCTGGAGTTCGTG	GGTTCAGGGGGAGGTGTGGG
pPEU17 and pPEU19	pPEU17 and pPEU19 MCS sequencing	TAATACGACTCACTATAGGG	TCGATCTCAGTGGTATTTGTG
<i>ADAT2</i>	Amplification of synthetic codon optimized ADAT2-RBS-ADAT3 genes for cloning into pOPINFS vector and expression in <i>E. coli</i>	AAGTTCTGTTTCAGGGCCCCGATGGAAG CAAAAGCAGCACCGAAACC	GGTGGCTCCAGCTAGCGGTATCCG GATCCAGCCAACGACAC
<i>ADAT3 V128M</i>	Quikchange site-directed mutagenesis primers for codon optimized ADAT3 V128M (GTG → ATG) in pOPINFS	CCGTTTCTGGTTCCGATGCCTGCCCGT CCGC	GCGGACGGGCAGGCATCGGAACCA GAAACGG

Table 2.2. Oligonucleotides (5' → 3') used in this study for RT-PCR analysis.

Gene/target	Purpose	Forward (5' → 3')	Reverse (5' → 3')
<i>ADAT2</i>	qRT-PCR	AAGGCTGAGATTGGAAGCTG	AGCAATTCACCATGAGCTGT
<i>ADAT3</i>	qRT-PCR	AGCCTGACCGTGGATTTTCAG	GACCGAGTCCAGGAAAGCTC
<i>DAG1</i>	qRT-PCR	TCGAGTGACCATTCCAACAGA	GCACACCCTTATCAGTGTCAA
<i>SDC3</i>	qRT-PCR	CCTGACATCCCTGAGAGGAG	GTGGTCAGGAAGGTCTCTGG
<i>MUC1</i>	qRT-PCR	GTGCCCCCTAGCAGTACCG	GACGTGCCCCCTACAAGTTGG
<i>MUC5AC</i>	qRT-PCR	CAACCCCTCCTACTGCTACG	CTGGTGCTGAAGAGGGTCAT
<i>MUC5B</i>	qRT-PCR	GACGCTGGTGTTGGCTCT	GGTGTGCCCTGCATTCTC
<i>SPDEF</i>	qRT-PCR	AAGTGCTCAAGGACATCGAGA	AGGAGCCACTTCTGCACATT
<i>AGR2</i>	qRT-PCR	ATGAGTGCCCACACAGTCAA	AGGAGGACAAACTGCTCTGC
<i>GCNT3</i>	qRT-PCR	CACGCCTTGTGAAGAGATCA	GAATGGGAGACAGGTGGAGA
<i>AREG</i>	qRT-PCR	TGGTGCTGTCGCTCTTGATA	ACTGTGGTCCCCAGAAAATG
<i>IRE1</i>	qRT-PCR	GCCGAAGTTCAGATGGAATC	AAGGCCGATGACAAAGTCTG
<i>DPP4</i>	qRT-PCR	CTCAGCTCAGTCACCAATGC	TTTCTTGTGTTGCCCATGTC
<i>COX19</i>	qRT-PCR	TTTTGAAAATGCTTTGTGCAG	CCAAATCCCAGTTTCTCCAA
<i>METTL7A</i>	qRT-PCR	TGCTCTGTGAAGAACCAGGA	GCCACATGCTCCATGAAATA
<i>T. thermophila</i> tRNA ^{Leu} _{AAG}	qRT-PCR	AGTGGCCGAGCGGTCTAAG	TGGTGATGAAGGCGAGATTC
<i>GAPDH</i>	qRT-PCR	ATGGGTGTGAACCATGAGAAG	CTAAGCAGTTGGTGGTGCAG
<i>SDHA</i>	qRT-PCR	TGGGAACAAGAGGGCATCTG	CCACCACTGCATCAAATTCATG
<i>XBP1</i>	RT-PCR	TTACGAGAGAAAACATGGC	GGGTCCAAGTTGTCCAGAATGC
<i>RPL19</i>	RT-PCR	ATGTATCACAGCCTGTACCTG	TTCTTGGTCTCTTCCTCCTTG

Table 2.3. Oligonucleotides (5' → 3') used in this study for *in vitro* transcription of tRNA^{Ala}_{AGC}.

Gene/target	Purpose	Sequence (5' → 3')
<i>H. sapiens</i> tRNA ^{Ala} _{AGC}	Oligo 1 for assembling tRNA ^{Ala} _{AGC} sequence into pUC19 vector / Forward primer for PCR amplification of <i>in vitro</i> transcribed tRNA ^{Ala} _{AGC}	AGCTTAATACGACTCACTATAGGGGGTGTGGCT
	Oligo 2 for assembling tRNA ^{Ala} _{AGC} sequence into pUC19 vector	CAGTGGTAGAGCGCGTGCTTAGCATGCACGAG
	Oligo 3 for assembling tRNA ^{Ala} _{AGC} sequence into pUC19 vector	GCCCCGGGTTCAATCCCCGGCACCTCCACCAGGG
	Oligo 4 for assembling tRNA ^{Ala} _{AGC} sequence into pUC19 vector	CGCTCTACCACTGAGCCACACCCCCTATAGTGAGTCGTATTA
	Oligo 5 for assembling tRNA ^{Ala} _{AGC} sequence into pUC19 vector	TTGAACCCGGGGCCTCGTGTCATGCTAAGCACG
	Oligo 6 for assembling tRNA ^{Ala} _{AGC} sequence into pUC19 vector / Reverse primer for PCR amplification of <i>in vitro</i> transcribed tRNA ^{Ala} _{AGC}	GATCCCCTGGTGGAGGTGCCGGGGA

Table 2.4. Oligonucleotides (5' → 3') used in this study for ribosome profiling.

Name	Purpose	Sequence (5' → 3')
28 nt RNA marker	Gel purification marker	AUGUACACGGAGUCGACCCGCAACGCGA
32 nt RNA marker	Gel purification marker	GGCAUUAACGCGAACUCGGCCUACAAUGUCGC
Linker	Randomized linker	rApp ¹ /NNNNCTGTAGGCACCATCAAT/3ddC ²
Standard RT primer	Reverse transcription	(Phos ³)AGATCGGAAGAGCGTTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGC(SpC18)CACTCA(SpC18 ⁴)TTCAGACGTGTGCTCTTCCGATCTATTGATGGTGCCTACAG
Forward PCR primer	PCR amplification	AATGATACGGCGACCACCGAGATCTACAC

¹rApp: Adenylation; ²ddC: Dideoxycytidine (ddC) is a 3' chain terminator that prevents 3' extension by DNA polymerases; ³Phos: Phosphorylation; ⁴SpC18: Hexa-ethyleneglycol spacer.

2.4. Antibodies

Table 2.5. Antibodies used in this study.

Antibody	Species	Source	Working dilution
6X His tag	Mouse	Abcam (ab18184)	WB 1:250
Actin	Mouse	DSHB (JLA20)	WB 1:1000
β -Actin	Mouse	Sigma-Aldrich (A1978)	WB 1:5000
ADAT2	Goat	Santa Cruz (sc-107385)	WB 1:1000, IF 1:100
ADAT3	Rabbit	Abcam (ab192987)	WB 1:1000, IF 1:50
AlexaFluor 488	Donkey (anti-mouse)	Molecular Probes (A21202)	Flow 1:250
AlexaFluor 488	Goat (anti-rabbit)	Molecular Probes (A11008)	IF 1:400
AlexaFluor 488	Donkey (anti-goat)	Molecular Probes (A11055)	IF 1:400
AlexaFluor 555	Donkey (anti-mouse)	Molecular Probes (A31570)	IF 1:400
Anti-goat HRP	Donkey (anti-goat)	Abcam (ab6885)	WB 1:10000
Anti-mouse HRP	Sheep (anti-mouse)	GE Healthcare (NA931)	WB 1:10000
Anti-rabbit HRP	Donkey (anti-rabbit)	GE Healthcare (NA934)	WB 1:10000
ATF4	Rabbit	Aviva Syst Biol (ARP37017_P050)	WB 1:10000
BiP	Rabbit	Cell Signal (1673177S)	WB 1:1000-
α DG	Mouse	Millipore (05-593)	WB 1:250
β DG	Mouse	DSHB (MANDAG1 (7A11))	WB 1:500
EGFR	Mouse	Thermo Fisher (AHR5062)	WB 1:250
phospho-EGFR	Rabbit	Cell Signal (2237)	WB 1:100
ERK	Rabbit	Cell Signal (9102)	WB 1:1000
phospho-ERK	Mouse	Cell Signal (9106)	WB 1:2000
GAPDH	Rabbit	Trevigen (2275-PC-100)	WB 1:10,000
GFP	Rabbit	Abcam (ab290)	WB 1:1000
MUC1 (VU4H5)	Mouse	Cell Signal (4538S)	WB 1:1000
MUC5AC (45M1)	Mouse	Thermo Fisher (MA1-38223)	IF 1:200, Flow 1:200, Dot blot 1:2000
PERK	Rabbit	Cell Signal (1673192S)	WB 1:1000
Strep-tag II	Rabbit	Abcam (ab76949)	WB 1:1000
Syndecan-3	Mouse	Santa Cruz (sc-398194)	WB 1:500
Tubulin	Mouse	DSHB (E7)	WB 1:1000
Vinculin	Mouse	Abcam (ab18058)	WB 1:5000

2.5. Vectors

Table 2.6. Constructs used in this study.

Vector	Description
pEGFP-C1 - ADAT2	N-ter eGFP-tagged ADAT2 transient overexpression in mammalian cells
pEGFP-C1 - ADAT3 wt	N-ter eGFP-tagged ADAT3 transient overexpression in mammalian cells
pEGFP-C1 - ADAT3 V144M	N-ter eGFP-tagged ADAT3 V144M transient overexpression in mammalian cells
pEGFP-N1 - ADAT2	C-ter eGFP-tagged ADAT2 transient overexpression in mammalian cells
pEGFP-N1 - ADAT3 wt	C-ter eGFP-tagged ADAT3 transient overexpression in mammalian cells
pEGFP-N1 - ADAT3 V144M	C-ter eGFP-tagged ADAT3 V144M transient overexpression in mammalian cells
pPEU17 - ADAT2	N-ter mCherry-tagged ADAT2 transient overexpression in mammalian cells
pPEU17 - ADAT3 wt	N-ter mCherry-tagged ADAT3 transient overexpression in mammalian cells
pPEU17 - ADAT3 V144M	N-ter mCherry-tagged ADAT3 V144M transient overexpression in mammalian cells
pPEU19 - ADAT2	C-ter mCherry-tagged ADAT2 transient overexpression in mammalian cells
pPEU19 - ADAT3 wt	C-ter mCherry-tagged ADAT3 transient overexpression in mammalian cells
pPEU19 - ADAT3 V144M	C-ter mCherry-tagged ADAT3 V144M transient overexpression in mammalian cells
pOPINFS - ADAT2/3 wt	Recombinant wt ADAT2/3 protein co-expression and purification in bacteria
pOPINFS - ADAT2/3 V128M	Recombinant ADAT2/3 V128M protein co-expression and purification in bacteria
MISSION non-mammalian shRNA control plasmid DNA (cat number: SHC002)	Stable expression of a non-target shRNA
MISSION ADAT2 shRNA plasmid DNA (cat number: SHCLND-NM_182503; clone id: TRCN0000050656)	Stable expression of an shRNA targeting ADAT2
VSV-G	3rd generation lentiviral envelope plasmid
PKG-PIR	3rd generation lentiviral packaging plasmid encoding Gag and Pol genes
RTR2	3rd generation lentiviral packaging plasmid encoding Rev gene

2.6. Cloning and site directed mutagenesis

For eGFP and mCherry fusion proteins, wild type human ADAT2 (NCBI: NM_182503.2) and ADAT3 (NCBI: NM_138422.3) were amplified from cDNA obtained from HEK293T cells using primers listed in **Table 2.1**. eGFP-tagged proteins were generated through directional cloning of PCR products into pEGFP-N1 (C-ter eGFP) and pEGFP-C1 (N-ter eGFP) vectors (BD Clontech). For mCherry-tagged proteins, the HTP In-Fusion Cloning system (BD Clontech) was used following the manufacturer's instructions. PCR products were cloned into pPEU17 (N-ter mCherry) and pPEU19 (C-ter mCherry) vectors (Berrow et al., 2007). For the generation of the fluorescent ADAT3 V144M construct, QuikChange Site-Directed mutagenesis kit (Agilent Technologies) was used with a set of primers (**Table 2.1**) on all four vectors encoding wt ADAT3. All constructs were validated by Sanger sequencing.

2.7. Cell culture and *in vitro* assays

2.7.1. Cell lines and culture conditions

Human cell lines HEK293T, SH-SY5Y, HeLa, HT-29 M6 and Cal27 were maintained in Dulbecco's Modified Eagle Medium (41966052, ThermoFisher) supplemented with 10% fetal bovine serum (FBS) (10270106, ThermoFisher), 100 U/mL Penicillin-Streptomycin (15140122, ThermoFisher) and 25 µg/mL plasmocin (ant-mpp, InvivoGen). FaDu cells were maintained in Eagle's Minimum Essential Medium (BE06-174G, Lonza) supplemented with 10% FBS, 100 U/mL Penicillin-Streptomycin, and 25 µg/mL plasmocin. Human bronchial epithelial (HBE1) cells were cultured in serum-free Ham's F12 medium (P03-4110, PAN Biotech) supplemented with 10 µg/mL insulin (I6634, Sigma-Aldrich), 10⁻⁶ M hydrocortisone (H0888, Sigma-Aldrich), 3.75 µg/mL endothelial cell growth supplement (E2759, Sigma-Aldrich), 25 ng/mL epidermal growth factor (E4127, Sigma-Aldrich), 3 × 10⁻⁸ M triiodothyronine (T6397, Sigma-Aldrich), 5 µg/mL transferrin (T0665, Sigma-Aldrich), 10 ng/mL cholera toxin (C8052, Sigma-Aldrich) and 100 U/mL Penicillin-Streptomycin. NCI-H292 cells were maintained in Roswell Park Memorial Institute (RPMI) 1640 Medium (A1049101, ThermoFisher) supplemented with 10% FBS, 100 U/mL Penicillin-Streptomycin and 25 µg/mL plasmocin. All cells were grown at 37 °C in a humidified atmosphere with 5% CO₂, and were periodically checked for mycoplasma contamination by PCR. The cell line HT-29 M6 was a gift from Dr Eduard Batlle (IRB, Barcelona), FaDu and Cal27 cell lines were a gift from Dr Salvador Aznar (IRB, Barcelona), HBE1 cell line was a gift from Dr James

Yankaskas (University of North Carolina at Chapel Hill) (Yankaskas et al., 1993), and NCI-H292 control and ADAT2 KD cells were obtained in collaboration with Dr Ana Pardo (CIMA, Universidad de Navarra).

2.7.2. Generation of lentivirus and stable cell lines

Lentivirus were produced in HEK293T cells seeded at 80-90% confluence in T75 flasks. After 8 to 10 h of plating, cells were transfected using 75 μ L of polyethylenimine (PEI) (23966-1, Polysciences) with a mixture of 1.5 μ g envelope vector (VSVG), 1.5 μ g Rev-expressing vector (RTR2), 4.5 μ g packaging vector (PKG-PIR), 7.5 μ g transfer vector and 1.5 mL 150 mM NaCl that was added to the cell culture media. 16 h later, cells were washed in PBS and the media was replaced. 48 h post-transfection, viral supernatants were harvested, cleared using a 0.45 μ m filter, supplemented with 10% FBS and 8 μ g/mL polybrene (hexadimethrine bromide) (H9268, Sigma-Aldrich), and added to the cell line to be infected. The same procedure was repeated 72 h post-transfection. Two days later, puromycin (ant-pr-1, Invivogen) at 2 μ g/mL was added to culture medium for selection of transduced cells. Vectors containing the shRNA sequences were obtained from the Sigma Human MISSION shRNA library: MISSION pLKO.1-puro non-mammalian shRNA control plasmid DNA (shCV; cat number: SHC002) and ADAT2 MISSION shRNA plasmid DNA (shADAT2; cat number: SHCLND-NM_182503; clone id: TRCN0000050656).

For the transduction of NCI-H292 cells, viral supernatants obtained from HEK293T cells were harvested as described above, cleared with a 0.45 μ m filter and concentrated by ultracentrifugation through a 20% sucrose cushion at 26,000 g for 2 h at 4 $^{\circ}$ C using a Beckman SW-28 rotor. Purified lentiviral particles were resuspended in PBS, aliquoted and stored at -80 $^{\circ}$ C until needed. Lentiviral titer was determined using QuickTiter™ Lentivirus Quantitation Kit (Cell Biolabs, VPK-107). NCI-H292 cells were infected at a MOI of 6 for 24 h, and puromycin at 2 μ g/mL was added to culture medium for selection of transduced cells two days later.

2.7.3. Amphiregulin treatment

For induction of mucous cell differentiation, NCI-H292 cells were seeded in 6-well plates at 6×10^5 cells per well and incubated at 37 $^{\circ}$ C for 3 days or until confluent. Cells were then

washed with PBS and incubated in the absence of puromycin with fresh medium containing either 50 nM amphiregulin (AREG) (A7080, ThermoFisher) or PBS for 2 days, when the treatment was repeated for another 2 days. After that period of time, cells were stained for immunofluorescence and flow cytometry analyses, and RNA was extracted for RT-qPCR measurements of gene expression.

2.7.4. Thapsigargin treatment

Thapsigargin (T9033, Sigma-Aldrich) was used as an activator of the unfolded protein response (UPR). HEK293T cells were incubated with 2 μ M thapsigargin in the absence of puromycin for time periods of 3 and 18 hours. Cells were subsequently lysed and protein extracts were loaded onto SDS-PAGE for Western blot analysis.

2.7.5. Cell growth assay

To measure cumulative cell growth, 1×10^6 cells were plated in triplicate in 10 cm plates with 8 mL of growth media. Every two days, and for a period of 10 days, cells were counted with a Countess automated cell counter (Invitrogen), diluted and redistributed again in new plates. The number of counted cells was added to that of the previous time point in the curve.

2.8. Protein expression, detection and isolation

2.8.1. Recombinant protein expression and purification

The cDNA sequences of ADAT2 and ADAT3 (NM_182503.2 and NM_138422.3, respectively) were codon optimized for expression in *E. coli* (ThermoFisher). Of the two different transcript isoforms of ADAT3 currently annotated in the NCBI (NM_138422.3 and NM_001329533.1), the shortest ADAT3 isoform was used given that the resulting protein had been previously validated for *in vitro* activity assays (Saint-Léger et al., 2016).

ADAT2 and ADAT3 were cloned into pOPINFS vector (Oxford Protein Production Facility) using In-Fusion cloning system (Clontech) with specific primers (**Table 2.1**). The final construct encodes an N-terminal His6-tagged ADAT2 and a C-terminal StrepII-tagged

ADAT3 under the control of independent promoters. ADAT mutant was generated by Quickchange II XL site-directed mutagenesis (200521, Agilent Technologies) using primers specified in **Table 2.1**. Sequences for both constructs were confirmed by Sanger sequencing.

BL21(DE3)pLysS chemically competent cells (ThermoFisher) transformed with pOPINFS-His₆-ADAT2-ADAT3-StrepII (wt ADAT3) or pOPINFS-His₆-ADAT2-ADAT3*-StrepII (ADAT3 mutant) were grown overnight in LB medium supplemented with 100 µg/mL ampicillin and 50 µg/mL chloramphenicol. The starting culture was diluted 1:20 in LB medium containing the same antibiotics, and grown at 37 °C with shaking to an OD₆₀₀ between 0.6 and 0.8. Protein expression was then induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 37 °C for 3 hours with shaking. Cells were harvested by centrifugation at 4,000 g at 4 °C for 20 minutes; the resulting cell pellet was washed once with PBS (40 mL PBS per L of culture; centrifuged at 4,000 g at 4 °C for 20 minutes) and stored at -80 °C.

ADAT2 and ADAT3 proteins were co-purified as heterodimers by Strep-tag affinity chromatography (Strep-tagged ADAT3-mediated purification). All procedures were done at 4 °C in buffers containing protease inhibitor cocktail (P8340, Sigma-Aldrich). Cell pellets were resuspended in StrepTrap binding buffer (100 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA), and 5 MU of bovine pancreas DNase I was added (260913, Merck). Cells were lysed in a cell disruptor (Constant Systems) at 25,000 psi. The cell lysate was clarified by centrifugation at 70,000 g for 1 hour, and the supernatant was filtered with a 0.45 µm syringe filter (25PE045, Frisette). The clarified lysate was then loaded onto a 5 mL StrepTrap HP column (GE Healthcare Life Sciences). The column was washed with 7 column volumes of binding buffer, and the bound protein was eluted with StrepTrap elution buffer (100 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 5 mM *d*-Desthiobiotin). Protein elution was monitored by measuring absorbance at 280 nm. Protein concentration was determined by measuring the absorbance at 280 nm with a NanoDrop ND-1000. The purified heterodimeric ADAT was stored at -80 °C in elution buffer with 10% (v/v) glycerol, and subsequently used for the appropriate analyses.

2.8.2. Size-exclusion chromatography analysis

For the evaluation of the stability of affinity-purified heterodimeric ADAT enzymes (ADAT2/3 wt and ADAT2/3 V128M), 100 µg of protein in 500 µL of SEC buffer (20 mM Tris-HCl pH 7.5, 1 M NaCl, 1 mM EDTA) were injected onto a Superdex 200 10/300 GL column

(GE Healthcare) using a Bio-Rad NGC FPLC system. Proteins were eluted in SEC buffer with a 0.5 mL/min flow rate, fractions of 0.5 mL were collected and analyzed by SDS-PAGE and WB for protein detection, and by 8 M urea-PAGE for RNA detection. Protein standards were used to estimate the molecular weight (M_w) of the eluted fractions by injecting 500 μ L of 8.9 mM lysozyme from chicken egg white (62970, Fluka), 8.4 mM ovalbumin from chicken egg white (A5503, Sigma-Aldrich), 12.5 mM lectin from *Phaseolus vulgaris* (Sigma-Aldrich), 1.5 mM catalase from bovine liver (C1345, Sigma-Aldrich), and 0.8 mM ferritin from equine spleen (178440, Calbiochem) in StrepTrap elution buffer.

SEC-MALS was performed using the same column and experimental conditions as above, in a Prominence liquid chromatography system (Shimadzu) connected to a DAWN HELEOS II MALS detector and an Optilab T-REX refractive index (RI) detector (Wyatt Technology). ASTRA 7 software (Wyatt Technology) was used for data processing and result analysis. A refractive index increment (dn/dc) value of 0.185 mL/g (typical for proteins) was assumed for calculations.

2.8.3. Mass Spectrometry analysis of proteins

Protein bands from polyacrylamide gel were reduced with DTT 10 mM for 30 min at 56 °C, and alkylated for 30 min in the dark with iodoacetamide (IAM) 55 mM. Then, in-gel digestion was performed with chymotrypsin (0.1 μ g/ μ L) in 50 mM NH_4HCO_3 at 25 °C overnight. The digestions were stopped by adding formic acid (FA). Peptides were extracted with 100% acetonitrile (ACN) and completely evaporated. The bands were reconstituted in 20 μ L 1% formic acid aqueous solution with 3% of ACN for MS analysis.

The nano-LC-MS/MS set up was as follows. Digested peptides were diluted in 1% FA + 3% ACN. Sample was loaded to 100 μ m \times 2 cm Acclaim PepMap100, 5 μ m, 100 Å, C18 (ThermoFisher) at a flow rate of 15 μ L/min using a Thermo Scientific Dionex Ultimate 3000 chromatographic system. Peptides were separated using a C18 analytical column (Acclaim PepMap® RSLC (75 μ m \times 50 cm, nanoViper, C18, 2 μ m, 100Å) (ThermoFisher)) with a 90 min run, comprising three consecutive steps with linear gradients from 1% to 35% B in 60 min, from 35% to 50% B in 5 min, and from 50% to 85% B in 2 min, followed by isocratic elution at 85% B in 5 min and stabilization to initial conditions (A= 0.1% FA in water, B= 0.1% FA in ACN). The column outlet was directly connected to an Advion TriVersa NanoMate (Advion) fitted on an Orbitrap Fusion Lumos™ Tribrid (ThermoFisher). The mass spectrometer was operated in a data-dependent acquisition (DDA) mode. Survey MS scans

were acquired in the Orbitrap with the resolution (defined at 200 m/z) set to 120,000. The lock mass was user-defined at 445.12 m/z in each Orbitrap scan. The top speed (most intense) ions per scan were fragmented in the CID cell (35% collision energy) and detected in the orbitrap (30,000 resolution). The ion count target value was 400,000 for the survey scan and 50,000 for the MS/MS scan. Target ions already selected for MS/MS were dynamically excluded for 15 s. Spray voltage in the NanoMate source was set to 1.70 kV. RF Lens were tuned to 30%. Minimal signal required to trigger MS to MS/MS switch was set to 50,000 and activation Q was 0.250. The spectrometer was working in positive polarity mode and singly charge state precursors were rejected for fragmentation.

A database search was performed with Proteome Discoverer software v2.1 (ThermoFisher) using Sequest HT search engine and SwissProt database [Human release 2018, E. coli, release 2017, user protein (ADAT.fasta) and contaminant database]. Searches were run against targeted and decoy databases to determine the false discovery rate (FDR). Peptide mass tolerance was 10 ppm and the MS/MS tolerance was 0.02 Da. Peptides with a q-value lower than 0.1 and a FDR < 1% were considered as positive identifications with a high confidence level.

2.8.4. Protein lysates and Western blotting

Cells were rinsed with PBS and collected in Radioimmunoprecipitation assay (RIPA) buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, 0.1% SDS, 1 mM Na₃VO₄, 5 mM Na₄P₂O₇, 50 mM NaF, and protease inhibitor cocktail (P8340, Sigma-Aldrich)) for the extraction of soluble proteins. Samples were incubated on ice for 30 min, spun down in a benchtop centrifuge at 4 °C for 30 min at maximum speed and quantified using Pierce™ BCA Protein Assay Kit (23227, ThermoFisher). Protein extracts were boiled for 5 min with 6× loading dye (375 mM Tris-HCl pH 6.8, 9% SDS, 50% glycerol, 9% beta-mercaptoethanol, and 0.03% bromophenol blue) (J61337.AC, Alfa Aesar) and resolved on SDS-PAGE. Protein gels were then either stained or transferred to polyvinylidene difluoride (PVDF) membranes (Immobilion-P, Millipore) at 250 mA for 90 minutes at 4 °C. Membranes were incubated with blocking buffer (TBS-T with 5% dry non-fat milk) for 45 min, and primary antibody diluted in blocking buffer was then added for either 1 h at room temperature or overnight at 4 °C. Membranes were washed for 10 min three times in TBS-T, incubated with secondary antibody in blocking buffer for 1 h at room temperature, and washed again three times for 10 min (a complete list of the antibodies used in this work is shown in **Table 2.5**). Finally, membranes were developed using the enhanced chemiluminescence method (ECL Select

Western Blotting Detection Reagent) (RPN2235, GE Healthcare) and visualized in an Odyssey Fc Imager (Li-Cor).

For soluble and insoluble protein extraction, cells were rinsed in PBS, collected using a cell scraper and counted. RIPA buffer was added to the harvested cells to obtain soluble proteins. The pellet resulting from RIPA homogenization, which contained insoluble proteins, was washed with RIPA buffer, centrifuged to remove the supernatant, and incubated with solubilization buffer (50 mM Tris-HCl pH 7.5, 50 mM DTT, 2% SDS, 8 M urea, and protease inhibitor cocktail) at 95 °C for 10 min. Soluble protein extracts were quantified as above, whereas the 8 M urea present in solubilization buffer made quantification of insoluble proteins incompatible with the kit specifications. Therefore, equal loading on SDS-PAGE of soluble and insoluble proteins was performed according to the soluble protein quantification obtained from the same cell pellet, and taking into account the volumes of lysis buffers used at each step. Soluble protein extracts were mixed with 2× soluble loading dye (100 mM Tris-HCl pH 6.8, 0.1% bromophenol blue, 20% glycerol, 2% SDS, and 8 M urea), while insoluble extracts were mixed with 2× insoluble loading dye (100 mM Tris-HCl pH 6.8, 0.1% bromophenol blue, and 20% glycerol). Proteins were separated on SDS-PAGE and further analyzed.

2.8.5. Dot blot analysis

Extracellular medium and cell lysates were spotted on nitrocellulose membranes (0.45 µm) using a Bio-Dot Microfiltration Apparatus (Bio-Rad) according to the manufacturer's guidelines. Membranes were then incubated with blocking solution (4% BSA, PBS-T) for 1 h at room temperature followed by incubation with anti-MUC5AC antibodies diluted in blocking solution overnight at 4 °C. Membranes were washed in PBS-T and incubated with secondary antibody in blocking solution for 1 h at room temperature. For the detection of β-Actin, cell lysates were separated on SDS-PAGE, transferred to nitrocellulose membranes, and processed as described above. Membranes were imaged in an Odyssey Fc Imager (Li-Cor). Quantification of dot blots was performed with ImageJ software (Schindelin et al., 2012).

2.8.6. Immunofluorescence

Immunofluorescence was performed on cells grown on glass coverslips. Cells were fixed with 4% paraformaldehyde for 15 min at room temperature, washed twice with PBS, and

permeabilized in blocking buffer (0.3% Triton-X100, 1% BSA, 1× PBS) for 20 min at room temperature. Staining with primary antibodies diluted in blocking buffer was carried out overnight at 4 °C, cells were then washed twice with PBS and incubated with Alexa Fluor conjugated secondary antibodies in the dark for 1h at room temperature. After 4',6-Diamidino-2-phenylindole dihydrochloride (DAPI) (D9542, ThermoFisher) staining, slides were mounted with Vectashield mounting medium (H-1000, Vector Laboratories). Images were acquired with a Zeiss LSM 780 confocal microscope and analyzed using ImageJ software (Schindelin et al., 2012). In the cases where brightness and contrast were adjusted, the changes were applied to all images equally for direct comparison.

For the quantification of signal intensities in nuclei and cytosol, at least 23 different images per condition were taken with a Plan-Apochromat 40×/1.3 Oil DIC M27 objective. Intensities in the two cellular compartments were quantified using custom-made macros in ImageJ. Quantification of MUC5AC signal was performed on at least 31 different images per condition taken with a Plan-Apochromat 10×/0.45 M27 objective, and using custom-made macros in ImageJ.

2.8.7. Live cell imaging

HEK293T and SH-SY5Y cells were plated at 9×10^5 cells/well in 6-well plates and, on the next day, were transfected with combinations of vectors expressing eGFP- or mCherry-tagged proteins (wt ADAT2, wt ADAT3 or ADAT3 V144M) at either C-ter or N-ter using Lipofectamine 2000 (11668019, ThermoFisher). Two days post-transfection, cells were incubated with 1 mL of 1.2 µg/mL Hoechst 33342 (B2261, Sigma-Aldrich) for 15 min at 37 °C for nuclei staining. 1× CellMask Deep Red Plasma Membrane Stain (C10046, ThermoFisher) was then added to cells for 5 min at 37 °C in order to visualize cell boundaries, a marker that would later be used to quantify single cell fluorescent intensities in nucleus and cytoplasm. Images were scanned sequentially for the mCherry, eGFP, Hoechst, and CellMask Deep Red channels in a Zeiss LSM 780 confocal microscope using a Plan-Apochromat 40×/1.3 Oil DIC M27 objective. For the quantification of the intensities of nuclear and cytosolic signals, 23 to 35 different images per condition were taken and analyzed using custom-made macros in ImageJ.

2.8.8. Flow cytometry

NCI-H292 cells were harvested using Ca⁺²-free PBS supplemented with 0.02% EDTA (131026.1209, Panreac) to avoid enzymatic degradation of extracellular proteins, washed in PBS, and resuspended in flow cytometry buffer (PBS containing 0.1% (w/v) saponin (S7900, Sigma-Aldrich), 1% (w/v) sodium azide (S2002, Sigma-Aldrich), and 10% FBS). Aliquots of 5×10^5 cells were then stained with primary antibody at 4 °C for 30 min, washed twice in flow cytometry buffer, and incubated with secondary antibody in the dark at 4 °C for 30 min. Cells were then washed twice in flow cytometry buffer, resuspended in cold PBS and analyzed on a FACSAria™ Fusion flow cytometer (BD Biosciences). Cells stained only with secondary antibody were used as negative control to set the gate.

2.9. RNA extraction, detection and synthesis

2.9.1. RNA isolation and cDNA synthesis

Total RNA was isolated from cells with TRIzol (15596026, ThermoFisher) according to the manufacturer's protocol. cDNA was obtained by retro-transcription from 0.5 µg of total RNA with High-Capacity RNA-to-cDNA Kit (438706, Applied Biosystems) and subsequently used to perform gene expression analyses by qRT-PCR, RT-PCR or PCR amplification for cloning purposes.

2.9.2. Real-time qPCR

Real-time quantitative PCR (RT-qPCR) was carried out on a QuantStudio® 6 Flex instrument (Applied Biosystems) using Power SYBR Green PCR Master Mix (4368702, Applied Biosystems) and specific primers listed in **Table 2.2**. Reactions were prepared following the manufacturer's instructions and 10 µL/well were loaded onto 384-well plates (4309849, Applied Biosystems). qRT-PCR cycle: 2min at 50 °C, 10 min at 95 °C (1 cycle); 15 s at 95 °C, 1 min at 60 °C (40 cycles; PCR step); 15 s at 95 °C, 1 min at 60 °C, 15 s at 95 °C (one cycle; Melting Curve step). Gene expression analysis was performed using the comparative ddCT method.

2.9.3. RT-PCR analysis of *XBP1* splicing

cDNA from cells incubated with or without the ER-stressor thapsigargin was used as a template for PCR amplification of a fragment of *XBP1* encoding the intron target of IRE1 ribonuclease activity. PCR reactions were performed with primers listed in **Table 2.2** under the following cycle conditions: 95 °C for 5 min; 95 °C for 1 min; 58 °C for 30 sec; 72 °C for 30 sec; 72 °C for 5 min with 35 cycles of amplification. Unspliced *XBP1* generates a 289 base pair (bp) amplicon, while spliced *XBP1* generates a 263 bp amplicon. Amplification of *RPL19* cDNA was used as a reference. PCR products were separated on 2% agarose gels run in 0.5× TBE buffer.

2.9.4. *In vitro* tRNA transcription and purification

Constructs encoding a tRNA^{Ala}_{AGC} gene preceded by a T7 promoter were obtained as previously described (Geslain et al., 2006). Briefly, a template for transcription of the tRNA was generated by assembly of six DNA oligonucleotides (**Table 2.3**) that were first annealed and ligated, and then inserted between HindIII and BamHI restriction sites of the pUC19 plasmid. BstNI-digested plasmids were then used for *in vitro* transcription using T7 RNA polymerase following standard protocols (Sampson and Uhlenbeck, 1988). Transcripts were separated on 8 M urea/10% polyacrylamide gel electrophoresis and visualized by ultraviolet (UV) shadowing. Full-length tRNA was excised from the gel, electroeluted, ethanol-precipitated and resuspended in nuclease-free water.

2.9.5. Electrophoretic mobility gel shift assay (EMSA)

In vitro purified tRNA^{Ala}_{AGC} was dephosphorylated with CIP (M0290, NEB), radiolabeled with γ -³²P (PerkinElmer) using T4 polynucleotide kinase (M0201, NEB) and refolded at 70 °C for 3 min. After cooling down at room temperature, the tRNA was incubated on ice in the presence of reaction buffer (50 mM HEPES pH 7.5, 1 mM MgCl₂, 50 mM KCl, 0.5 mM DTT, 2% (v/v) glycerol, and 2.5 μ M oligo(dT)₂₅) and increasing concentrations of ADAT (0, 1 or 3.5 μ M) for 15 minutes. Samples were separated on a 10% non-denaturing polyacrylamide (19:1) gel run in 0.5× TBE buffer (44.5 mM Tris-HCl pH 8.3, 44.5 mM boric acid, and 1 mM EDTA pH 8) at 150 V for 2 h at room temperature. The gel was dried for 1 h at 80 °C, exposed

to a PhosphorImager screen, and the signal was analyzed using ImageQuant software (GE Healthcare).

2.9.6. Deamination assays

Inosine formation on tRNA^{Ala}_{AGC} by purified human ADAT was evaluated using the Restriction Fragment Length Polymorphism (RFLP) based method as in (Wulff et al., 2017). Briefly, *in vitro* transcribed tRNA was refolded at 95 °C for 3 min followed by gradual cooling at room temperature. Folded tRNA at 3.3 μM was incubated at 37 °C for the indicated times in a 30 μL reaction mix containing deamination buffer (50 mM Hepes-KOH, 100 mM KCl, 1 mM MgCl₂, 0.1 mM EDTA, pH 8) and varying amounts of ADAT. tRNA was then purified using the miTotal RNA Extraction Miniprep System (VTR1002, Viogene) and eluted in 30 μL from the purification column with nuclease-free water. Reverse transcription of 100 ng purified tRNA was performed with 10 pmol of RT primer (**Table 2.3**) and High Capacity cDNA Reverse Transcription Kit (4368814, ThermoFisher) following manufacturer's instructions. 2 μL of cDNA were used in a 20 μL PCR reaction with Taq DNA Polymerase (10342046, ThermoFisher) and specific primers (**Table 2.3**), at 60 °C annealing temperature and 30 second extension for 35 cycles. PCR amplicons were digested with 10 U Bpu1102I (ER0091, ThermoFisher) in Tango buffer (BY5, ThermoFisher) for 12 h at 37 °C, and separated on 2% agarose gels. Gel bands were quantified using ImageJ software, and the ratio of cut to non-cut PCR product was obtained. Potential DNA contamination was ruled out by using as PCR templates: material from an RT reaction lacking reverse transcriptase, material from an RT reaction lacking tRNA template, or nuclease-free water.

2.10. Mucin isolation and glycosylation quantification

2.10.1. Mucin purification

Mucins were purified from HT-29 M6 shCV and shADAT2 cells at 21 days post-confluence by isopycnic density gradient as described in (Davies et al., 2012). Cells were washed once with PBS, followed by addition of 0.6 mL/cm² of ice-cold 6 M guanidinium chloride (GuHCl) extraction buffer (6 M GuHCl (A0860.0500, Panreac), 5 mM EDTA, 10 mM sodium phosphate buffer pH 6.5, 5 mM *N*-ethylmaleimide (NEM) (E3876, Sigma-Aldrich), and 0.1 mM phenylmethylsulphonylfluoridate (PMSF) (78830, Sigma-Aldrich). Cells were

then harvested using a cell scraper, transferred to a new tube, and stirred overnight at 4 °C. On the next day, extracts were centrifuged at 23,000 *g* for 45 min at 4 °C, supernatants were collected and dialyzed against ten volumes of 6 M GuHCl extraction buffer. Then, cesium chloride (CsCl) (A1126.0100, Panreac) was added to achieve a starting density of 1.39 g/mL and the solution was stirred at 4 °C until fully dissolved. The amount of CsCl needed to give the correct density was calculated according to following equation:

$$x = v(2.347\rho - 0.0318M - 1.347)$$

Where *x* is CsCl weight (g), *v* is the final volume (mL), ρ is the starting density (g/mL), and *M* is the molarity of GuHCl (4 M or 0.5 M).

Samples were transferred to polycarbonate tubes (355654, Beckman) and centrifuged at 36,000 rpm for 90 h at 15 °C, using a 50.2 Ti rotor (Beckman) in a Beckman Optima L-90K ultracentrifuge. After centrifugation, 26 fractions of 1 mL were collected from the top of the gradients and analyzed for density by weighting 0.5 mL of every fraction. To determine which fractions were enriched in mucins or in other proteins, a portion of each sample was ethanol precipitated to remove GuHCl, run on SDS-PAGE gels, and stained accordingly. Briefly, 50 μ L of each sample were mixed with 9 volumes of ice-cold ethanol 100%, incubated at -20 °C for at least 1 h, and centrifuged at 15,000 *g* for 10 min, 4 °C. Supernatants were carefully discarded and pellets were washed with ice-cold ethanol 100%, vortexed and centrifuged again at 15,000 *g* for 10 min at 4 °C. Pellets were then air dried, resuspended in protein loading dye, and loaded onto 6% SDS-PAGE gels. The presence of glycoproteins and mucins in the gels was determined with Pierce Glycoprotein Staining Kit (24562, ThermoFisher) following manufacturer's guidelines. Gels were then counterstained with BlueSafe (MB15201, NZYTech) for detection of all proteins. Fractions rich in mucins, and depleted of other proteins, were pooled and dialyzed against 0.5 M GuHCl buffer (0.5 M GuHCl, 5 mM EDTA, 10 mM sodium phosphate buffer, pH 6.5) before being subjected to a second CsCl density gradient centrifugation with a starting density of 1.5 mg/mL. Samples were centrifuged at 36,000 rpm for 90 h at 15 °C using a 50.2 Ti rotor in a Beckman Optima L-90K ultracentrifuge. Following centrifugation, 26 fractions were taken and analyzed as described above. Mucin-containing fractions were pooled, dialyzed in PBS, and concentrated for further analysis.

2.10.2. Glycoprofiling of mucin-type O-glycans

O-glycan release and procainamide labelling

Purified mucins (~ 200 µg) from shCV and shADAT HT-29 M6 cells, 250 µg of fetuin glycoprotein (GCP-FET-250U, Ludger) as positive control and 200 µL of Milli-Q water as negative control were used for detection of O-glycans. Samples were buffer exchanged into 0.1% trifluoroacetic acid (TFA) by washing with 0.1% TFA using centrifugal filtration tubes with a molecular weight cut-off of 10 kDa. O-glycans were released by hydrazinolysis at 60 °C for 6 h, re-N-acetylated, cleaned up with LudgerClean CEX cartridges (LC-CEX-H-01, Ludger), and dried. Released O-glycans were then labelled with procainamide (PROC) using LudgerTag Procainamide glycan labeling kit (LT-KPROC-24, Ludger) followed by clean up by passing the samples through LudgerClean S cartridges (LC-S-A6, Ludger) and elution in 1 mL water. Samples were dried and resuspended in 30 µL water.

Mass spectrometry

Procainamide labelled samples were analyzed by hydrophilic interaction chromatography ultra-high-performance liquid chromatography (HILIC-UHPLC) coupled to electrospray ionization mass spectrometry (ESI-MS) with fluorescence detection. Samples were injected in 15% aqueous/85% acetonitrile (injection volume 25 µL) onto an ACQUITY UPLC® BEH-Glycan 1.7 µm, 2.1 x 150 mm column at 40°C on a Ultimate 3000 UHPLC instrument with a fluorescence detector ($\lambda_{\text{ex}} = 250\text{nm}$, $\lambda_{\text{em}} = 428\text{nm}$; ThermoFisher). The running conditions used were: Buffer A was 50 mM ammonium formate made from LudgerSep N Buffer stock solution, pH4.4 (LS-N-BUFFX40, Ludger); Buffer B was acetonitrile (Acetonitrile 190 far UV/gradient quality) (H039, Romil). Gradient conditions were: 0 to 10 min, 85% B; 10 to 95 min, 85 to 57% B; at flow rate 0.4 mL/min; 95 to 98 min, 57 to 10% B at a flow rate from 0.4 to 0.2 mL/min; 98 to 99 min, 10% B at a flow rate of 0.2 mL/min. 99 to 100 min, 10 to 85% B at a flow rate 0.2 mL/min; 100 to 103 min, 85% B at a flow rate of 0.2 mL/min; 103 to 115 min, 85% B at a flow rate 0.4 mL/min. The UHPLC system was coupled on-line to an Amazon Speed ETD electrospray mass spectrometer (Bruker Daltonics, Bremen, Germany) with the following settings: source temperature 250 °C, gas flow 10 L/min; capillary voltage 4500 V; ICC target 200,000; maximum accumulation time 50 ms; rolling average 2; number of precursors ions selected 3, release after 0.2 min; positive ion mode; scan mode: enhanced resolution; mass range scanned, 200-1600; target mass, 600. A glucose homopolymer ladder labelled with procainamide (CPROC-GHP-30, Ludger) was used as a system suitability standard as well as an external calibration standard for GU allocation. All standards

were run in the same injection volume and conditions as the samples analyzed within the same sample set.

Data processing

ESI-MS and MS/MS data analysis was performed using Bruker Compass DataAnalysis 4.1 software. Glycan compositions were allocated by mass matching to the number of: Hexose (H), N-acetylhexosamine (N), Fucose (F), sialic acid Neu5Ac (S) and additional O-acetyl group (O-Ac) on sialic acid. MS/MS fragmentation was used to allocate the sequence of monosaccharides within the structure. Specific diagnostic ions were used to indicate the presence of specific groupings of monosaccharides. Procainamide labelled glucose homopolymer (CPROC-GHP-30, Ludger) was used as a system suitability standard as well as an external calibration standard for glucose unit (GU) allocation. Chromeleon 7 software (ThermoFisher) was used to allocate GU values to peaks. In some cases, the software split peaks that were not fully resolved, in these cases the % areas were added together so that the same peaks were compared throughout the analysis.

2.11. Analysis of protein translation

2.11.1. Polysome profiling

HT-29 M6 or HEK293T control and ADAT2 silenced cells grown in 15 cm plates were incubated with fresh DMEM containing 100 µg/mL cycloheximide (C4859, Sigma-Aldrich) at 37 °C for 2 minutes before harvesting. When specified, cells were treated with 100 µg/mL puromycin at 37 °C for 5 minutes prior to cycloheximide incubation. Cells were then washed with ice-cold PBS supplemented with 100 µg/mL cycloheximide and lysed in 450 µL of lysis buffer (10 mM Tris-HCl pH 7.4, 100 mM KCl, 10 mM MgCl₂, 1% Triton-X100, 2 U/mL Turbo DNase (AM2239, ThermoFisher), 2 mM DTT and 100 µg/mL cycloheximide). Lysates were incubated on ice for 5 minutes and then centrifuged at 8000 *g* for 5 minutes at 4 °C. Supernatants were transferred to clean tubes and absorbance at 260 nm (A₂₆₀) was measured with a NanoDrop ND-1000. Ten A₂₆₀ units of cell lysate were loaded onto a 10-50% (w/v) sucrose gradient composed of lysis buffer containing 0.5 mM DTT and lacking Triton-X100 and Turbo DNase, and generated by mixing on a Gradient Master (BioComp Instruments). Gradients were centrifuged at 35,000 rpm for 3 h at 4 °C using a Beckman SW41 rotor. Polysomes were monitored by following the absorbance at 245 nm (A₂₄₅) with an Econo UV monitor (BioRad), and 15 fractions of 800 uL were obtained. Prior to RNA extraction, fractions were spiked-in with 2 ng of *T. thermophila* tRNA^{Leu}_{AAG} for normalization

purposes. RNA was then purified with TRIzol (15596, ThermoFisher), retro-transcribed and further analyzed by RT-qPCR.

2.11.2. Ribosome profiling

Control and ADAT2 KD cells were cultured in 10 cm plates until they reached 90% confluency. Fresh medium was added before harvesting to ensure good cell fitness and, 2 hours later, cycloheximide at 100 µg/mL was added to the plates for 3 min at 37 °C. Cells were then washed once with ice-cold PBS supplemented with 100 µg/mL cycloheximide, and 200 µL of lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM MgCl₂, 100 mM NaCl, 1% Triton-X100, 1 mM DTT, 100 µg/mL cycloheximide) were added. Cells were harvested using a cell scraper, collected in a tube kept on ice and centrifuged at 10,000 rpm for 3 min at 4 °C in a benchtop centrifuge. The supernatant was transferred to a clean tube and RNA absorbance at 260 nm was measured. Twenty A260 units of cell lysate were digested with 700 U RNaseI (100 U/µL) (AM2294, ThermoFisher) (35 U RNase I/A260 unit of lysate) at 25 °C for 45 min with continuous mixing. Lysates were then immediately supplemented with 10 µL of SUPERase In (20 U/µL) (AM2696, ThermoFisher) and loaded onto linear 10-50% (w/v) sucrose gradients prepared in 20 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 100 mM NH₄Cl, 1 mM DTT, and 100 µg/mL cycloheximide on a Gradient Master (BioComp Instruments). Gradients were centrifuged at 35,000 rpm for 3 h at 4 °C using a Beckman SW41 rotor, and fractionated by upward displacement at 0.75 mL/min with continuous monitoring of absorbance at 254 nm. Fractions of 0.4 mL were collected, and those containing the monosome were supplemented with SDS at 1%, flash-frozen and stored at -80 °C. RNA was isolated from monosome fractions using the hot acid phenol method. Briefly, one volume of 65 °C warm acid phenol (A1624, Panreac) was added to a sucrose sample and vortexed, 5 µL of glycogen (AM9510, ThermoFisher) and 1/5 volume of chloroform were then added, the sample was centrifuged at 10,000 rpm for 5 min at room temperature and the aqueous phase was extracted. This procedure was repeated three times, and at the final step only chloroform was used for extraction followed by ethanol precipitation of the RNA. 20 µg of RNA were used for ribosome-protected footprint purification by size selection in 8 M urea/15% polyacrylamide gels. Gels were run at 400 V for 4 h in 1× TBE, stained with SYBR Gold (S11494, ThermoFisher) and bands between RNA oligonucleotide makers of 28 nt and 32 nt were excised. Ribosomal RNA was depleted from samples using Ribo-Zero Gold rRNA Removal Kit (MRZG12324, Illumina) according to the manufacturer's instructions, but using only half of the recommended amounts of reagents.

For RNA-seq, a fraction of the lysates used for ribosome-profiling was treated with 100 µg/mL proteinase K (AM2546, Ambion) for 10 min at 60 °C. Total RNA was then extracted with acid phenol and incubated with TURBO DNase (AM2238, Ambion) at 37°C for 30 min with shaking. Poly(A) RNA was isolated by two rounds of purification with Poly(A)Purist MAG kit (AM1922, Ambion) following the manufacturer's guidelines. Next, purified poly(A) RNA was fragmented by alkaline hydrolysis in 50 nM sodium bicarbonate (pH 9.2) and 1 mM EDTA for 20 min at 95 °C. Subsequent purification of hydrolyzed RNA was performed by ethanol precipitation and fragments of 50-80 nt were excised from 8 M urea/15% polyacrylamide gels run in 1× TBE.

Sequencing libraries were prepared essentially as in (Ingolia et al., 2012). Briefly, RNA was dephosphorylated at the 3' end using T4 Polynucleotide Kinase (M0201L, NEB) in the presence of 20 U SUPERase In for 1 h at 37 °C, extracted with acid-phenol:chloroform (5:1) and ethanol precipitated for 1 h at -80 °C. 3'-adapter ligation was then performed by first denaturing RNAs mixed with 0.75 µg adapter at 80 °C for 90 seconds, and then ligation at 22 °C for 4 h in a reaction mix containing T4 RNA ligase 2, truncated (M0242L, NEB), 25% PEG 8000, and 10 U SUPERase In. RNA was ethanol precipitated at -80 °C and separated on 8 M urea/15% polyacrylamide gels run at 250 V for 2 h in 1× TBE. Gels were stained with SYBR Gold and ligation products of around 50 nt were excised, purified, denatured at 80 °C for 90 seconds mixed with reverse transcription (RT) primer, and reverse transcribed with SuperScript III Reverse Transcriptase (18080044, ThermoFisher) at 55 °C for 40 min. Reverse transcription products of around 135 nt were purified from 8 M urea/10% polyacrylamide gels run at 300 V for 2.5 h in 1× TBE. Libraries were then circularized with CircLigase I ssDNA Ligase (CL4115K, Epicentre) at 60 °C for 3 h, and the enzyme was inactivated at 80 °C for 10 min. PCR amplification was performed with Phusion High-Fidelity DNA Polymerase (F530L, ThermoFisher), forward primer (**Table 2.4**) and standard index primer (NEBNext® Multiplex Oligos for Illumina) (E7335S, NEB), and 8 to 11 PCR cycles were used. PCR products of about 175 nt were isolated from non-denaturing 8% polyacrylamide gels run at 250 V for 1 h in 1× TBE. Purified libraries were quantified with Qubit dsDNA HS Assay (Q32851, ThermoFisher) and sequenced using a HiScanSQ system (Illumina), single read, 50 nt at the Max Plank Institute for Molecular Biomedicine in Münster, Germany.

2.11.3. Ribosome profiling sequencing data analysis

Sequencing data from ribosome footprints was pre-processed by trimming linker sequences from the 3' end of reads and alignment to human rRNA gene sequences, obtained from the NCBI RefSeq database, using Bowtie 1.0.0 (Langmead et al., 2009). Unmapped reads with lengths above 25 nt were then mapped to a reference transcriptome consisting of UCSC canonical transcripts of the *H. sapiens* genome hg19 build. In order to retain reads corresponding to initiating and terminating ribosomes during mapping, 18 nt of the 5' UTR and 3' UTR sequences were included for each transcript. One mismatch was allowed in the seed sequence during the alignment and only reads with unique matches to the references were retained for further analysis. Most mapped footprint reads were 29 to 33 nt long. Of the 21 – 35 million reads present in each library, 6 – 10 million mapped uniquely to CDSs.

Only reads with no mismatches were used for codon occupancy measurements. The offset of the P-site codon from the 5' end of the reads was inferred by examination of the cumulative distribution of reads 29 to 32 nt long mapping to the annotated AUG start codons for each transcript (i.e. reads corresponding to initiating ribosomes positioned at the P-site). Read coverage showed an offset of 12 to 13 nt upstream of the start codon, therefore we selected reads with a 5'-nt in the 0 and 2 frames, discarded the first nucleotide of reads beginning in the 2 frame, and used an offset of 12 nt from the 5'-end of each read to determine the P-site codon. Because cycloheximide is known affect ribosome density at the beginning of CDSs by allowing initiating ribosomes to bind mRNAs but inhibiting their elongation (Ingolia et al., 2009, 2011), therefore generating an artifactual accumulation of footprints in that area, we excluded reads mapping within the first 15 codons of CDSs. Codon frequencies at each ribosomal site were normalized to their average frequency in positions downstream of the A-site that have not yet been translated and, therefore, should have negligible effects on ribosome occupancy (Nedialkova and Leidel, 2015; Stadler and Fire, 2011).

Differential ribosome occupancy and mRNA abundance for gene-level measurements was performed with DESeq2 (Love et al., 2014). Translation efficiency analyses were done with the Xtail R-package (Xiao et al., 2016). Statistical significance was defined by a Benjamini-corrected *P*-value < 0.1, and low coverage genes were filtered out by requiring a mean value or normalized reads between compared samples of at least 20.

2.12. Statistical analysis

Statistical analysis was performed using Prism 8 and R. A two-tailed t-test was performed when comparing two groups with normal distributions. A Kruskal-Wallis test was performed when comparing populations without normal distributions. A non-linear regression was used to compare growth curves. Microscopy data in **Figure 2.9** was statistically analyzed to assess group differences using a linear model fitted to the contribution to total intensity of nuclear wt ADAT3 or ADAT3 V144M.

3. OBJECTIVES

The main goal of this thesis is to examine the influence of decoding with wobble inosine tRNA modification, catalyzed by ADAT, on the human proteome composition, with a special focus on the synthesis of extracellular matrix (ECM) proteins.

The specific aims that have been pursued are the following:

- Characterize the enzymatic activity, cellular distribution, and dynamics of human ADAT
- Define the impact of the ADAT3 V144M mutation on the enzyme biology
- Determine the effect of I34 deficiency in human cells
- Investigate the mechanisms whereby ADAT contributes to gene translation
- Examine the importance of ADAT for ECM protein synthesis in human cell lines

4. RESULTS

4.1. ACTIVITY AND STABILITY OF ADAT ARE COMPROMISED BY A MUTATION LINKED TO HUMAN INTELLECTUAL DISABILITY

4.1.1. Effect of an ADAT3 mutation on heterodimer structure and tRNA affinity

Given that the ADAT3 V144M mutation had been shown through molecular modelling to alter the tertiary structure of the ADAT3 protein (Alazami et al., 2013), we decided to examine its impact on the quaternary structure of human ADAT with the hypothesis in mind that an altered heterodimer formation could be critical to enzymatic activity, as previously reported in yeast (Tsutsumi et al., 2007). We initially did so by size-exclusion chromatography (SEC) analysis of affinity purified wild type ADAT (wt ADAT: wt ADAT2 / wt ADAT3) and mutant ADAT (ADAT*: wt ADAT2 / ADAT3 V144M) (**Figure 4.1** and **Figure 4.2**). Calibration of the SEC column was performed in order to provide a standard curve for molecular weight (Mw) estimation of the eluted ADAT species (**Figure 4.1 B**). As shown in **Figure 4.1 A**, wt ADAT eluted in two peaks at the expected elution volume, and the presence of both subunits was confirmed by SDS-PAGE and Western blot (**Figure 4.1 C**; fractions 23-32). Urea-PAGE analysis revealed the presence of nucleic acids of tRNA-like size in fractions 31-34 (**Figure 4.1 D**), suggestive of the presence of *E. coli* tRNA in the input injected into the SEC column which had been co-purified bound to ADAT in the previous affinity capture purification step. SEC-coupled Multi-Angle Light Scattering (MALS) analysis assigned a molecular weight (Mw) of roughly 60 kDa to the predominant molecular species in the first peak, consistent with the expected mass of the ADAT heterodimer, and a Mw of 24 kDa to the predominant species in the second peak, consistent with the mass of a tRNA molecule (**Figure 4.1 E**).

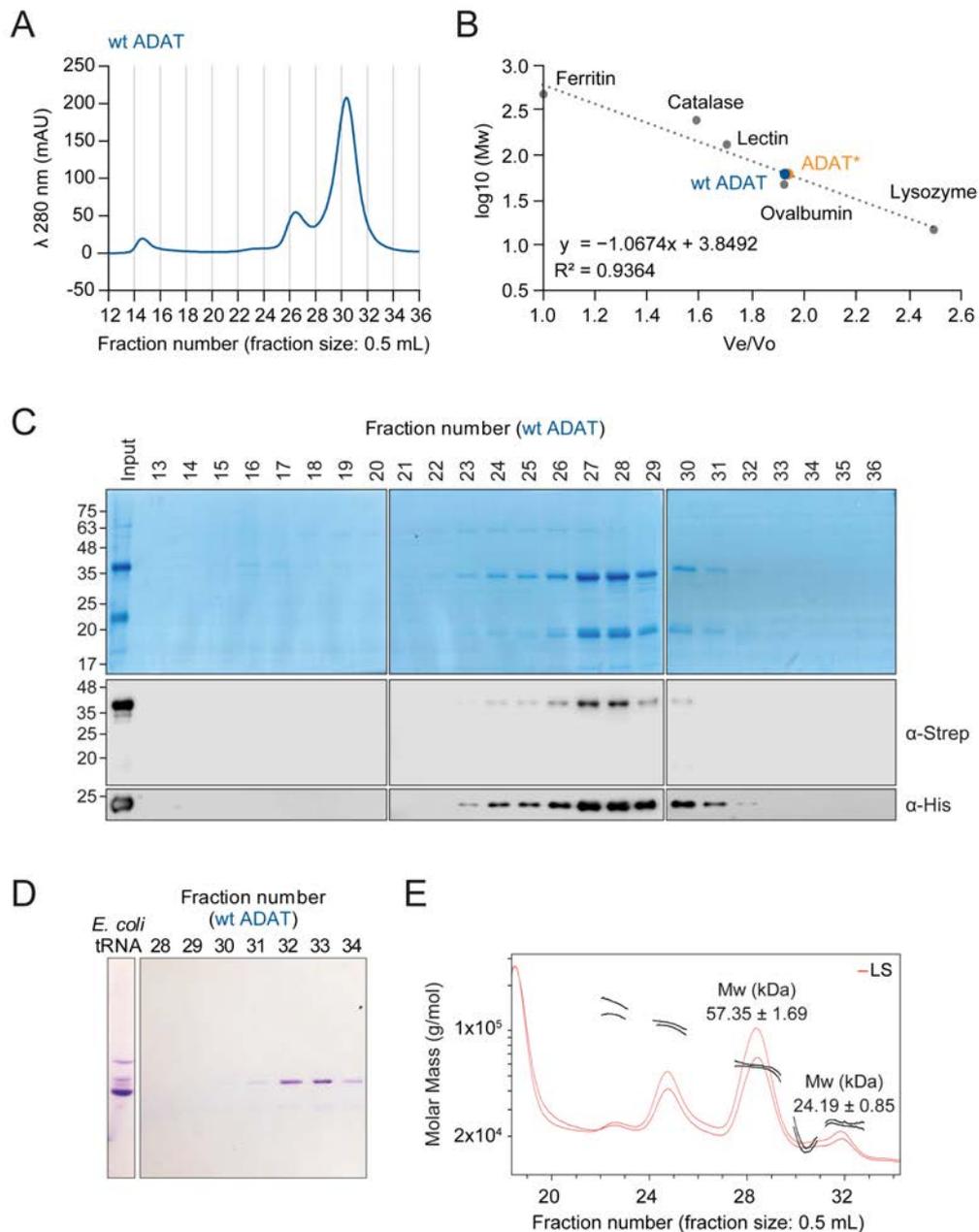


Figure 4.1. Analysis of wt ADAT quaternary structure. (A) Size-exclusion chromatogram of wt ADAT. Elution fraction numbers (fraction size: 0.5 mL) and UV absorbance at 280 nm are represented on x- and y-axis, respectively. (B) Calibration of size-exclusion chromatography (SEC) column and molecular weight estimation for ADAT. Shown are the log₁₀ (Mw) values against the observed relative elution volumes (V_e/V_o) for protein standards (grey), wt ADAT (blue), and ADAT* (orange). Void volume (V_o) corresponds to the elution volume (V_e) of ferritin (7.7 mL). (C) SDS-PAGE and Western blot of SEC-eluted fractions. Lanes are labelled with the fraction number. Mw for ADAT2 and ADAT3 are 23 kDa and 39 kDa, respectively. ADAT2 is His₆-tagged and ADAT3 is StrepII-tagged. (D) Urea-PAGE of SEC-eluted fractions. Lanes are labelled according to fraction number. *E. coli* total tRNA (average size: 75 nt) was used as Mw marker control (first lane). (E) Molecular weight determination of the primary species present in each fraction by SEC-MALS. Shown are the normalized output from light scattering signal (LS) (red) and molecular weight estimates across peaks (black) for two replicates. Fraction numbers are represented on the x-axis and Molar mass (g/mol) is represented on the y-axis. (Data for figure panels B and D-E was provided by Dr Helena Roura).

The mutant heterodimeric enzyme (ADAT*) was similarly detected as a double peak eluting in fractions 23-32 (**Figure 4.2 A**), with the late fractions co-eluting with tRNA-like size molecules as well (**Figure 4.2 B**). SDS-PAGE and Western blot analyses revealed that the first peak corresponded to the full-length ADAT heterodimer, while the second peak contained ADAT2 bound to a truncated form of mutant ADAT3 of a similar size to that of ADAT2 (**Figure 4.2 C**). The detection by Western blot of the C-terminal Strep-tag indicated an N-terminal truncation of mutant ADAT3, which we confirmed by mass spectrometry (**Figure 4.3**). The fact that the truncated mutant ADAT3 was already present in the protein sample injected into the SEC column (**Figure 4.2 C**; "Input" lane) ruled out any degradation events occurred during the chromatographic analysis, and suggested that *E. coli* proteases were responsible for the cleaved mutant ADAT3. Based on the observations made by Alazami and colleagues (Alazami et al., 2013) that the mutation altered the local structure of the protein, and on recent data obtained from *T. brucei* ADAT3 that indicates that the conserved Val residue is located in a region of low solvent access (McKenney, 2018), it is possible that the substitution by a Met residue induces the exposure of a novel proteolytic site for *E. coli* proteases. A large third peak was also observed at the void volume ($V_0 = 7.7$ mL, fractions 15-16), but neither SDS-PAGE nor Western blot analyses detected significant levels of ADAT2 or ADAT3 in those fractions (**Figure 4.2 C**).

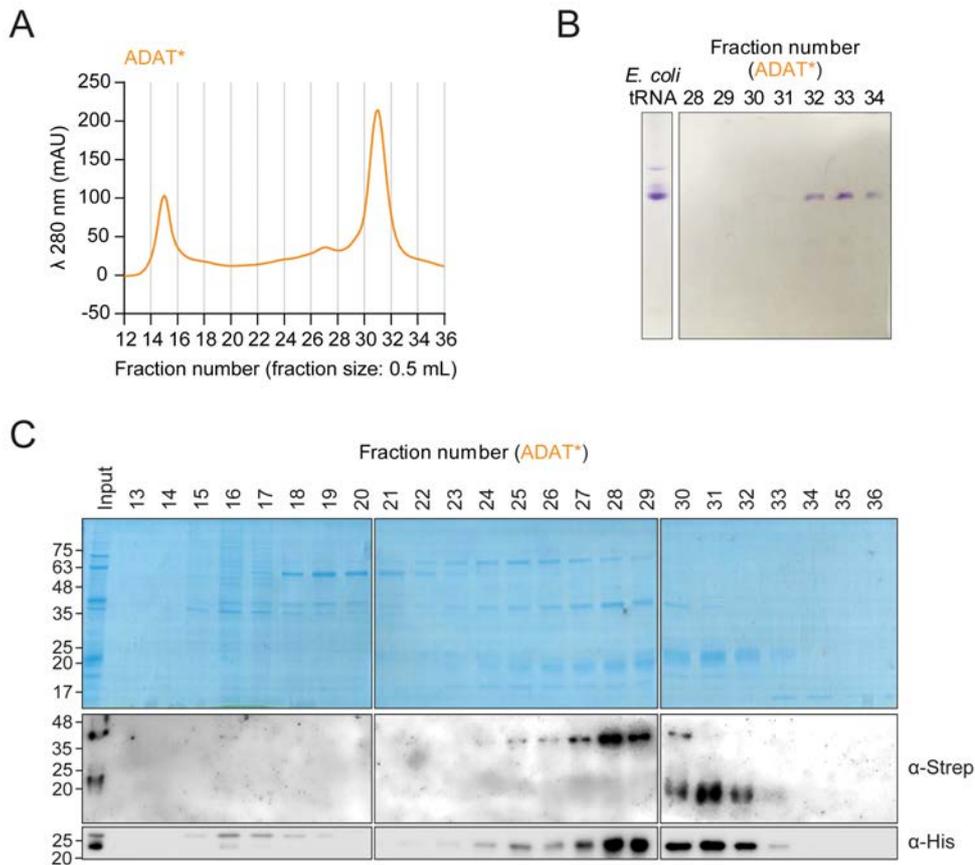


Figure 4.2. Analysis of mutant ADAT quaternary structure. (A) Size-exclusion chromatogram of ADAT*. Elution fraction numbers (fraction size: 0.5 mL) and UV absorbance at 280 nm are represented on x- and y-axis, respectively. (B) Urea-PAGE of SEC-eluted fractions. Lanes are labelled according to fraction number. *E. coli* total tRNA (average size: 75 nt) was used as Mw marker control (first lane). (Figure panel B was provided by Dr Helena Roura) (C) SDS-PAGE and Western blot of SEC-eluted fractions. Lanes are labelled with the fraction number. Mw for ADAT2 and ADAT3 V144M are 23 kDa and 39 kDa, respectively. ADAT2 is His₆-tagged and ADAT3 V144M is StrepII-tagged.



Figure 4.3. Identification of the proteolytic cleavage site in ADAT3 V144M. Mass spectrometry analysis of ADAT*. Left panel: BlueSafe stained SDS-PAGE of Strep-tag affinity purified ADAT*; bands containing full-length ADAT3* (band 1) and truncated ADAT3* (bands 2 and 3) were cut from the gel, digested with chymotrypsin and analyzed by mass spectrometry. Right panel: ADAT3* protein sequence coverage by peptides detected for each band. Peptides mapping to the N-terminal region of ADAT3* are not detected for truncated ADAT3* samples (bands 2 and 3).

We next evaluated the binding affinity of the ADAT proteins for tRNA by electrophoretic mobility shift assays (EMSA). Similar binding affinities of both wt ADAT and ADAT* for tRNA^{Ala}_{AGC} were observed (**Figure 4.4 A**). In addition, we found that the full-length (fractions 25-29) and truncated (fractions 31-33) forms of the mutant heterodimer bound tRNA^{Ala}_{AGC} to similar extents (**Figure 4.4 B**). Altogether, these results suggest that the V144M mutation does not affect the quaternary structure of ADAT, but may rather destabilize the tertiary structure of ADAT3, making it prone to proteolytic cleavage without affecting binding affinity for tRNA.

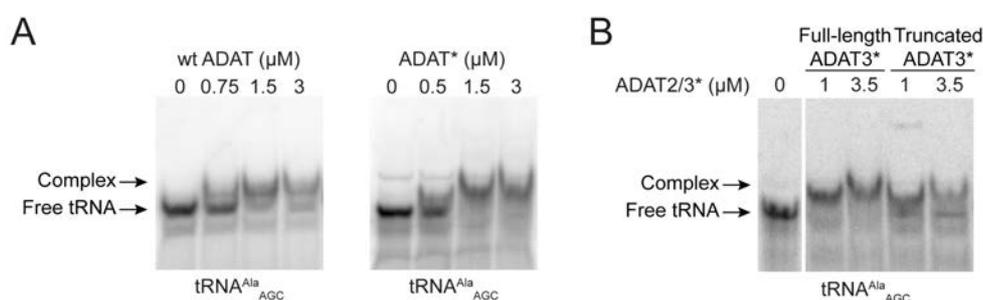


Figure 4.4. ADAT binding affinity for tRNA. (A) Electrophoretic mobility shift assays (EMSAs) with increasing amounts (0 to 3 μM) of wt ADAT (left panel) and ADAT* (right panel). Both heterodimers bind tRNA^{Ala}_{AGC} with similar affinity. (B) EMSAs for ADAT2-ADAT3* heterodimers corresponding to fractions 25-29 (ADAT2/full-length ADAT3*) and fractions 31-33 (ADAT2/truncated ADAT3*). Both mutant heterodimers bind tRNA^{Ala}_{AGC} with similar affinity. (Data obtained in collaboration with Noelia Camacho).

4.1.2. Characterization of the enzymatic activity of human ADAT

The direct impact of the ADAT3 V144M mutation on A34-to-I34 editing has not been demonstrated thus far. We set to characterize the deamination activity of purified wt ADAT and ADAT* on tRNA^{Ala}_{AGC} by the restriction fragment length polymorphism (RFLP) method (Wulff et al., 2017). This method takes advantage of the fact that, due to its similarity to guanosine, inosine directs the incorporation of cytidine by reverse transcriptase during cDNA synthesis. In brief, the procedure consists in reverse-transcription and PCR amplification of human tRNA^{Ala}_{AGC} transcripts that generate DNA fragments (~100 bp) containing a BlnI restriction site that is obliterated when the tRNAs are modified to I34 (Wulff et al., 2017). Thus, deamination of tRNA^{Ala}_{AGC} by ADAT results in a decrease in BlnI digestion products (~50 bp) that can be visualized by agarose gel electrophoresis (**Figure 4.5**). Using RFLP, we studied the deamination activity of wt ADAT and ADAT* at different concentrations over time. Complete deamination (92.3 ± 10.8 %) was obtained after 4 hours of incubation with

50 nM wt ADAT (**Figure 4.5 A, C**). However, a sixteen-fold higher ADAT* concentration was required to obtain the same result (**Figure 4.5 B**).

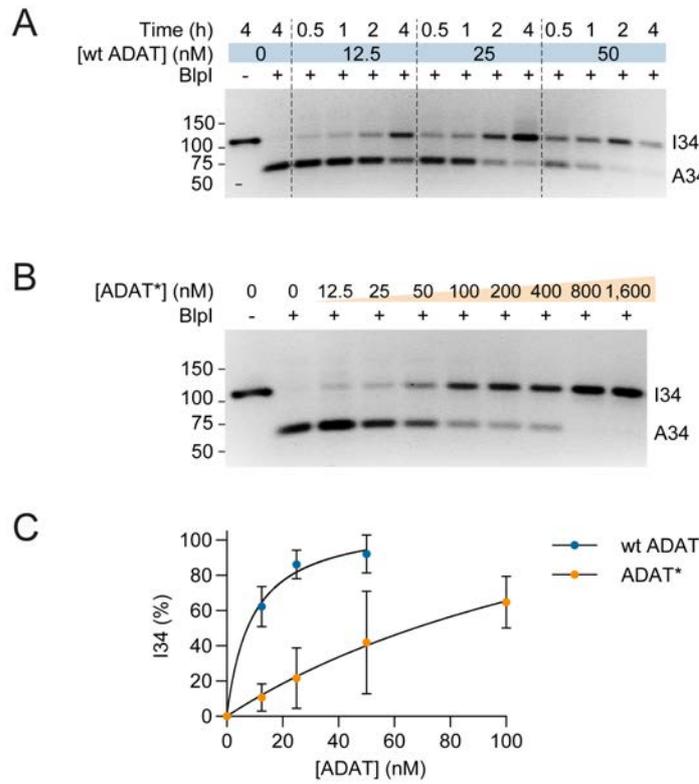


Figure 4.5. *In vitro* enzymatic activity of wt ADAT and ADAT* for tRNA^{Ala}_{AGC} by RFLP. (A) RFLP to assess the deamination activity of wt ADAT with 3.3 μ M of tRNA^{Ala}_{AGC}. Inosine formation was monitored at the indicated enzyme concentrations and time points. (B) RFLP to assess the deamination activity of ADAT* with 3.3 μ M of tRNA^{Ala}_{AGC}. Inosine formation was monitored at the indicated enzyme concentrations upon 4 hours of incubation. (C) Percentage of I34 formation against enzyme concentration after 4 hours of incubation for wt ADAT and ADAT*. Shown are mean values and associated SD of three independent experiments. Curve fittings were performed following a nonlinear hyperbolic regression model (wt ADAT: $R^2 = 0.9574$; ADAT*: $R^2 = 0.7356$). (Data provided by Dr Helena Roura Frigol ).

To further characterize the kinetics of wt and mutant ADAT, we analyzed the deamination rate for each heterodimer at increasing substrate concentrations while keeping the enzyme concentration constant (**Figure 4.6**). We decided to use 20 nM wt ADAT (**Figure 4.6 A, B**) and 100 nM ADAT* (**Figure 4.6 C, D**) in the deamination reactions because those were the concentrations at which either enzyme had catalyzed the conversion of about 60 % I34 (**Figure 4.5**), and were therefore at comparable contexts. Determination of kinetic constants of wt ADAT (K_m : 2.98 μ M, k_{cat} : 5.64 min^{-1}) produced values similar to those reported for *T. brucei* ADAT and *E. coli* TadA (Kim et al., 2006; Spears et al., 2011). We calculated a 10-fold decrease in k_{cat}/K_m for ADAT* with respect to wt ADAT (**Table 4.1**), a reduction caused

by a 5-fold drop in k_{cat} for ADAT* (K_m : 5.81 μM , k_{cat} : 1.07 min^{-1}). Although we cannot rule out an impact of the presence of the truncated form of ADAT3* in our activity assays, its estimated concentration (Figure 4.2 C; "Input" lane) cannot account for the observed 5-fold drop in k_{cat} calculated for ADAT*.

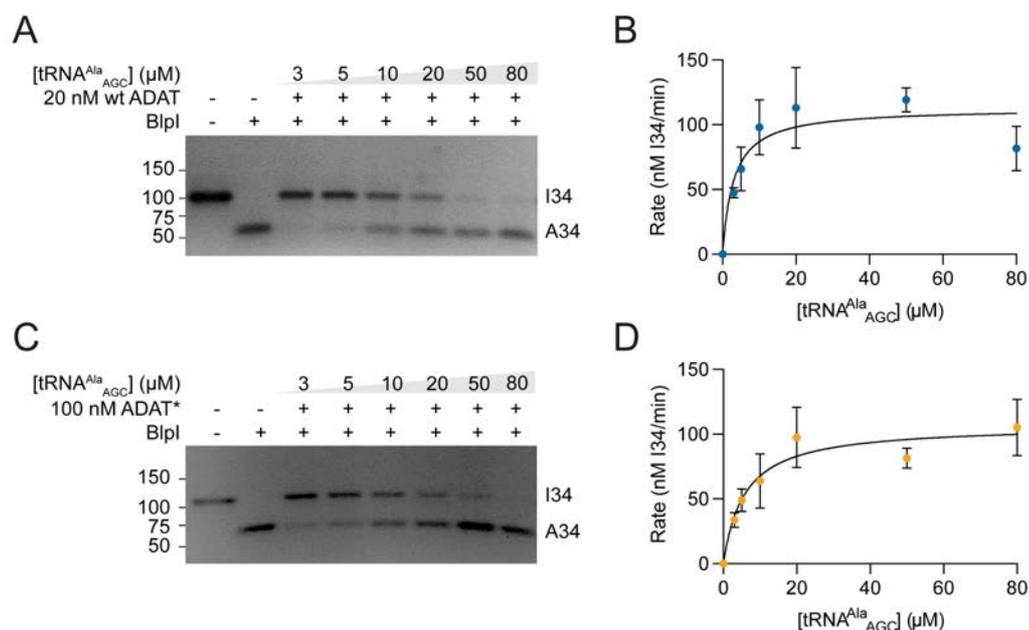


Figure 4.6. *In vitro* deamination assays to determine kinetic constants of wt and mutant ADAT for tRNA^{Ala}_{AGC} by RFLP. (A, C) Representative RFLP experiments to monitor the deamination activity of wt ADAT (A) or ADAT* (C) at varying tRNA^{Ala}_{AGC} substrate concentrations. Inosine formation was monitored after 1 hour of incubation with 20 nM wt ADAT or 100 nM ADAT*. (B, D) Deamination rate is plotted as a function of substrate concentration given in μM for wt ADAT (B) or ADAT* (D). Mean values and associated SD from three independent experiments are shown (blue: wt ADAT; orange: ADAT*). Curve fitting was performed following a nonlinear hyperbolic regression model ($R^2 = 0.7687$ for wt ADAT; $R^2 = 0.8347$ for ADAT*). (Figure panels A and C were provided by Dr Helena Roura Frigolé).

Table 4.1. Deamination kinetic parameters of wt and mutant ADAT for tRNA^{Ala}_{AGC}.

	wt ADAT	ADAT*	FC (wt/mut)
k_{cat} (min^{-1})	5.64	1.07	5.27
K_m (μM)	2.98	5.81	0.51
V_{max} (nM/min)	112.90	107.10	1.05
k_{cat}/K_m ($\mu\text{M}^{-1} \cdot \text{min}^{-1}$)	1.90	0.18	10.29

4.1.3. Subcellular localization of ADAT and impact of the V144M mutation

The location of human heterodimeric ADAT has not been directly determined so far. While in *T. brucei*, ADAT2/3 has been shown to be mostly cytoplasmic, with a small portion of the enzyme found in the nucleus (Gaston et al., 2007; Rubio et al., 2007), in plants it appears to be the other way around, with most of ADAT2/3 being located in the nucleus, and a minority appearing in the cytoplasm (Zhou et al., 2014). A slightly different situation occurs in yeast, where the homologs TAD2 and TAD3 have been found in both nucleus and cytoplasm (Huh et al., 2003). In human cells, the extremely low endogenous protein levels of both ADAT subunits hampered an initial analysis of their cellular distribution under native conditions in HEK293T cells. To overcome this problem, we identified a cell line (HBE1) with higher endogenous levels of both proteins and (**Figure 4.7 A**) used it to determine the subcellular localization of ADAT2 and ADAT3 by immunofluorescence. We found that endogenous human ADAT2 and ADAT3 localized predominantly in the nucleus and, to a much lesser extent, to the cytoplasm (**Figure 4.7 B, C**). This is in agreement with evidence indicating that precursor-tRNAs (pre-tRNAs) (i.e. tRNAs that have not yet undergone full 5' and 3' processing), which are thought to be processed mainly in the nucleus, are already modified with inosine at position 34 in their anticodon (Gogakos et al., 2017; Torres et al., 2015). Thus, human ADAT2/3 is primarily a nuclear enzyme.

We tried to generate a genetically edited cell line that encoded the ADAT3 V144M mutation using CRISPR/Cas9, but we failed to obtain viable cells in which the mutation was present in homozygosis (data not shown). In the absence of a cell line expressing endogenous levels of the mutated ADAT3 form, and considering that alterations in its native local structure could also impair detection by antibodies, we decided to develop an *in vitro* overexpression system to investigate the effect of the ADAT3 V144M mutation upon the enzyme's intracellular distribution. We generated in-frame fusions with either eGFP or mCherry chromophore at the N- or C-terminus of each of the three genes (ADAT2, wt ADAT3, and ADAT3 V144M). By doing so, we could rule out potential localization artifacts due to the presence of a particular type of chromophore or to the location of the tag. Live cell imaging confocal microscopy of transfected HEK293T cells with the wild type constructs showed that eGFP- and mCherry-tagged ADAT2 localized mainly to the nucleus and partially to the cytoplasm (**Figure 4.8**; upper panel), consistent with the results obtained by immunofluorescence. Unexpectedly, either eGFP- or mCherry-tagged wt ADAT3 localized mainly to the cytoplasm and partially to the nucleus (**Figure 4.8**; central panel). A possible explanation for this behavior is that, due to the heterodimeric nature of the enzyme, the

localization of its subunits is likely to be dependent on their stoichiometry. Therefore, we co-expressed eGFP-tagged ADAT2 together with mCherry-tagged ADAT3 and found that, only when both subunits were co-expressed (i.e. cells that had been successfully transfected with the two plasmids), they co-localized in the nucleus (**Figure 4.8**; lower panel), which suggests that ADAT3 translocates to nucleus in an ADAT-dependent manner.

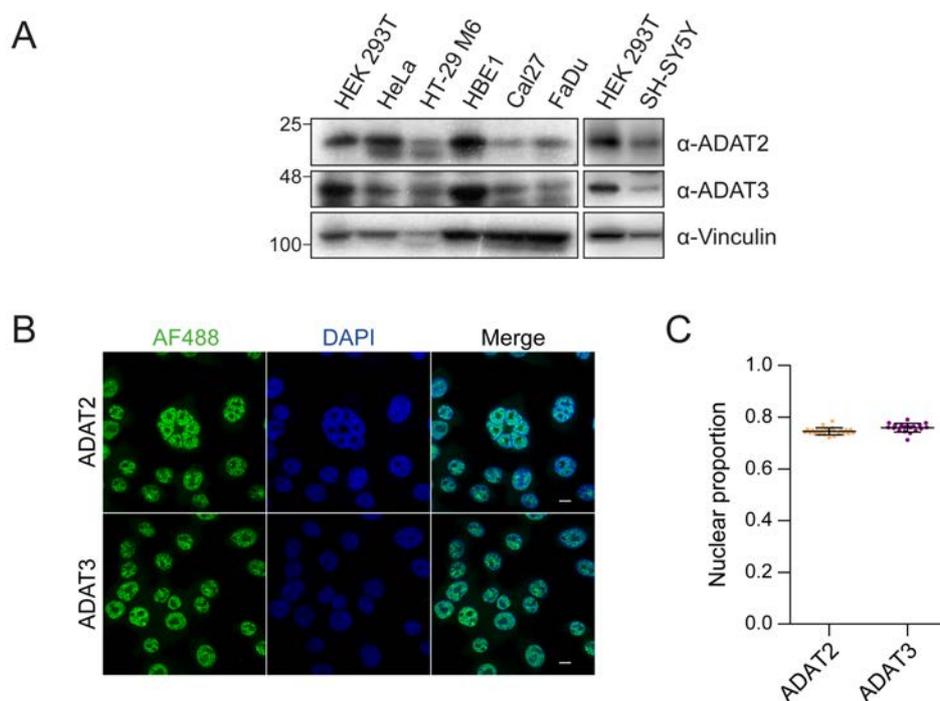


Figure 4.7. Subcellular localization of endogenous ADAT2 and ADAT3. (A) Western blot analysis of ADAT2 and ADAT3 protein levels in different cell lines. 3.6×10^6 cells were harvested per condition and lysed in 100 μ L of RIPA buffer at 4 $^{\circ}$ C. 15 μ L of the obtained protein extract of each cell line was loaded in 12% SDS-PAGE and blotted with specific antibodies; vinculin was used as internal control. (B) Immunofluorescence microscopy of HBE1 cells, staining for endogenous ADAT2 and ADAT3 proteins (green). DAPI was used to identify the nucleus. Scale bar corresponds to 10 μ m. (C) Quantification of the proportional nuclear distribution of endogenous ADAT2 and ADAT3 shown in (B). A minimum of 25 different microscopy images were quantified per condition. Shown are the mean and associated SD.

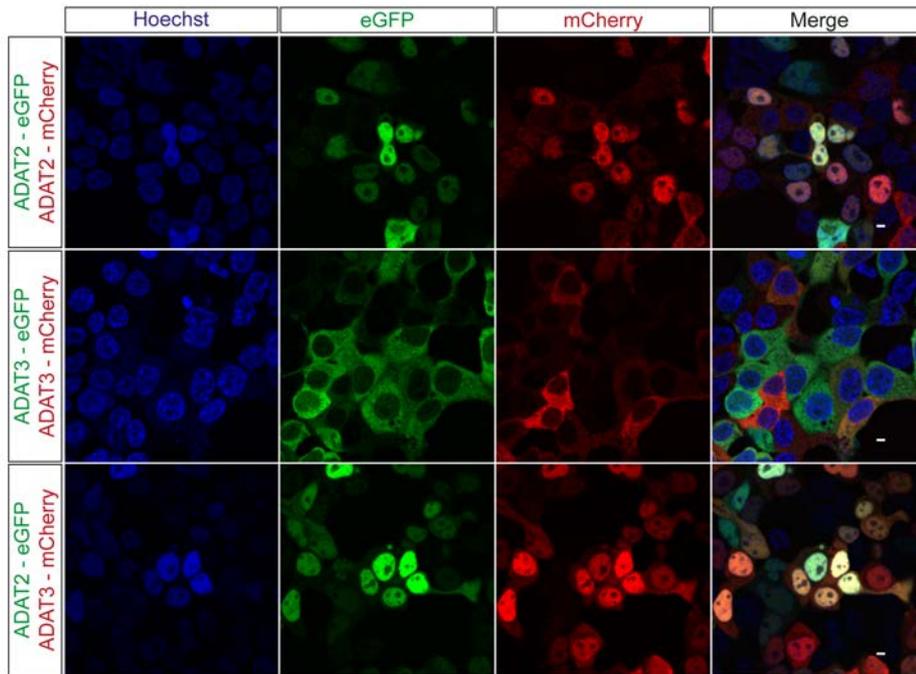


Figure 4.8. Subcellular localization of co-expressed wild type ADAT2 and ADAT3. Live cell imaging confocal microscopy of HEK293T cells 48 h after co-transfection of constructs encoding eGFP- or mCherry-tagged human ADAT2 and ADAT3 proteins. Scale bar corresponds to 10 μ m. Shown are the results for C-ter tagged ADAT proteins. Adapted from (Torres et al., 2015).

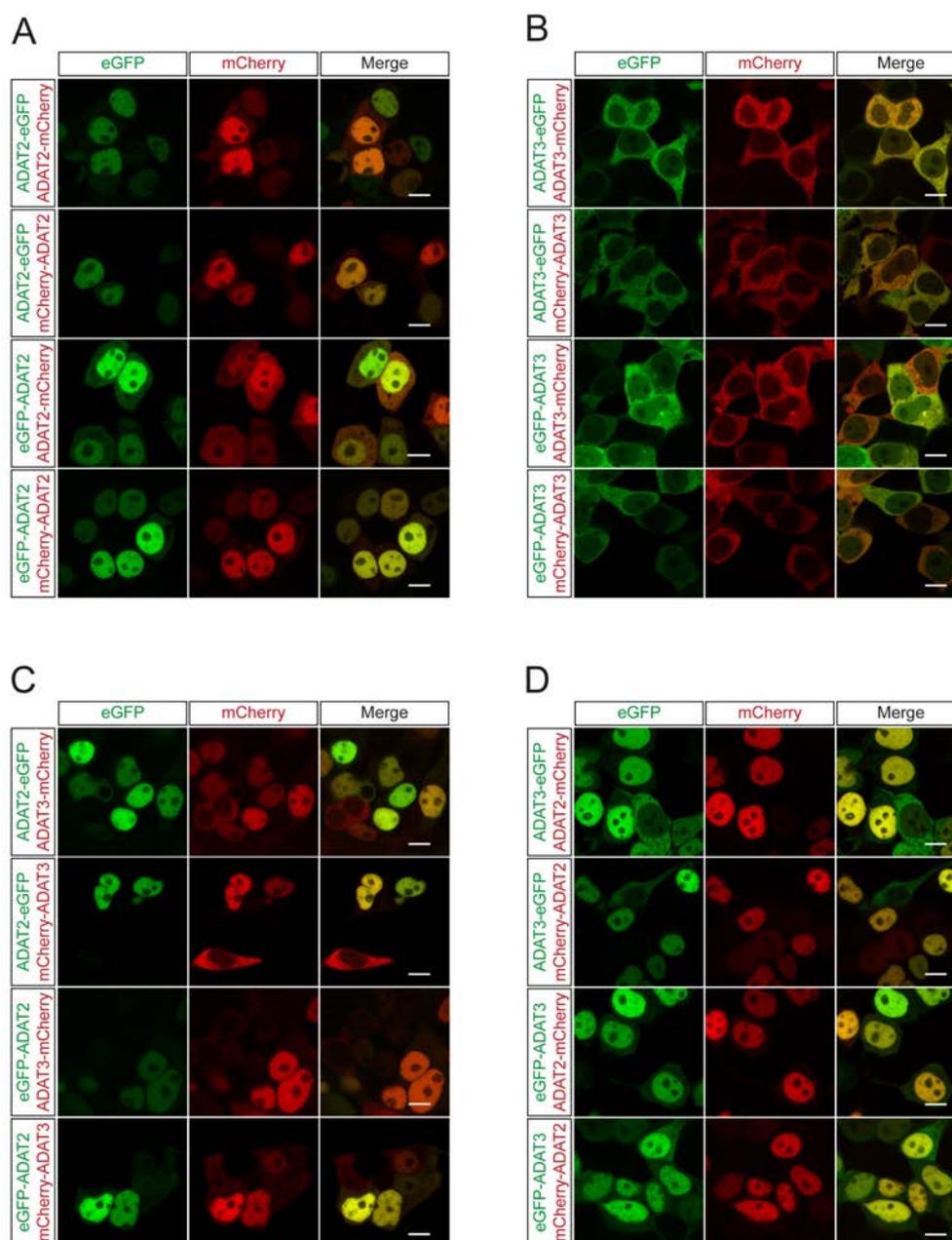


Figure 4.9. Subcellular localization of wild type ADAT2 and ADAT3 using different fluorescent tags inserted at either end of the proteins. Live cell imaging confocal microscopy of HEK293T cells 48 hours after co-transfection of plasmids expressing C- or N-terminal eGFP- or mCherry-tagged ADAT2 (**A**), ADAT3 (**B**), or combinations of both subunits (**C**, **D**). The localizations of ADAT2 and ADAT3 are not dependent on the fluorescent tag used nor on their insertion position. Note that, in panels C and D, cells that have only received the plasmid encoding ADAT3, but not the one carrying ADAT2, still show ADAT3 present mainly in the cytoplasm. Scale bar corresponds to 10 μ m.

These results agree with those obtained for endogenous ADAT2 and ADAT3, are further supported by the fact that equivalent outcomes were obtained independently of the tag used or its location within the protein (**Figure 4.9**), and validate this system for the study of how the V144M mutation might affect the enzyme's subcellular localization. We included the wt

ADAT3 construct as a control in each experiment conducted. As opposed to what we had previously observed for wt ADAT3, eGFP-ADAT* was detected uniformly in both the nucleus and the cytosol (**Figure 4.10 A, B**). However, when co-expressed with mCherry-tagged ADAT2, both wt and mutant ADAT3 equally translocated to nucleus (**Figure 4.10 A, C**). The same behavior was observed using the neuroblastoma cell line SH-SY5Y (**Figure 4.11**).

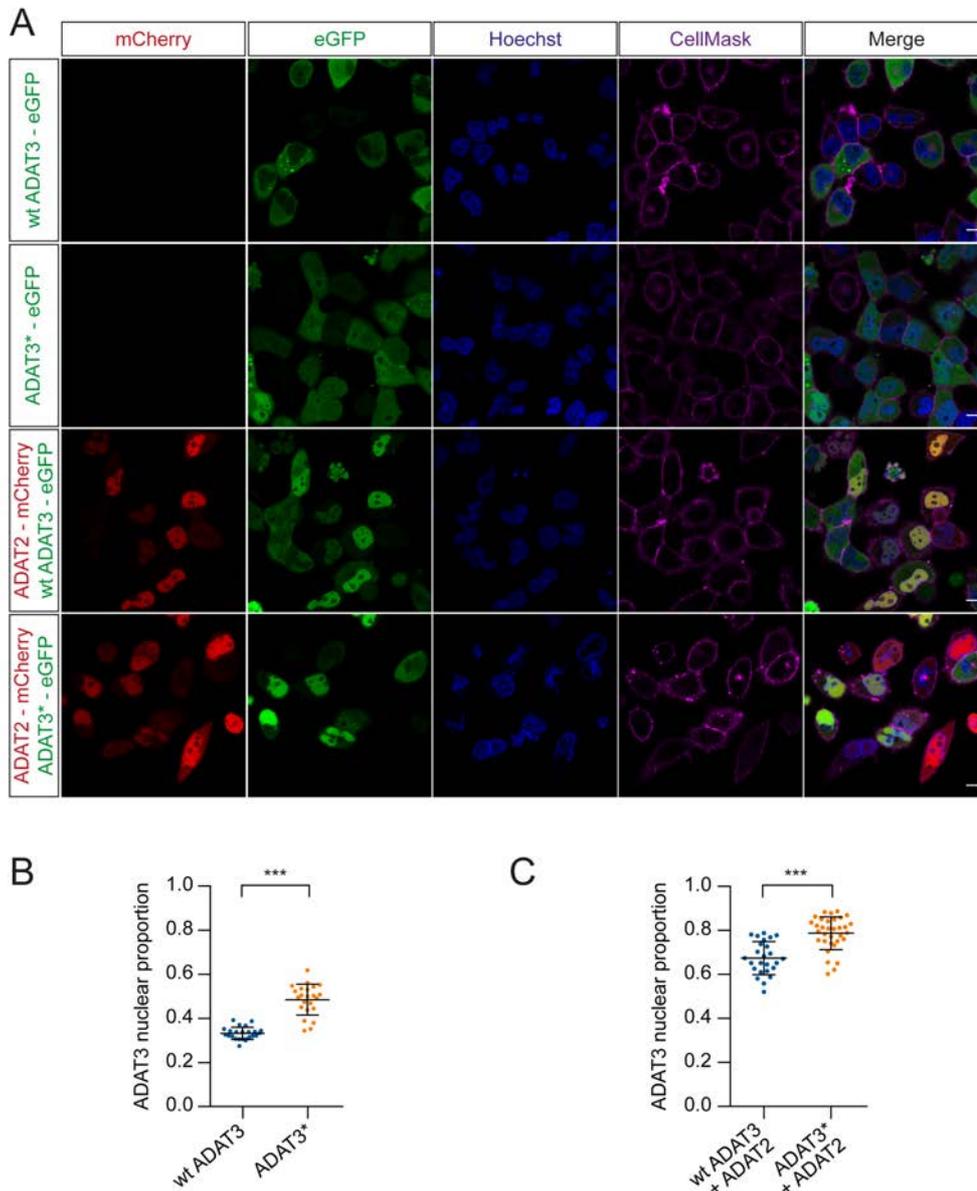


Figure 4.10. Impact of the V144M mutation upon the subcellular localization of ADAT3. (A) Live cell imaging confocal microscopy of HEK293T cells 48 hours post-transfection with plasmids expressing eGFP-tagged wt ADAT3 or eGFP-tagged ADAT3*; with or without co-transfection of mCherry-tagged ADAT2. Scale bar corresponds to 10 μ m. Hoechst staining was used to identify the nucleus and CellMask was used to identify the plasma membrane. (B, C) Proportional nuclear distribution of wt ADAT3 and ADAT3* when overexpressed alone (B) or upon co-transfection with ADAT2 (C). A minimum of 23 different images as the ones shown in (A) were quantified per condition. Statistical analyses were performed as described in Experimental procedures (see also Table S2). Shown are the mean and associated SD. *** $P \leq 0.001$.

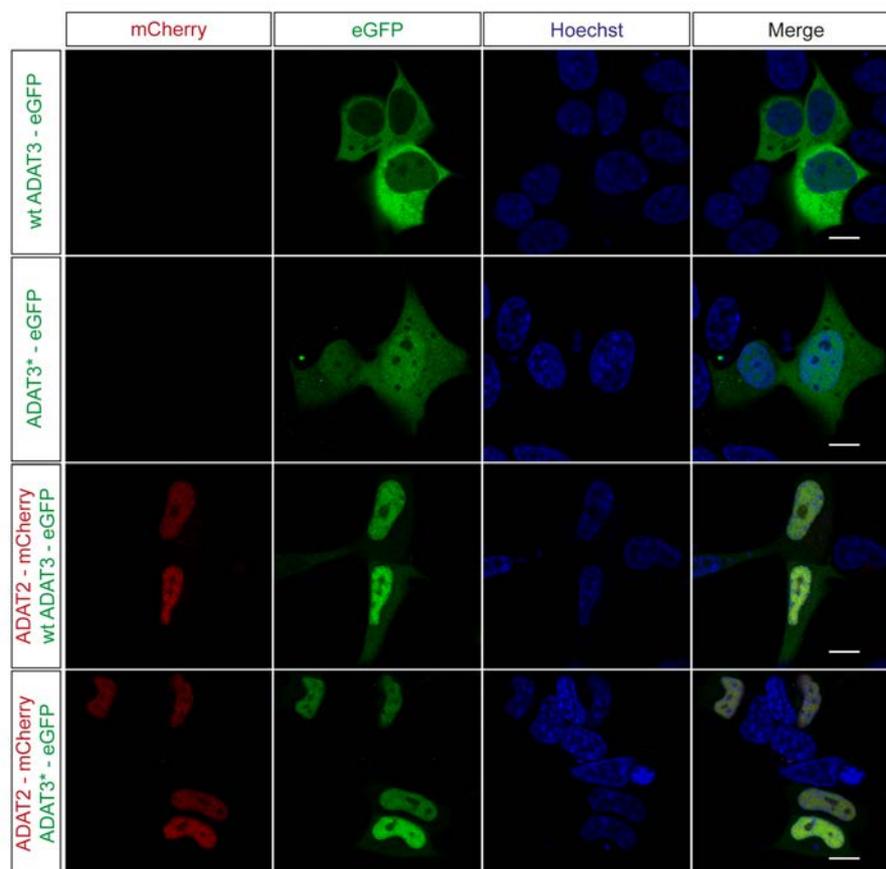


Figure 4.11. Impact of the V144M mutation upon subcellular localization of ADAT3 expressed in neuroblastoma cells. Live cell imaging confocal microscopy of SH-SY5Y cells 48 hours post-transfection with plasmids expressing eGFP-tagged wt ADAT3 or eGFP-tagged ADAT3*; with or without co-transfection of mCherry-tagged ADAT2. Scale bar corresponds to 10 μ m. Hoechst staining was used to label the nucleus.

Considering that ADAT2 appears to be similarly required for the translocation of either wt ADAT3 or ADAT3* to nucleus, it is likely that the ADAT2-ADAT3* interaction within the cell is not altered. We ruled out that the observed effects of the V144M mutation in the localization of ADAT3 were due to the tag used or its position in the protein by expressing mCherry- or eGFP-tagged ADAT3* at either C- or N-terminus (**Figure 4.12**). Additionally, because fluorescent tags present in fusion proteins can be partially cleaved in cells (Huang et al., 2014; Shaner et al., 2005), we verified by Western blot that eGFP-tagged ADAT3 proteins were expressed to similar extents and that the degree of partial cleavage of the tag was identical for all fusion proteins (**Figure 4.13 A**).

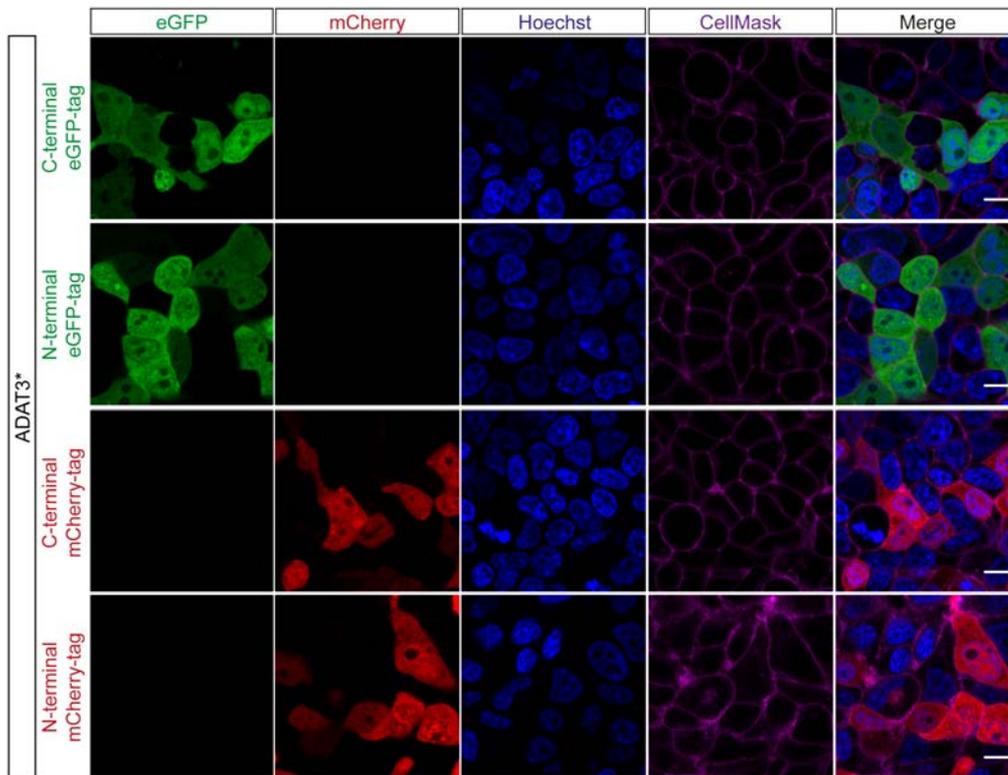


Figure 4.12. Subcellular localization of ADAT3* using different fluorescent tags inserted at either end of ADAT3*. Live cell imaging confocal microscopy of HEK293T cells 48 hours post-transfection with plasmids expressing ADAT3* tagged with eGFP or mCherry at its C- or N-terminus. Scale bar corresponds to 10 μ m. Hoechst staining was used to identify the cell nucleus and CellMask to label the plasma membrane. The abnormal ubiquitous localization of ADAT3* is not dependent on the fluorescent tag used nor on its insertion position.

A tendency for wt ADAT3 to aggregate has been described (Delker et al., 2013). The inherent instability of ADAT3 when not in complex with its partner ADAT2 may explain why the protein levels of the overexpressed wt ADAT3 and ADAT3* detected by Western blot were lower than those of ADAT2 (**Figure 4.13 A**). In agreement with this report, we also noted that cells overexpressing wt ADAT3 occasionally presented aggregation foci, which were larger and more abundant in cells expressing ADAT3* (**Figure 4.13 B**). While these observations suggest an increased aggregation propensity of ADAT3*, its physiological relevance is unclear given the low proportion of cells that displayed this phenotype.

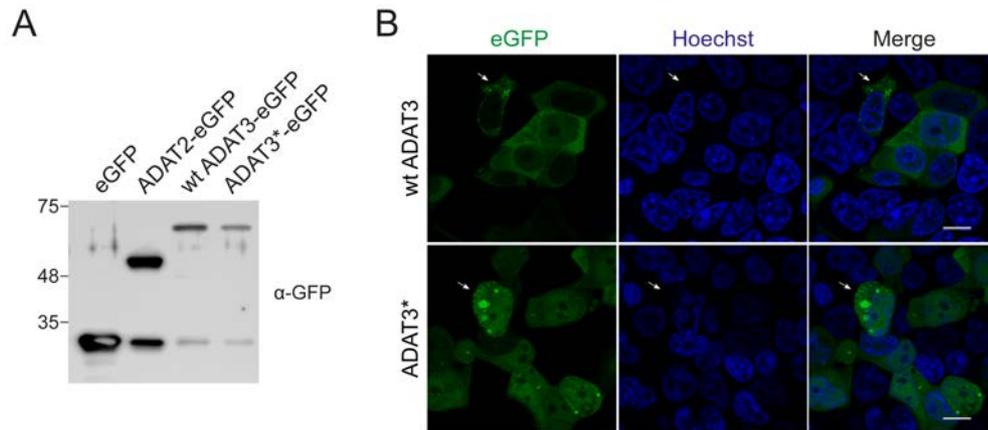


Figure 4.13. Expression levels of eGFP fusion proteins and insoluble nature of ADAT3. (A) Western blot analysis of extracts from HEK293T cells transfected with plasmids expressing eGFP (28.3 kDa), eGFP-tagged ADAT2 (49.3 kDa), eGFP-tagged wt ADAT3 (68 kDa), or eGFP-tagged ADAT3 V144M (68 kDa). Note that comparable expression levels are observed for both ADAT3 constructs, and that presence of free eGFP is observed for both ADAT3 constructs to similar extents. (B) Representative images of cells expressing eGFP-tagged wt ADAT3 or ADAT3* that present foci of these proteins (white arrows). Scale bar corresponds to 10 μ m. Hoechst staining was used to identify the nucleus.

4.2. ROLE OF ADAT IN THE TRANSLATION OF ECM PROTEINS

4.2.1. ADAT2 silencing impairs cell proliferation

In order to investigate the role of human ADAT in translation, we generated several cell lines stably expressing either a short-hairpin RNA (shRNA) against the *ADAT2* mRNA or an shRNA with a non-target sequence (negative control). Thus, we obtained four cell lines, including the previously described HEK293T cell line generated by CRISPR/Cas9 technology, in which the knockdown of ADAT2 was confirmed by Western blot (**Figure 4.14 A**) and RT-qPCR (**Figure 4.14 B**). A significant reduction of ADAT2 both at the protein and at the mRNA levels was observed for all cells with the sole exception of HEK293T ADAT2 KD cells, whose *ADAT2* mRNA levels were expectedly unchanged as the mutation introduced generates a truncated protein form without altering mRNA stability (data not shown). Of note, mRNA levels of *ADAT3* remained the same in all cells.

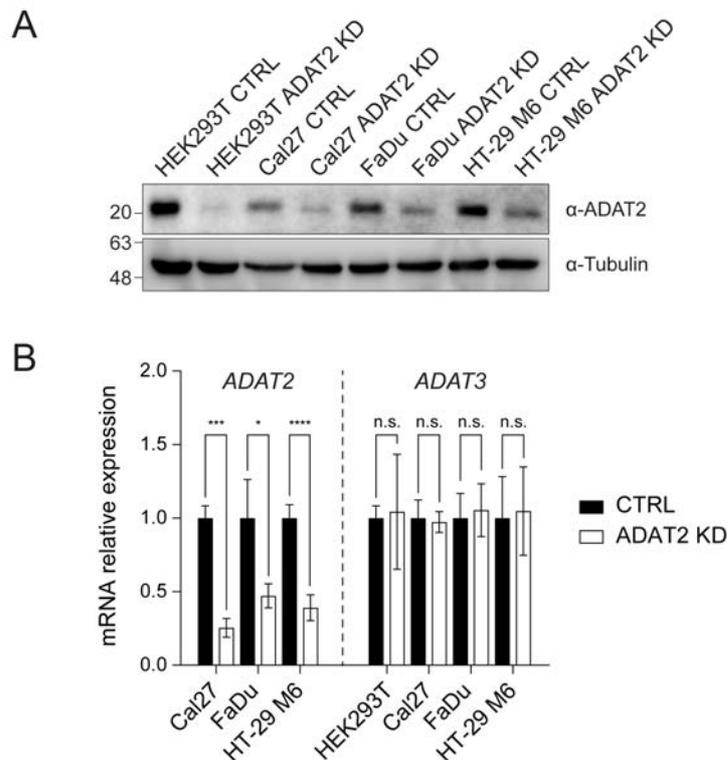


Figure 4.14. ADAT2 depletion in several cell lines. (A) Western blot showing reduced protein levels of ADAT2 in silenced cells. (B) mRNA levels for *ADAT2* decreased in silenced cells, but transcript levels for *ADAT3* remain the same. Statistical significance was determined using a t-test. * $P < 0.05$; *** $P < 0.001$; **** $P < 0.0001$; n.s., non-significant.

A decrease in the extent of the I34 tRNA modification carried out by ADAT has been previously reported for the same shRNA-mediated knockdown of ADAT2 in HEK293T cells (Torres et al., 2015). It is therefore reasonable to assume that the knockdown of ADAT2 in these newly generated cell lines also leads to a reduction in I34 levels. This assumption was confirmed for the following results.

Based on initial observations showing that silencing of ADAT2 leads to a global delay in cell cycle progression (**Figure 1.15 C**), we hypothesized that it could also result in decreased cellular proliferation. We assessed this idea by measuring cumulative cellular growth for CTRL and ADAT2 KD cells (**Figure 4.15**). Significant reductions in cell growth rates were displayed by all ADAT2 KD cell lines relative to their control counterpart, suggesting that ADAT is required for proper cellular proliferation.

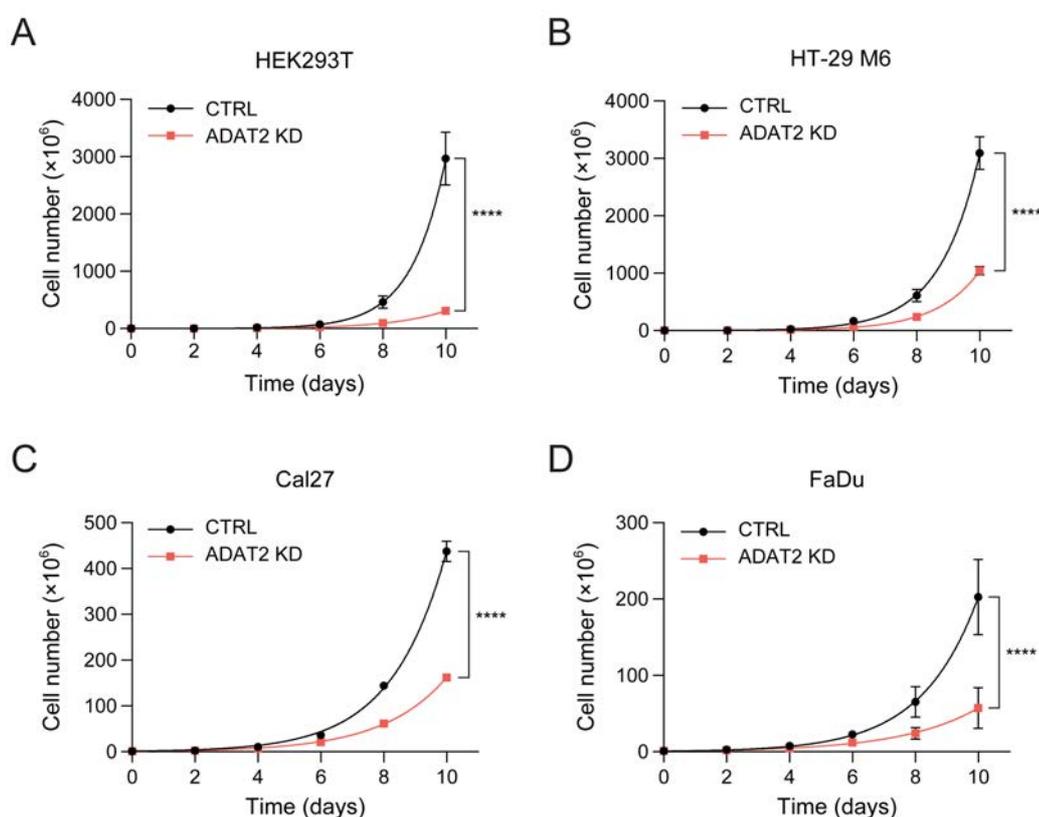


Figure 4.15. Effect of ADAT2 silencing on cellular growth. Cumulative cell growth curves for CTRL and ADAT2 KD in HEK293T (**A**), HT-29 M6 (**B**), Cal27 (**C**), and FaDu (**D**) cells. Shown are the average and associated SD at the indicated time points for three independent replicates. Curves were obtained by least squares fit of the raw data. Statistical significance was determined using a F-test. **** $P < 0.0001$. (Figure panel A was provided by Dr Adrián G. Torres).

4.2.2. Depletion of ADAT2 does not activate the unfolded protein response

Several studies have reported an association between the loss of tRNA modifications at the anticodon loop and the activation of ER-stress and the unfolded protein response (UPR), presumably due to defects in protein translation that result in misfolded or unfolded proteins (Pereira et al., 2018). Genes containing ADAT-stretches are enriched in ECM components, which implies that most of their encoded proteins must be synthesized in the ER in order to be secreted to the plasma membrane. We wondered whether the silencing of ADAT2 could also cause misfolding of these proteins leading to UPR activation.

In mammalian cells, ER-stress is transduced via the activation of three ER transmembrane proteins, namely IRE1, PERK, and ATF6, which induce UPR signaling (Karagöz et al., 2019). Depletion of the I34 modification did not lead to the activation of PERK via phosphorylation, which results in eIF2 α phosphorylation and upregulation of ATF4 (**Figure 4.16 A**). Neither did it cause splicing of *XBP1* mRNA that occurs downstream of IRE1 stimulation by the Hsp70-type chaperone BiP, whose levels were also unaltered (**Figure 4.16 B**). Importantly, UPR was not impaired in ADAT2 KD cells as activation using the ER stressor thapsigargin lead to increased levels of phospho-PERK, BiP, ATF4 and splicing of *XBP1*. Additionally, UPR activation in ADAT2 KD cells was comparable to that in CTRL cells, suggesting that silenced cells were not hypersensitive to ER stress and that their ability to mount stress responses was not impaired by the deficiency of I34.

Despite a lack of IRE1- or PERK-driven UPR signaling in ADAT2 KD cells, we could not rule out at this point that the depletion of the enzyme did not cause misfolding of newly synthesized proteins. On the one hand, activation of ATF6 in these cells is yet to be investigated. And on the other hand, previous reports have shown that loss of tRNA wobble U modifications can lead to protein aggregation without activation of UPR (Nedialkova and Leidel, 2015).

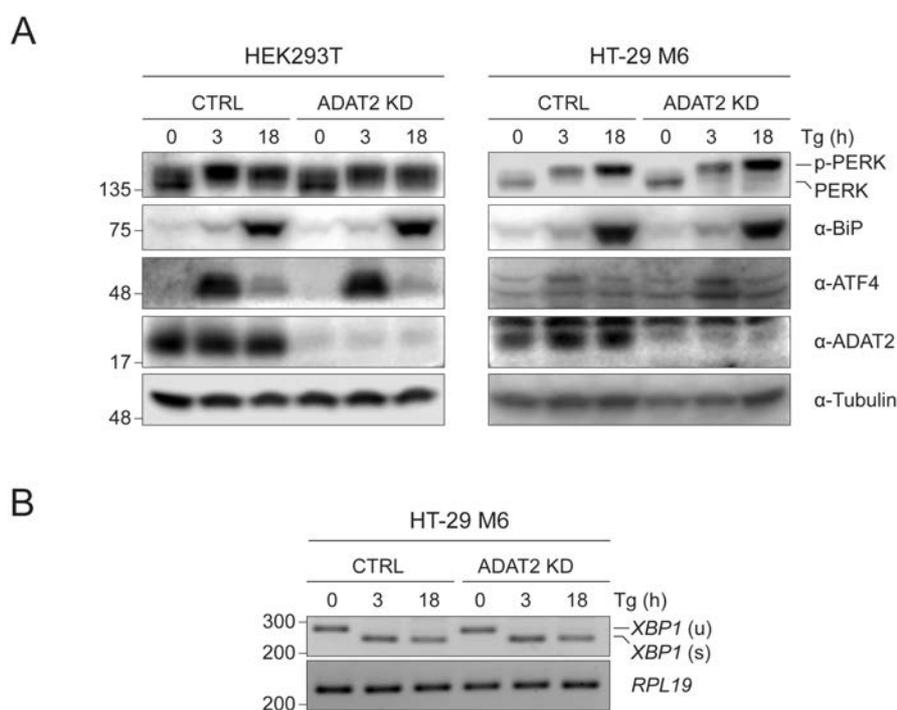


Figure 4.16. UPR activation in CTRL and ADAT2 KD HEK293T and HT-29 M6 cells. (A) Western blot detection of the activation of UPR-related proteins in CTRL and ADAT2 KD for HEK293T and HT-29 M6 cells. 2 μ M thapsigargin was used as an activator of UPR. **(B)** RT-PCR to examine *XBP1* mRNA splicing in the same HT-29 M6 cells as in (A). *XBP1* unspliced (u) and spliced (s) forms are indicated. Numbers correspond to bp.

4.2.3. Assessing the impact of ADAT2 KD in translation by polysome profiling

If the translation of genetic regions enriched in ADAT-stretches is indeed dependent on the pool of I34-modified tRNAs, one would expect that the number of ribosomes associated to these gene transcripts might be altered when the size of the available pool is decreased. We addressed this possibility by analyzing the distribution of transcripts with or without ADAT-stretches on polysome sucrose gradients. The colorectal cancer HT-29 M6 cell line spontaneously differentiates at culture post-confluence to acquire an epithelial mucus-secreting phenotype that is reminiscent of goblet cells. As mucins were identified to be highly enriched in ADAT-stretches, we used cellular extracts from control and ADAT2 silenced HT-29 M6 cells, before and after post-confluency (i.e. undifferentiated and differentiated cells) (**Figure 4.17**). Chromatograms for pre-confluent cultures showed the standard polysome profile consisting in two relatively small peaks for the 40S and 60S ribosome subunits, followed by the 80S ribosome peak, and a series of peaks with increasing A254 intensity that correspond to actively translating polyribosomes (**Figure 4.17 A**). Because over confluent

cells are known to reduce their translational activity, it was not surprising to observe an enlargement of the 80S peak and decreased polyribosome peaks when the cells were grown to post-confluency (**Figure 4.17 B**). However, no differences were noted in the profiles of control and ADAT2 KD cells, neither at pre-confluence nor at post-confluence, which is in agreement with our previous observations that general translation in ADAT2 silenced cells is unaffected.

The distribution of mRNAs with or without ADAT-stretches along sucrose gradients was analyzed by RT-qPCR (**Figure 4.17 C, D**). *COX19* and *METTL7A* were used as negative controls because they encode very few ADAT-sensitive codons and, therefore, are not expected to be differentially translated by ADAT2 KD cells as compared to control cells. A marked decrease in transcript abundance in heavy polysomes found at late fractions was observed for most mRNAs upon culture in post-confluency. This reduction was accompanied by a higher presence of those transcripts in earlier fractions, and it is indicative of decreased translation. The fact that this shift was not as obvious for *MUC1* and *MUC5AC* mRNAs correlates well with their sustained and increased expression upon cellular differentiation. However, we found no reproducible differences in the profiles of mRNAs containing ADAT-stretches between ADAT2 KD and control cells, neither under pre-confluent nor post-confluent conditions (**Figure 4.17 C, D**). A caveat of this assay is that some of the transcripts found to be associated with polyribosomes may not be undergoing active translation. Previous reports have shown situations where significant translation inhibition of a given protein is not correlated with a decrease in the number of ribosomes associated to its encoding mRNA (Braat et al., 2004; Darnell et al., 2011; Nottrott et al., 2006; Petersen et al., 2006). Therefore, the signal to noise ratio may prevent detection of differential load of actively translating ribosomes onto mRNAs between control and ADAT2 KD cells.

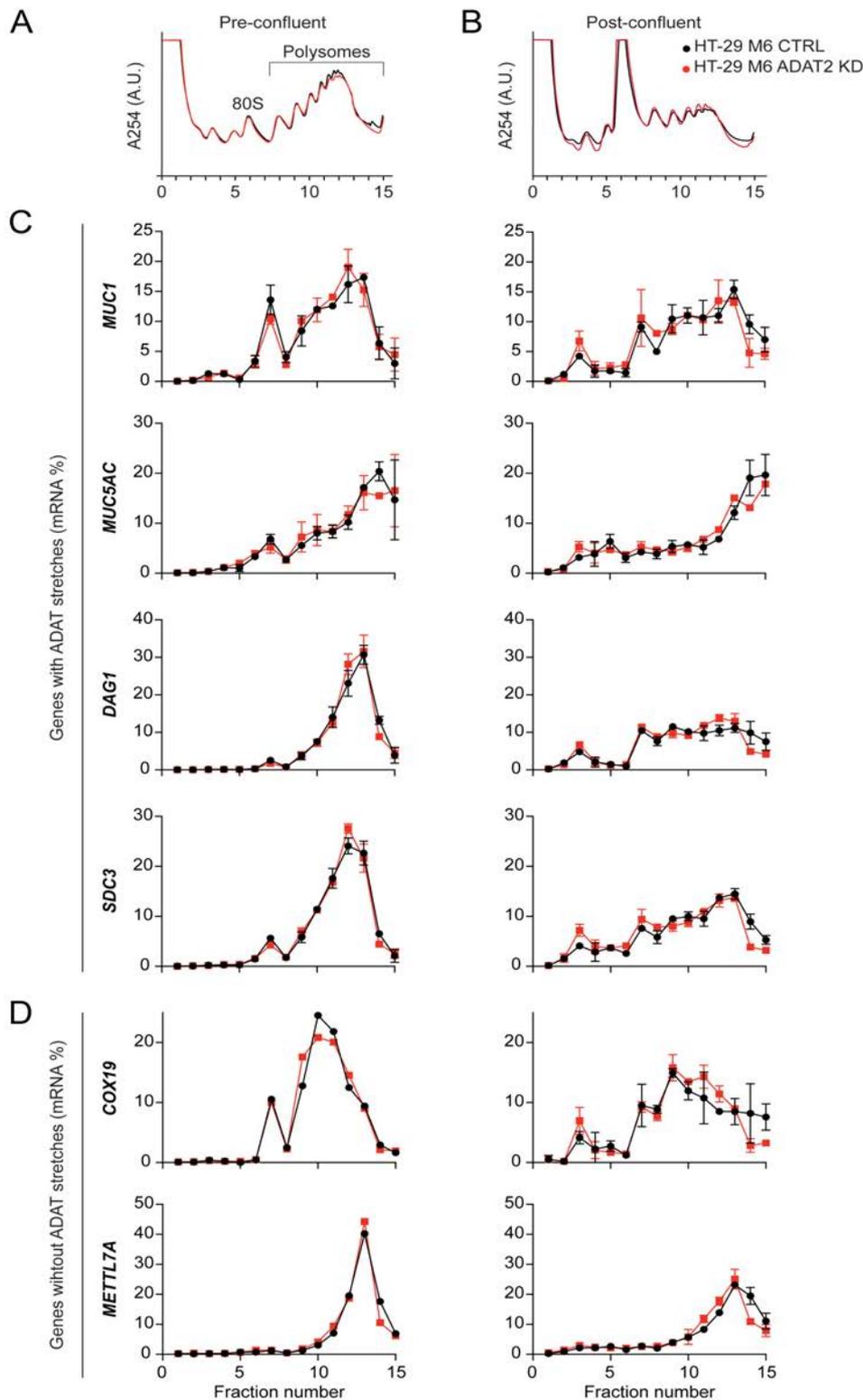


Figure 4.17. Polysome profiles and associated mRNAs in pre- and post-confluent cultured HT-29 M6 cells. (A–B) Polysome profiles of control and ADAT2 depleted HT-29 M6 cells at pre-confluency (A) and at post-confluency (B). The 80S and polysome peaks are indicated. Control cells are shown in black, and ADAT2 KD cells, in red. (C–D) Distribution of transcripts with (A) or without (B) ADAT-stretches across the 15 sucrose gradient fractions as quantified by RT-qPCR. Pre-confluent cells are shown on the left, and post-confluent cells are shown on the right. CTRL and ADAT2 KD samples are color coded as in (A–B). Data are plotted as the percentage of total mRNA on the gradient; error bars represent standard deviation from two biological replicates, except for pre-confluent *COX19* and *METTL7A*, where a single sample was used.

In order to overcome that possibility, we used the translation inhibitor puromycin to perform ribosome run-off assays and then assess the distribution of mRNAs along polyribosome gradients as described above. Puromycin is a non-functional analog of aminoacyl-tRNA that, when incorporated into the polypeptide nascent chain during elongation, leads to premature termination of translation via release of ribosomal subunits off the mRNA. Therefore, treatment of cells with puromycin would result in dissociation of actively translocating ribosomes and mRNAs, while ribosomes that were stalled or elongating at reduced rates should be less sensitive to the drug and remain loaded on transcripts.

We carried out ribosome run-off assays in HEK293T cells because they are not resistant to puromycin, as opposed to all our cell lines that stably express and shRNA and use the *N*-acetyl-transferase as a selection marker. Polysome profiles in HEK293T untreated cells were similar to those obtained for HT-29 M6 cells, although in this case a small decrease in A254 intensity could be detected for ADAT2 KD cells at heavy polysomes (**Figure 4.18 A**). Ribosomal run-off was obvious upon incubation of cells with puromycin, with an increase in the peak corresponding to the 80S monosome and a decrease in polysome peaks (**Figure 4.18 B**). *MUC1* and *MUC5AC* transcripts were excluded from this analysis as they are not expressed in HEK293T cells, which was confirmed by RT-qPCR (data not shown). Instead, we tested the distribution of *SDC3* and *DAG1*, as well as the previously used negative controls *COX19* and *METTL7A* (**Figure 4.18 C, D**). While a clear shift from heavy to light polysomes could be observed for *DAG1*, *COX19*, and *METTL7A*, this was less certain for *SDC3*. Unfortunately, the high variability present within biological replicates prevented us from observing differences between the relative amounts of mRNAs associated to ribosomes in control and ADAT2 KD cells.

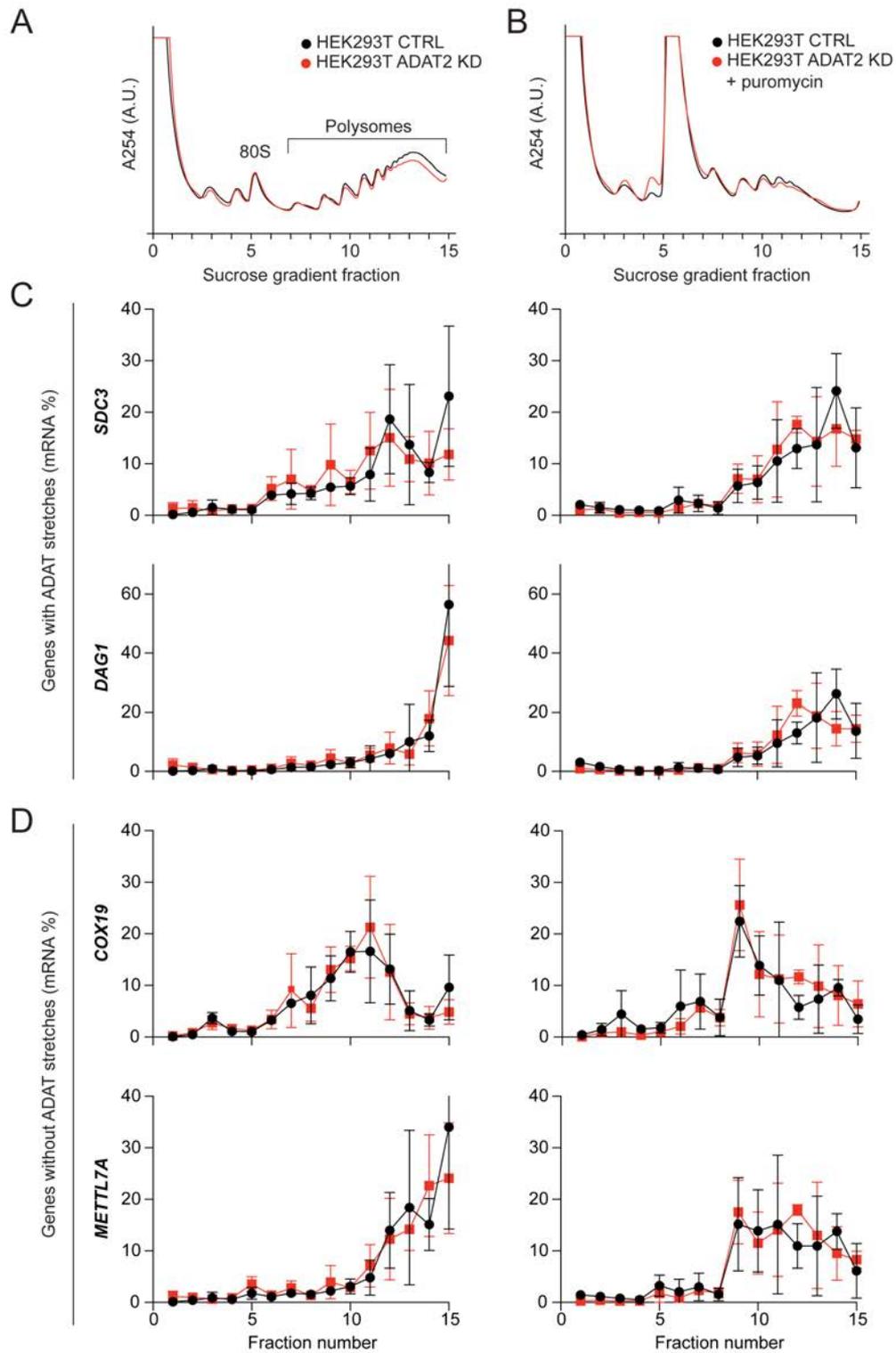


Figure 4.18. Polysome profiles and associated mRNAs in HEK293T under basal conditions and puromycin run-off assays. (A–B) Polysome profiles of control and ADAT2 KD HEK293T cells in steady state (A) and upon treatment with puromycin (B). The 80S and polysome peaks are indicated. Control cells are shown in black, and ADAT2 KD cells, in red (C–D) Distribution of transcripts with (C) or without (D) ADAT-stretches across the 15 sucrose gradient fractions as quantified by RT-qPCR. Untreated cells are shown on the left, and puromycin treated cells are shown on the right. CTRL and ADAT2 KD samples are color coded as in (A–B). Data are plotted as the percentage of total mRNA on the gradient; error bars represent standard deviation from two biological replicates.

4.2.4. Ribosome profiling does not show decoding defects in ADAT2 KD cells

Polysome analysis was unable to determine whether depletion of the I34 modification lead to elongation defects. Therefore, we addressed this question by examining the transcriptome-wide landscape of ribosome occupancy in control and ADAT2 KD cells using ribosome profiling (Ingolia et al., 2009). By sequencing ribosome-protected mRNA fragments (RPFs), or ribosome footprints, ribosome profiling provides quantitative information about the identity of the codons located in each of the ribosomal sites (E-, P-, and A-site) along endogenous transcripts (Ingolia et al., 2009; Stadler and Fire, 2011). Global analysis of relative ribosome occupancy yields thousands of instances of decoding for each codon, thus allowing the identification of codon-specific effects on the rate of translation elongation.

We investigated the consequences of the knockdown of ADAT2 in translation using four different cells lines. In addition to the above-mentioned HEK293T and HT-29 M6 cells, we performed ribosome profiling on two squamous cell carcinoma cell lines in which we had silenced ADAT2. As opposed to HT-29 M6, these two cell lines do not express high amounts of mucins. Instead, they synthesize several components of the keratin protein family which do not contain ADAT-stretches, but exhibit a relatively high abundance of ADAT-sensitive codons.

Evaluation of the sequencing data quality revealed a strong correlation of per-gene normalized read count values between most biological replicates (**Figure 4.19 A-D**). However, one of the HEK293T CTRL mRNA replicates showed an elevated presence of outliers (**Figure 4.19 A**), which would negatively affect downstream evaluation of differential gene expression. Therefore, we excluded HEK293T mRNA samples from our analyses. We confirmed a reduction of at least 40% of the transcript levels of ADAT2 in the mRNA data from HT-29 M6, Cal27, and FaDu cells (**Figure 4.20 A**). All RPF libraries showed good replicability, as well as the expected three-nucleotide periodicity that is characteristic of ribosome footprints (**Figure 4.20 B**). Most RPF reads were found to have a 5'-end nucleotide positioned in the 0 and 2 frames (**Figure 4.20 C**), which is likely due to inefficient RNA digestion as the majority of the reads should be located in frame 0. However, this was not considered to be problematic, since previous studies have successfully used similar data to discern codon effects during translation (Nedialkova and Leidel, 2015; Wang et al., 2018).

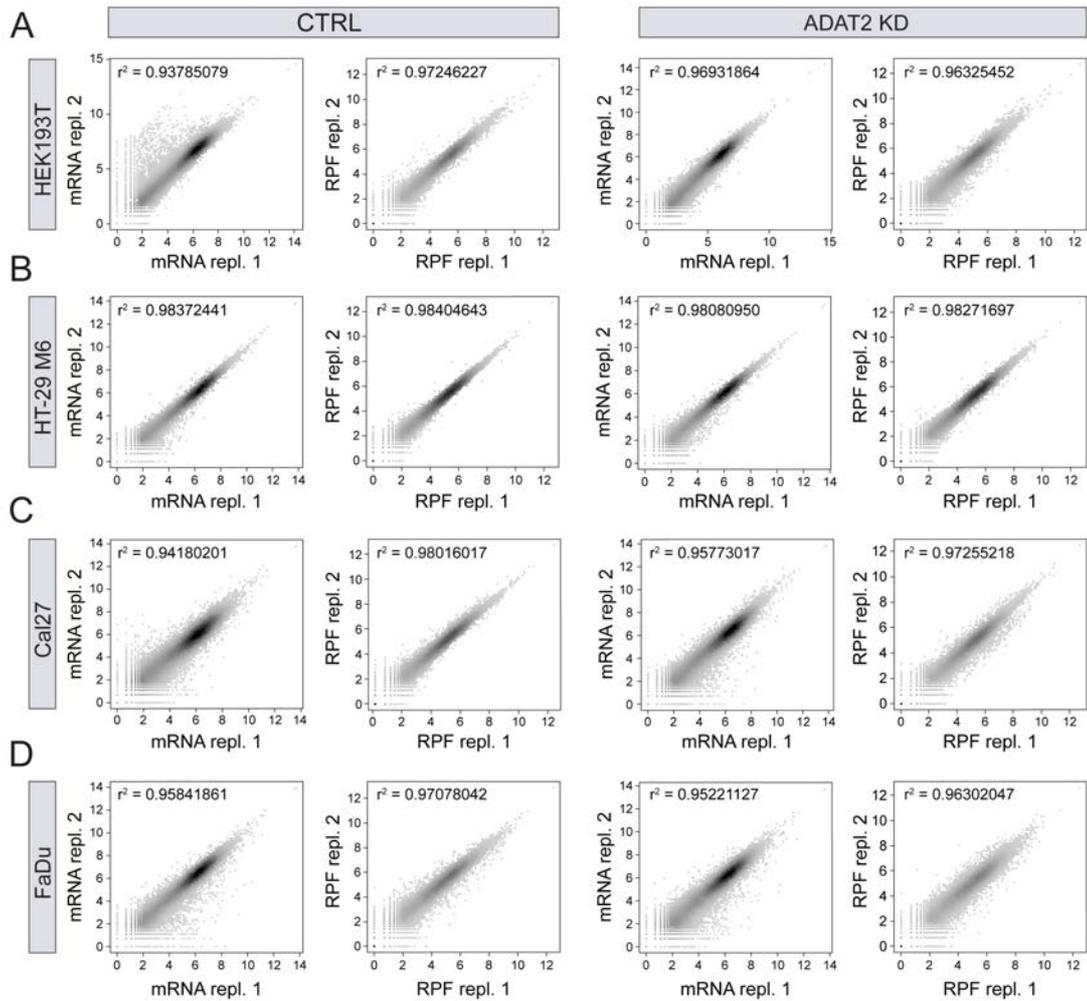


Figure 4.19. Replicability of mRNA and ribosome footprint data. (A-D) Inter-replicate correlation of log-transformed normalized read counts for control and ADAT2 KD in HEK293T (A), HT-29 M6 (B), Cal27 (C), and FaDu (D) cells. (Data was obtained in collaboration with Dr Eva M. Novoa).

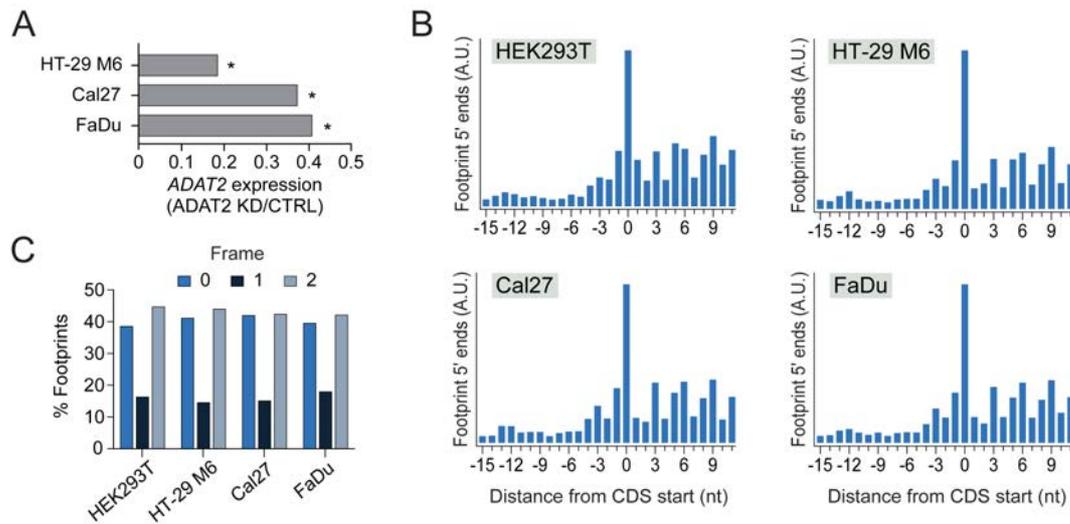


Figure 4.20. Quality control of ribosome profiling data. (A) Fold change transcriptional (RNA-seq) expression of *ADAT2* in *ADAT2* KD cells versus control cells. * $P < 0.05$. (B) Aggregated coverage of the 5'-end nucleotide (nt) from ribosome footprints mapped around the start codon across all transcripts in HEK293T, HT-29 M6, Cal27, and FaDu cells. Shown are reads of lengths comprised between 29 and 32 nt mapped without mismatches. A three-base periodicity is observed within the CDS. The small peak located 12-13 nt upstream of the start site is inferred to correspond to ribosomes positioned for translation initiation with the mRNA AUG codon located in their P-site. (C) Proportion of footprint reads with a 5' located in each frame throughout CDSs.

To test whether the silencing of *ADAT2* lead to variations in translational efficiency, we used the Xtail R-package (Xiao et al., 2016), which identifies genes whose ribosome occupancy cannot be explained by expression changes in the corresponding mRNA. However, this analysis yielded no statistically significant results. It is possible that our read coverage was not sufficiently high to detect small changes in translation efficiency. Alternatively, strong ribosome pauses occurred during translation might lead to simultaneous accumulation of ribosomes on a given mRNA and ribosome drop-off events, which together could result in an unaltered number of RPF read counts due to counterbalance between these two factors. In order to examine the possibility of ribosome stalling upon *ADAT2* silencing, we compared codon-specific ribosome occupancies in *ADAT2* KD and control cells. Following the approach depicted in **Figure 4.21 A**, we analyzed the normalized frequency of all 61 sense codons at each footprint site. The three non-tRNA binding positions downstream of the A-site (+1 to +3) are yet to be translated, therefore their codon identity should have negligible effects on ribosome occupancy (Nedialkova and Leidel, 2015; Stadler and Fire, 2011). Instead, they are rather a reflection of the mRNA codon composition within each sample, and thus can correct for differences in mRNA expression. Using this metric, we were able to detect an enrichment of stop codons at the A-site and downstream positions (**Figure 4.21 B**), suggesting that ribosomal sites had been correctly assigned.

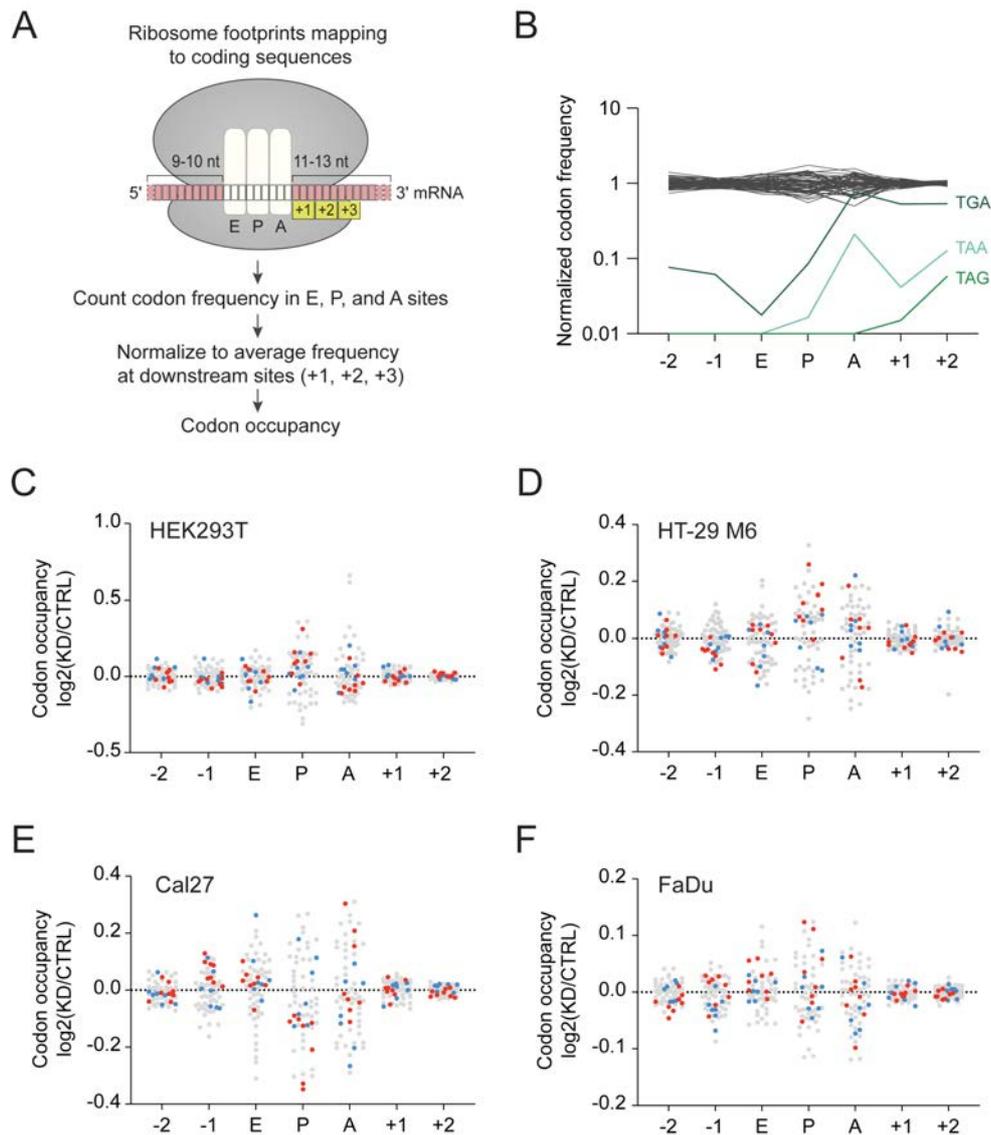


Figure 4.21. Global ribosomal codon occupancy. (A) Schematic representation of the approach used to determine codon occupancy at ribosomal sites. Adapted from (Nedialkova and Leidel, 2015). (B) Normalized frequencies at each footprint position for all 64 codons. Stop codons (TAA, TAG, and TGA) are shown in green, and the remaining 61 codons, in gray. Data corresponds to HEK293T samples. (C-F) Per-codon ribosome occupancy across footprint sites for ADAT2 KD relative to control cells. Shown is the average of two biological replicates for HEK293T (B), HT-29 M6 (C), Cal27 (D), and FaDu (E) cells. C-ended codons specifying ADAT amino acids are color coded in red, and G-ended codons corresponding to ADAT amino acids, in blue.

Ribosome occupancy for each of the 61 codons at E-, P-, and A-sites, as well as at the two adjacent up- and downstream positions, was determined for ADAT2 silenced cells relative to control cells (Figure 4.21 C-F). Although most of the variation in codon frequencies occurred within the A- and P-sites, which could indicate altered decoding rates and peptide-bond formation, these were relatively small changes with no statistical significance and were not reproducible across the different cells lines. This might be a consequence of too low

sequencing coverage combined with the fact that, based on our previous findings, we expect the effect of the depletion of I34 to be rather weak. Nevertheless, we decided to explore whether our data could reveal any trends suggestive of a translational defect in ADAT2 KD cells.

The decoding of C-ended (NNC) codons specifying ADAT amino acids is expected to be altered upon reduction of the I34-containing tRNA pool, as no cognate tRNAs exist for these codons in the human genome. On the other hand, G-ended (NNG) codons specifying ADAT amino acids cannot be efficiently decoded by neither I34- nor A34-containing tRNAs and, therefore, should not present decoding alterations resulting from changes in the levels of I34. We thus analyzed the relative abundance of ADAT NNC codons along footprint sites for ADAT2 KD cells, and used ADAT NNG codons as a negative control (**Figure 4.21 C-F**). A small bias towards an enrichment of some NNC codons in the P-site and their depletion in the A-site could be observed for ADAT2 KD HEK293T, HT-29 M6, and FaDu cells. However, NNG codons seemed to follow a similar pattern, with HT-29 M6 cells being the only ones displaying a wider separation between the two populations in the P-site. Additionally, these differences appeared to correlate better with amino acid identity than with codon identity. For example, increased occupancy of isoleucine codons at the P-site, or of leucine and valine codons at the A-site was present in all three cell lines with little codon-specific variation (**Figure 4.22, Figure 4.23, and Figure 4.25**). Strikingly, Cal27 cells exhibited the exact opposite behavior consisting in ADAT NNC codons being depleted in the P-site and enriched in the A-site (**Figure 4.21 F and Figure 4.24**). This fitted well our initial hypothesis whereby higher ribosome occupancy of ADAT NCC codons at the A-site would reflect slow decoding due to the decrease in I34 levels. However, Cal27 showed conflicting trends in general codon occupancy when compared with the other three cell lines, as well as an elevated variability in the E-site not observed in the other cases. These observations suggested that it was likely that the noted codon effects in this cell line were due to technical issues arising during sample preparation or sequencing. Altogether, our results for global codon occupancy seemed to point towards amino acid-specific variations rather than codon-driven ones.

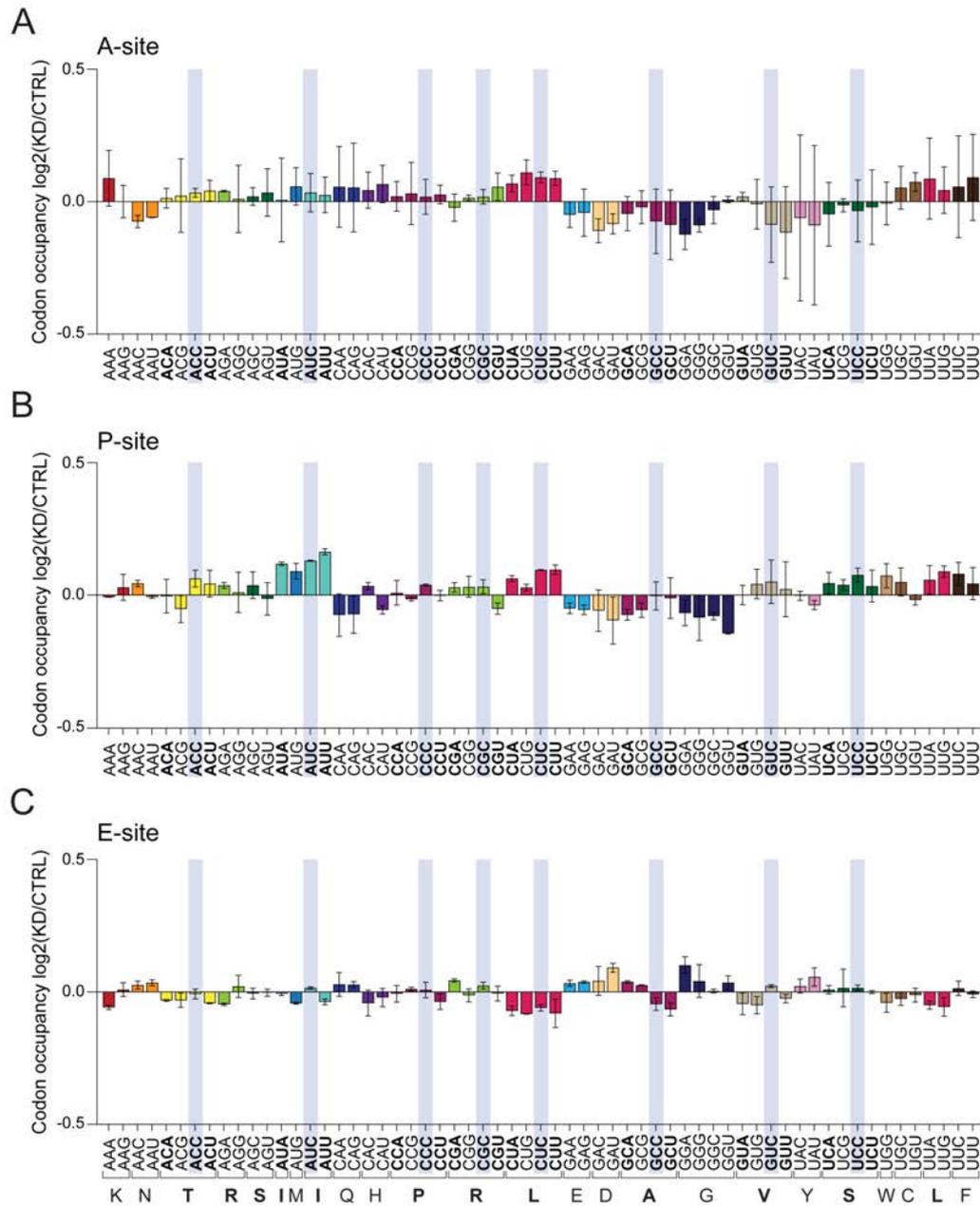


Figure 4.23. Global ribosomal codon occupancy in HT-29 M6 cells. Levels of ribosomal occupancy at the A-site (A), P-site (B), and E-site (C) for each of the 61 sense codons are shown. Values represent the mean and associated SD for two biological replicates. Bars are color-coded based on amino acid identity. ADAT amino acids and ADAT-sensitive codons are shown in bold. C-ended codons specifying ADAT amino acids are indicated in blue boxes.

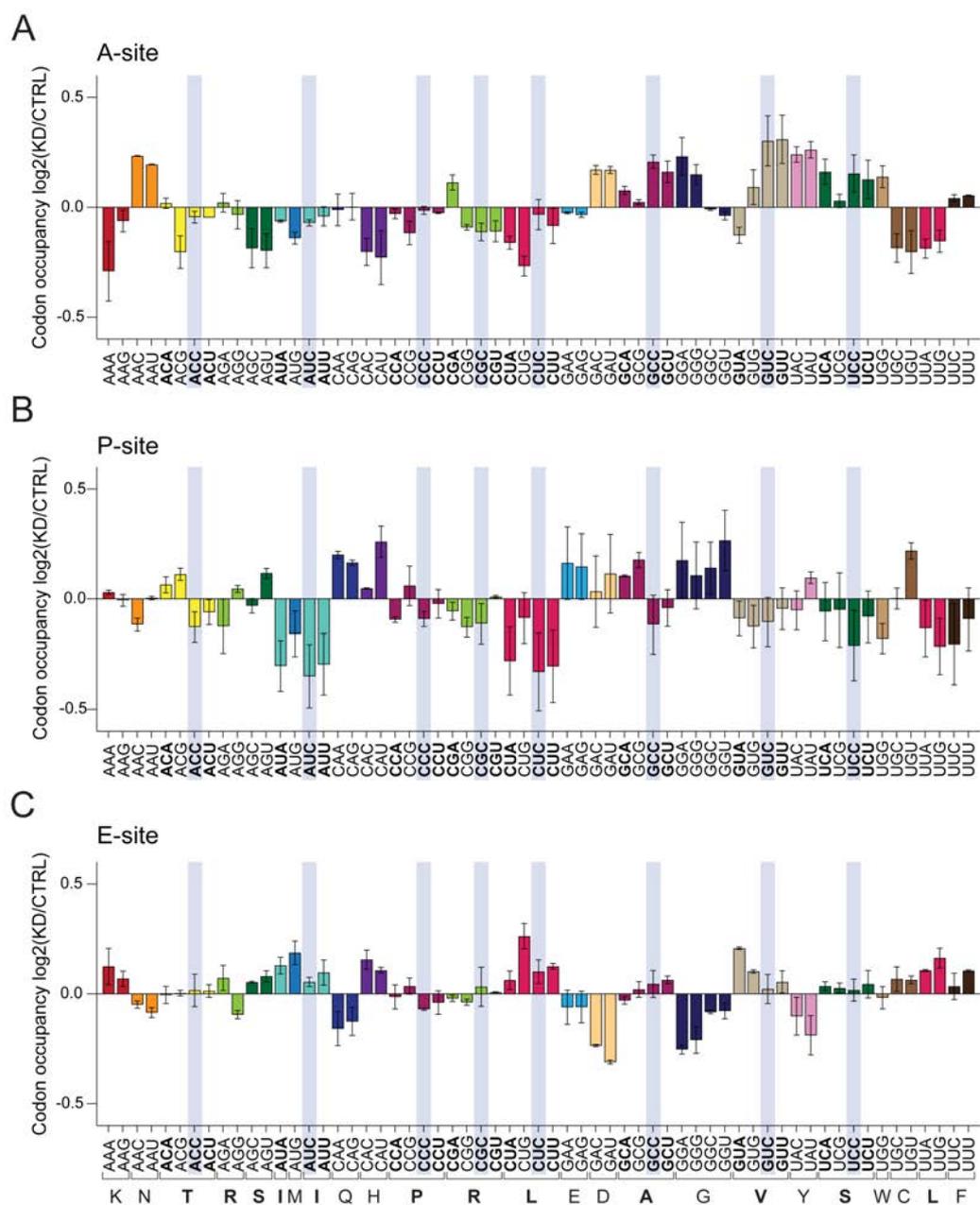


Figure 4.24. Global ribosomal codon occupancy in Cal27 cells. Levels of ribosomal occupancy at the A-site (A), P-site (B), and E-site (C) for each of the 61 sense codons are shown. Values represent the mean and associated SD for two biological replicates. Bars are color-coded based on amino acid identity. ADAT amino acids and ADAT-sensitive codons are shown in bold. C-ended codons specifying ADAT amino acids are indicated in blue boxes.

codon occupancy this time including only reads that had been mapped to genes with ADAT-stretches (**Figure 4.26**). Unfortunately, no clear patterns were observed and NNC and NNG codons showed similar trends to those found in global codon occupancy (**Figure 4.21 C-F**). One limiting factor in this analysis was the considerably smaller sample size, which lead to higher variability as it can be seen in the wider spread of codon values across all footprint positions, not limited to P- and A-sites. Another consideration that must not be omitted is the challenge that the repetitive nature of the ADAT-stretches sequences poses for precise read mapping. In our codon occupancy analyses, we used uniquely mapped reads with no mismatches in order to avoid artifacts. Therefore, we might miss ribosome footprint reads located within ADAT-stretches.

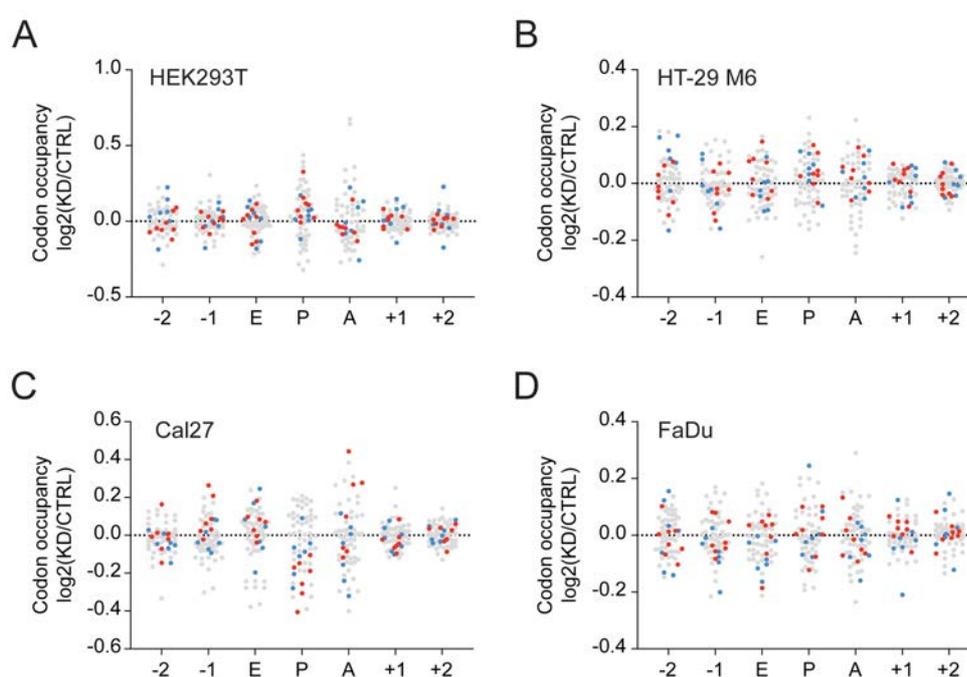


Figure 4.26. Ribosome codon occupancy in genes containing ADAT-stretches. Per-codon ribosome occupancy was computed as in **Figure 4.21**, except in this case only reads mapping to transcripts that encode ADAT-stretches were used. Aggregated data for each ribosomal position in ADAT2 KD relative to control HEK293T (**A**), HT-29 M6 (**B**), Cal27 (**C**), and FaDu (**D**) cells is shown as the mean of two biological replicates. C-ended codons specifying ADAT amino acids are color coded in red, and G-ended codons corresponding to ADAT amino acids, in blue.

It is possible that a reduction in the I34 tRNA pool has mixed effects on protein synthesis. While reduced decoding rates could be expected for ADAT NNC codons, evidence for A34 tRNAs being capable of decoding other codons than NNU has been reported (Borén et al., 1993). Additionally, modeling approaches have also suggested that wobble pairs with an A at position 34 within the ribosomal P-site can destabilize the adjacent A-site codon-anticodon

duplex (Lim, 1995). Altogether, these features may create confounding signatures that challenge the detection and interpretation of the contributions of I34 to translation elongation.

4.2.5. ADAT2 KD has variable impact on the synthesis of extracellular matrix proteins

Because no defects in general protein synthesis had been previously observed, and our ribosome profiling data did not show a clear picture of the impact of I34 in translation, we set out to investigate its effect on the synthesis of particular proteins that had been identified to contain stretches of ADAT-sensitive codons. As previously mentioned, the colorectal cancer HT-29 M6 cell line spontaneously differentiates at culture post-confluence. Its differentiated mucus-secreting phenotype is characterized by the synthesis of several brush-border associated enzymes such as dipeptidyl-peptidase IV (DPPIV), secreted mucins (MUC3 and MUC5AC), and transmembrane mucin 1 (MUC1) (Lesuffleur et al., 1990, 1993). While MUC1 expression is known to be present in pre-confluent cells, MUC5AC appears later in the differentiation process (Lesuffleur et al., 1993). We thus analyzed the expression of these two mucins before and after differentiation in control and ADAT2 silenced cells (**Figure 4.27**).

In agreement with previous reports, we found that MUC1 was already expressed at pre-confluent cultures (**Figure 4.27 A**, day 0) and that its protein levels increased with time in post-confluent culture, suggestive of a correct cellular differentiation. Unexpectedly, ADAT2 KD cells showed higher protein levels of MUC1 relative to control cells at all time points (**Figure 4.27 A**, upper panel). We wondered whether this increment could be accompanied by an accumulation of MUC1 in aggregate forms due to defective protein synthesis or processing in ADAT2 silenced cells. However, no appreciable detection of insoluble MUC1 was found in these cells (**Figure 4.27 A**, lower panel). Additionally, the pattern and extent of insoluble proteins detected by Coomassie staining was comparable between CTRL and ADAT2 KD cells at all time points. These results, together with the absence of UPR signaling, indicate that the lack of I34 does not result in accumulation of detectable protein aggregates. Although a modest increment in *MUC1* mRNA was observed at all time points for ADAT2 KD cells relative to control cells (**Figure 4.27 B**), it is unlikely that this can explain by itself the large amount of protein detected in these cells. Instead, this transcriptional response might constitute a compensation mechanism for defective protein synthesis.

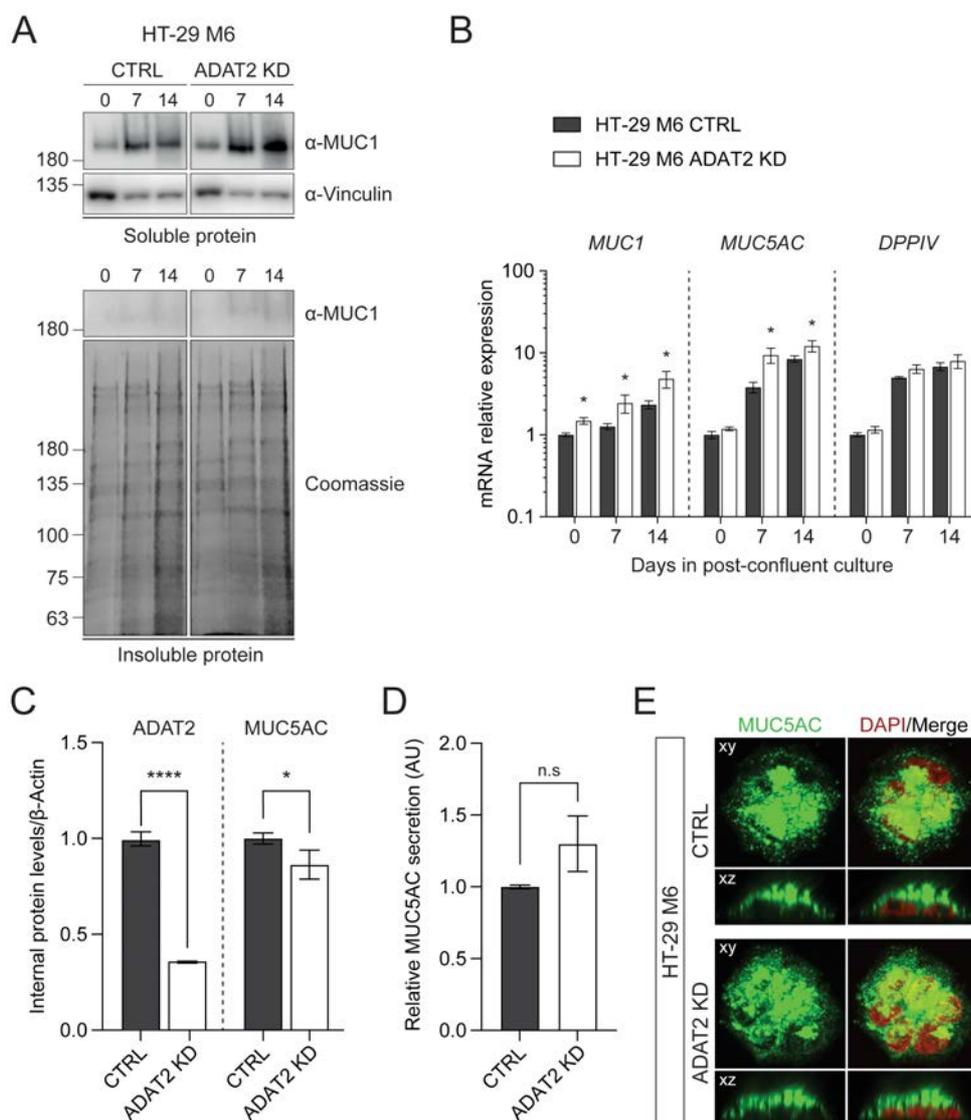


Figure 4.27. Mucin expression in HT-29 M6 cells. (A) Western blot detection of MUC1 protein levels in CTRL and ADAT2 KD HT-29 M6 cells. Coomassie staining of proteins in insoluble extracts was used as loading control. Numbers indicate the days the cells were in post-confluent culture, day 0 corresponds to pre-confluent cultures. (B) Relative mRNA levels of *MUC1*, *MUC5AC*, and *DPPIV* in CTRL and ADAT2 KD HT-29 M6 cells upon differentiation. Significant differences relative to CTRL cells at each time point were assessed using a t-test. * $P < 0.05$. (C) Intracellular ADAT2 and MUC5AC protein levels for CTRL and ADAT2 KD HT-29 M6 cells. Differentiated cells were lysed and analyzed by dot blot with anti-ADAT2, anti-MUC5AC, and anti- β -Actin antibodies. Intensities of ADAT2 and MUC5AC spots were quantified with ImageJ and normalized to β -Actin levels. Shown are the mean and associated SD for three independent biological replicates comprising 5 technical replicates each. Statistical significance was assessed using a t-test. * $P < 0.05$; **** $P < 0.0001$. (D) Secreted MUC5AC levels for CTRL and ADAT2 KD HT-29 M6 cells. Media of differentiated cells were collected, analyzed by dot blot with anti-MUC5AC antibodies, and quantified using ImageJ. Values are normalized to CTRL cells. Shown are the mean and associated SD for three independent biological replicates comprising 5 technical replicates each. Statistical significance was determined with a t-test. n.s., non-significant (E) Differentiated CTRL and ADAT2 KD HT-29 M6 cells were analyzed by immunofluorescence with anti-MUC5AC antibodies (green) and DAPI (red). Images represent a single plane (xy) and an orthogonal view (xz). (Data for figure panels C-E was obtained in collaboration with Dr Gerard Cantero).

MUC5AC mRNA was also found to be upregulated in ADAT2 KD cells compared to CTRL cells upon differentiation (**Figure 4.27 B**). However, neither *MUC1* nor *MUC5AC* mRNA increase was due to enhanced cellular differentiation, since the mRNA levels of the well-known differentiation marker *DPPIV* remained unchanged between the two conditions. Measurement of *MUC5AC* extra- and intra-cellular protein levels revealed a reduction of ~15% in the internal levels of *MUC5AC* (**Figure 4.27 C**) and a non-significant ($p=0.053$) increase of ~30% of secreted *MUC5AC* (**Figure 4.27 D**) in ADAT2 KD relative to control cells. These observations suggest that the amounts of *MUC5AC* synthesized by control and ADAT2 silenced cells are probably comparable, but that the latter secretes slightly more *MUC5AC* protein to the extracellular space through unknown mechanisms. Additionally, immunohistochemistry analysis of the mucin granules in these cells showed no major changes between ADAT2 KD and CTRL cells, with both conditions exhibiting normal granule size and apical distribution (**Figure 4.27 E**). Altogether, our results might indicate that the translation of *MUC5AC* is more affected than that of *MUC1* by the depletion of ADAT2, as an increase in its mRNA levels leads to similar amounts of protein to those of control cells. Consistent with this is the fact that *MUC5AC* is the human protein containing the largest ADAT-stretch identified in our transcriptome-wide analysis, and thus its translation might be further compromised when the I34-containing tRNA pool is altered.

In addition to the effects observed for mucins, we also explored the status of other proteins containing ADAT-stretches. Dystroglycan was identified to contain an ADAT-stretch in the mucin-domain of its alpha subunit. Detection of α DG with the glycan-specific IH6 antibody revealed a reduction in protein levels in ADAT2 KD relative to control undifferentiated HT-29 M6 cells (**Figure 4.28 A**). Because the two dystroglycan subunits are encoded by a single mRNA as an N-ter α DG and a C-ter β DG (**Figure 1.18 A**), we also analyzed the protein expression of the β subunit, as defects in the translation of α DG are likely to affect the synthesis of the downstream β DG. Albeit less pronounced, a decrease in β DG was also observed (**Figure 4.28 A**). While it is possible that the disparity in the levels of α DG and β DG could be due to alterations in the glycosylation patterns of α DG that are detected by the antibody, these changes would be expected to be rather modest given that the bands run at comparable heights. Considering that *DAG1* mRNA levels did not change between conditions (**Figure 4.28 B**), it is likely that the decrease detected by Western blot arises from lower translational activity, which might lead to further defects in the post-translational processing of dystroglycan.

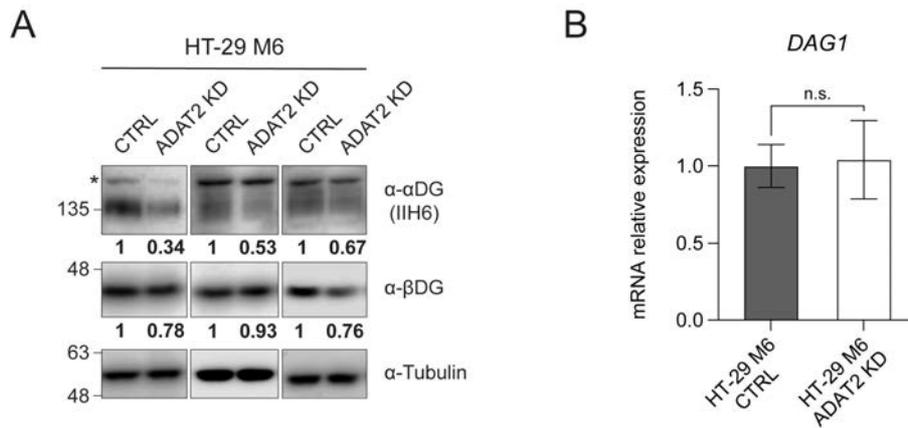


Figure 4.28. Dystroglycan expression in HT-29 M6 cells. (A) Western blot detection of protein levels of dystroglycan subunits with anti-αDG-glycan antibody IIH6 and anti-βDG for three independent biological replicates in CTRL and ADAT2 KD HT-29 M6 cells. Quantification of their relative signal intensity to Tubulin is shown in bold. Asterisk indicates a non-specific band detected by the primary antibody. **(B)** Relative mRNA levels of *DAG1* in CTRL and ADAT2 KD HT-29 M6 cells. Statistical significance was tested with a t-test. n.s., non-significant.

We next identified a third candidate gene encoding an ADAT-stretch whose protein levels were affected by the lack of ADAT2 (**Figure 4.29**). Expression of syndecan-3 was examined in HEK293T cells due to the absence of a clear signal detected by Western blot in the HT-29 M6 cell line. Syndecan-3 was found to be downregulated by ~ 30% at the protein level in ADAT2 KD HEK293T cells (**Figure 4.29 A**), while its mRNA levels remained unaltered (**Figure 4.29 B**). Altogether, these results suggest that the lack of ADAT2 might lead to defects in translation of genes encoding stretches of ADAT-sensitive codons that, under certain circumstances, could be compensated by transcriptional upregulation.

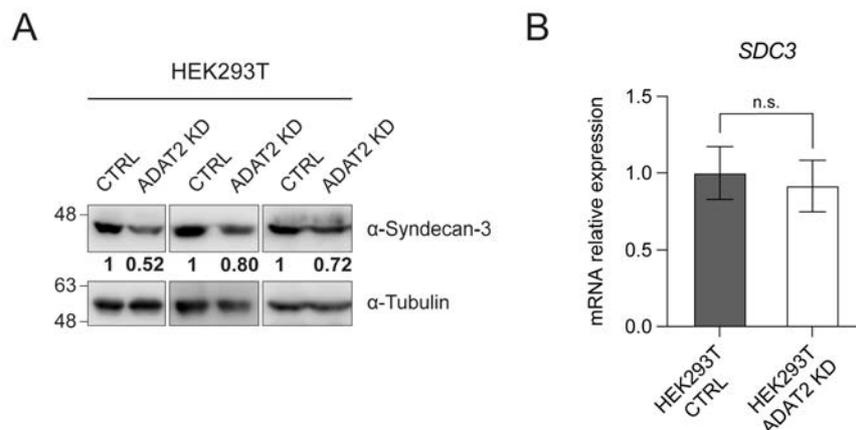


Figure 4.29. Syndecan-3 expression in HEK293T cells. (A) Western blot detection of Syndecan-3 protein levels for three independent biological replicates for CTRL and ADAT2 KD HEK293T cells. Quantification of their relative signal intensity to Tubulin is shown in bold. **(B)** Relative mRNA levels of *SDC3* in CTRL and ADAT2 KD HEK293T cells. Statistical significance was tested with a t-test. n.s., non-significant.

4.2.6. ADAT2 KD does not alter mucin glycosylation patterns

Based on the variable effect of the knockdown of ADAT2 on the synthesis of proteins containing ADAT-stretches, we tested whether small defects introduced at the translation level, such as reduced elongation rates or mistranslation events, could have an impact on the post-translational fate of these proteins. The most relevant post-translational modification undergone by proteins enriched in ADAT-stretches is glycosylation at serine or threonine residues, both of which are ADAT amino acids. Therefore, misreading of their codons could lead to decreased glycosylation at those positions. Similarly, altered elongation rates could result in moderate modulations of protein folding that, despite not being sufficient to activate UPR or produce large aggregates, may lead to altered glycosylation processing in the ER and Golgi apparatus. We tested this hypothesis by analyzing the glycan content of mucins given that they are the most heavily glycosylated proteins amongst those examined in this study.

Purification of mucins was performed by solubilization of cellular extracts in guanidinium chloride (GuHCl) followed by double isopycnic density-gradient centrifugation in CsCl (**Figure 4.30**) (Davies et al., 2012). Due to their heavy carbohydrate content, mucins have a buoyant density that is higher than that of most proteins and lower than that of proteoglycans and nucleic acids. Density-gradient centrifugation in CsCl thus allows separation of mucins from proteins and DNA. In the first step, centrifugation in 4M GuHCl/CsCl with a starting density of 1.41 g/mL generates gradients of ~ 1.2 – 1.6 g/mL that allow separation of mucins from proteins (**Figure 4.30 A**). Fractions collected from the gradients were separated on SDS-PAGE followed by staining with periodic acid-Schiff (PAS) and Coomassie. PAS specifically detects glycoproteins, while Coomassie has low reactivity towards them due to the sugars surrounding the protein backbone (Tytgat et al., 1994). Non-mucin proteins stained by Coomassie had densities between 1.2 and 1.3 g/mL. In contrast, high molecular weight species that were positive for PAS staining and had densities (1.32 – 1.46 g/mL) consistent with that of mucins (Strous and Dekker, 1992) were detected in fractions 9 – 19. Because mucins are known to co-migrate with DNA to densities between 1.4 and 1.5 g/mL, a second purification step was required to ensure removal of DNA. Fractions 9 – 19 were pooled, dialyzed against 0.5 M GuHCl, and subjected to density-gradient centrifugation with a starting density of 1.5 g/mL that gives gradients of 1.35 – 1.65 g/mL (**Figure 4.30 B**). Notably, no Coomassie staining was detected in any of the fractions, thus confirming a correct separation in the previous purification step. Fractions 7 – 17, which had densities of 1.4 – 1.5 g/mL and were determined to contain mucins by PAS staining, were then pooled, dialyzed against PBS, and concentrated for further analysis.

Mass spectrometry analyses of glycan structures present in purified mucins from CTRL and ADAT2 KD HT-29 M6 cells showed similar O-glycan profiles consisting of a mixture of core types 1 and 2 O-glycans. Both samples exhibited a high proportion of sialylated glycans and the presence of O-acetylation on sialic acid (Neu5Ac) (**Figure 4.31**). The major O-glycans identified were Neu5Ac α 2-3Gal β 1-3GalNAc (peak number 12), its peeled product Neu5Ac α 2-3Gal (peak number 3) originated during sample processing, and Gal α 1-3GalNAc (peak number 2), which was consistent with previous analysis of mucin O-glycans in HT-29 M6 cells (Hennebicq-Reig et al., 1998; Huet et al., 1995; Lesuffleur et al., 1993). While no differences were observed in the structures identified in control and ADAT2 KD cells, a slight relative increase in sialylated glycans was noted for ADAT2 KD cells (peak number 3, 10, and 12) (**Figure 4.31 B**). We interpreted this modest increment as a reflection of the small sample size used in the experiment and not as a physiological process derived from the silencing of ADAT2. This idea was supported by the fact that no shifts in the electrophoretic mobility of MUC1, which could be attributed to an increased negative charge due to higher abundance of sialic acid molecules, had been observed by Western blot in ADAT2 KD cells relative to control cells (**Figure 4.27 A**). Additionally, we also considered the possibility that the activity of sialyltransferases, the enzymes responsible for the addition of sialic acid to glycan residues, or that of neuraminidases (NEU1, NEU2, and NEU4) that cleave sialic acid from mucins, could be affected by the silencing of ADAT2 leading to higher sialylation of O-glycans in other mucins than MUC1. However, neither the genes encoding sialyltransferases that attach sialic acid to Gal residues at position α 2-3 or to GalNAc at position α 2-6, nor those coding for mucin neuraminidases, were enriched in ADAT-sensitive codons, thus suggesting that defects in their translation that might result in altered enzymatic activity upon ADAT2 KD were unlikely. Therefore, we believe that it is plausible that the differences in glycan sialylation observed between control and ADAT2 KD cells were a result of the intrinsic variability of the assay rather than a biological consequence of the depletion of I34-containing tRNAs.

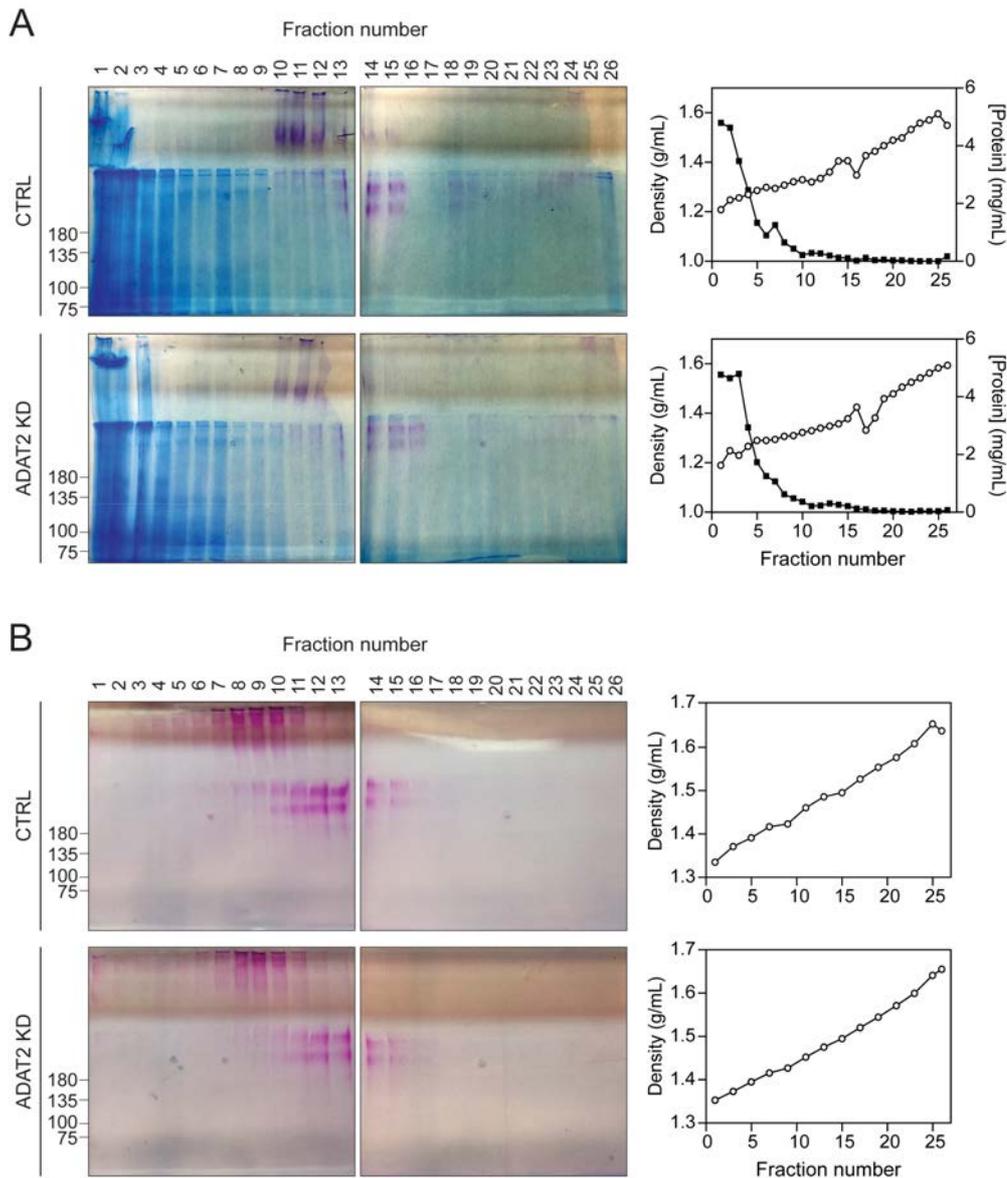


Figure 4.30. Purification of mucins from HT-29 M6 cells by two-step CsCl isopycnic density-gradient centrifugation. (A) Representative PAS- and Coomassie-stained SDS-PAGE of fractions from isopycnic density-gradient centrifugation in 4 M guanidinium chloride/CsCl of extracts from differentiated CTRL (upper panel) and ADAT2 KD (lower panel) HT-29 M6 cells. Gradients with a starting density of 1.41 g/mL were used to separate mucins from proteins. Fractions were analyzed for density (open circle) and protein concentration (filled square) (B) Representative PAS- and Coomassie-stained SDS-PAGE of fractions from isopycnic density-gradient centrifugation in 0.5 M guanidinium chloride/CsCl of fractions 9-19 from (A) for CTRL (upper panel) and ADAT2 KD (lower panel) HT-29 M6 cells. Gradients with a starting density of 1.5 g/mL were used to separate mucins from nucleic acids. Fractions were analyzed for density as in (A). Fractions 7-15 were pooled, dialyzed against PBS, and concentrated for further analysis.

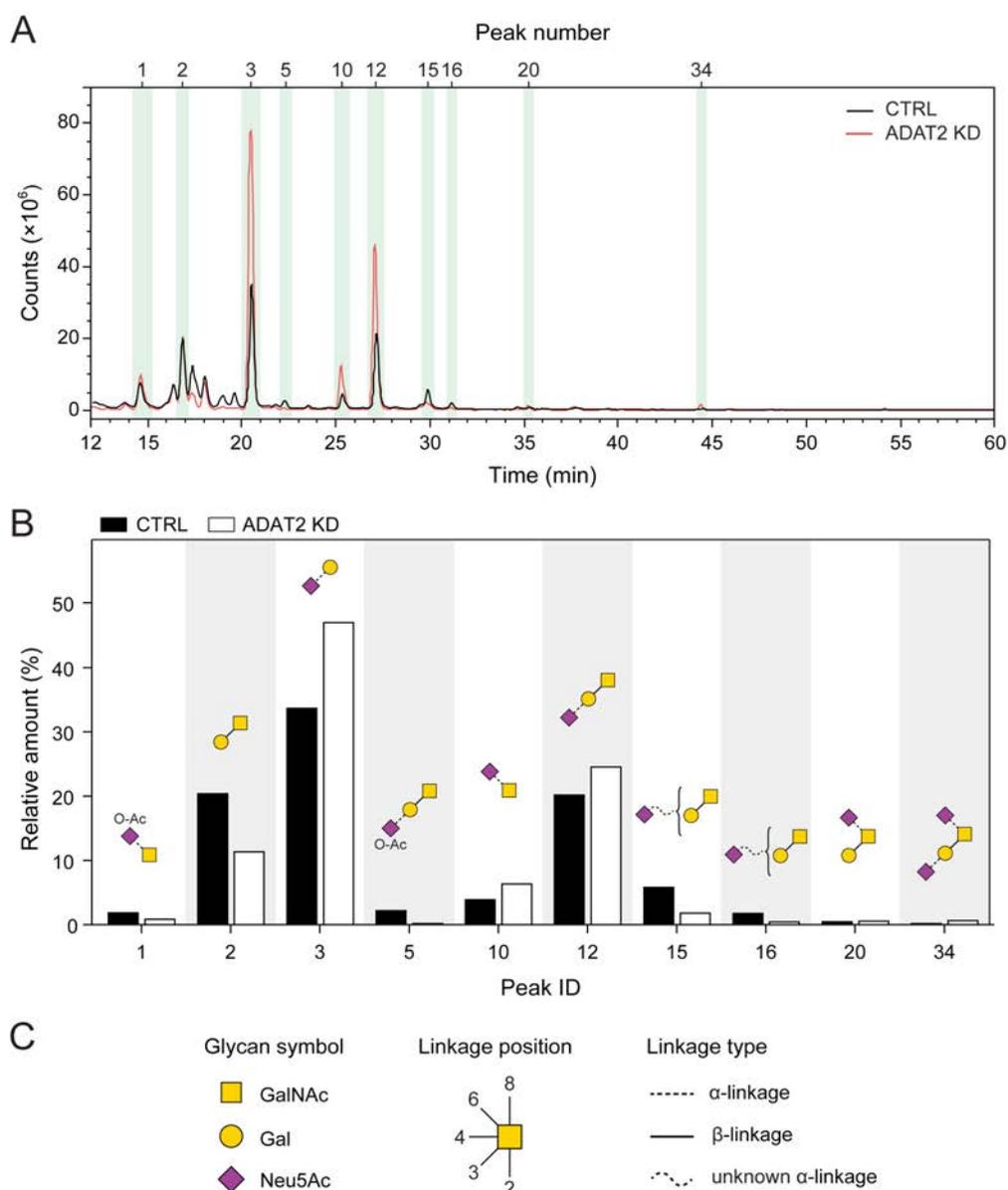


Figure 4.31. Mass spectrometry analysis of glycan structures present in mucins from HT-29 M6 cells. (A) HILIC-UHPLC chromatographic profiles for procainamide labelled O-glycans released by hydrazine from purified mucins present in CTRL and ADAT2 KD HT-29 M6 cells. The most abundant glycan species for which a structure was determined are labeled with their peak number and shadowed in green. (B) Relative abundance of the major identified species in CTRL and ADAT2 KD cells as calculated by quantification of integrated peak areas. The structure of each glycan molecule is shown using the standard symbol nomenclature for glycans (Varki et al., 2015). Values correspond to $n=1$. (C) Abbreviations for glycan structure symbols, linkage position and linkage type used in (B).

4.2.7. ADAT2 is required for MUC5AC protein synthesis in an *in vitro* cellular model of airway remodeling

The mild effects of the ADAT2 KD on mucin protein and mRNA levels observed in HT-29 M6 cells might suggest that the lack of I34 leads to the activation of transcriptional compensatory mechanisms (El-Brolosy and Stainier, 2017). Accordingly, if the translation of certain genes was defective, an upregulation of their corresponding transcripts could serve as a cellular adaptation to that insult, thus enhancing their protein output. Furthermore, we reasoned it was plausible that, under circumstances in which the demand for I34-modified tRNAs was not sufficiently high, the requirement of these tRNAs for translation of proteins enriched in ADAT-stretches might not be detectable. This could also explain the absence of clear signals for codon decoding alterations in polysome and ribosome profiling analyses. Therefore, we sought an alternative system that would yield a sudden and considerable amplification of this demand, while still resembling a physiological state.

It has been extensively reported that stimuli including inflammatory cytokines, bacterial products, growth factors, and environmental pollutants are involved in goblet cell hyperplasia (increased number of goblet cells responsible for mucus production) in models of hypersecretory airway diseases (Rose and Voynow, 2006; Thai et al., 2008). Many of these factors have been shown to rapidly induce MUC5AC expression both *in vivo* and *in vitro* on airway epithelial cells by activation of epithelial growth factor receptor (EGFR), which initiates the RAS/MEK/ERK signaling pathway leading to increased transcription of *MUC5AC* (Burgel and Nadel, 2008; Lemjabbar-Alaoui et al., 2011; Lemjabbar et al., 2003). One such factor is amphiregulin (AREG), an EGF-like growth factor that plays a key role in tissue remodeling in asthma and COPD by promoting increased cell proliferation and mucin production (Manzo et al., 2012; Shoyab et al., 1989). AREG has been previously used to induce mucous cell metaplasia (the process of differentiation into mucous-producing cells) in human mucoepidermoid lung carcinoma NCI-H292 cells, a situation thought to be analogous to the milieu of airway epithelia in hypersecretory diseases (Enomoto et al., 2009). We decided to use this cellular model to study the effect of the lack of ADAT2 in MUC5AC translation upon stimulation with AREG as it provides a physiologically relevant stimulation of the synthesis of large amounts of this protein.

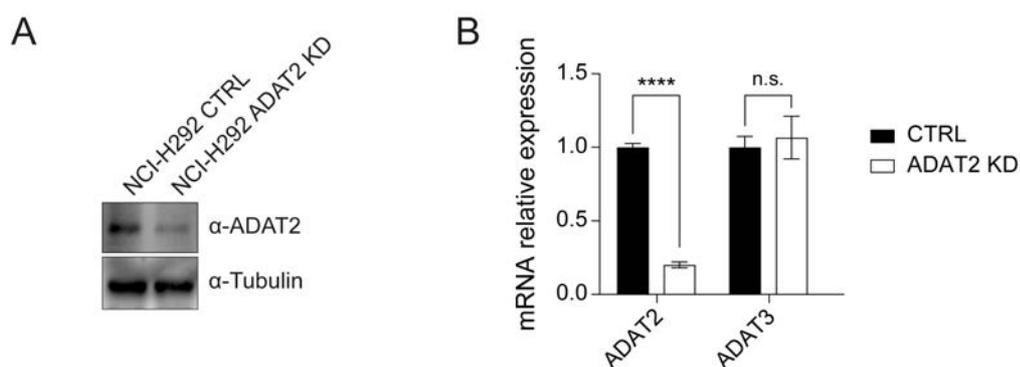


Figure 4.32. ADAT2 silencing in NCI-H292 cells. (A) Western blot showing reduced protein levels of ADAT2 in NCI-H292 ADAT2 silenced cells with respect to control cells. (B) mRNA levels for ADAT2 decrease in ADAT2 KD, but transcript levels for ADAT3 remain the same. Statistical significance was determined using a t-test. **** $P < 0.0001$; n.s., non-significant.

We generated control and ADAT2 silenced NCI-H292 cells using the same shRNAs previously described (Figure 4.32). Cells grown in the presence or absence of 50 nM AREG over a period of four days were tested for MUC5AC expression by immunofluorescence (Figure 4.33 C). As expected, MUC5AC was barely present in untreated cells, whereas a significant increase in MUC5AC protein expression was detected upon treatment with AREG (Figure 4.33 C, D). Most striking was the roughly 3-fold reduction in MUC5AC staining observed for treated ADAT2 KD cells relative to treated control cells. Flow cytometry analysis of AREG-treated cells showed that not only was the mean signal intensity of MUC5AC markedly reduced when ADAT2 was silenced (Figure 4.34 A, C), but also the number of cells expressing this protein decreased (Figure 4.34 B, D). These results were consistent with the 3-fold reduction in MUC5AC expression observed by immunocytochemistry for ADAT2 KD cells.

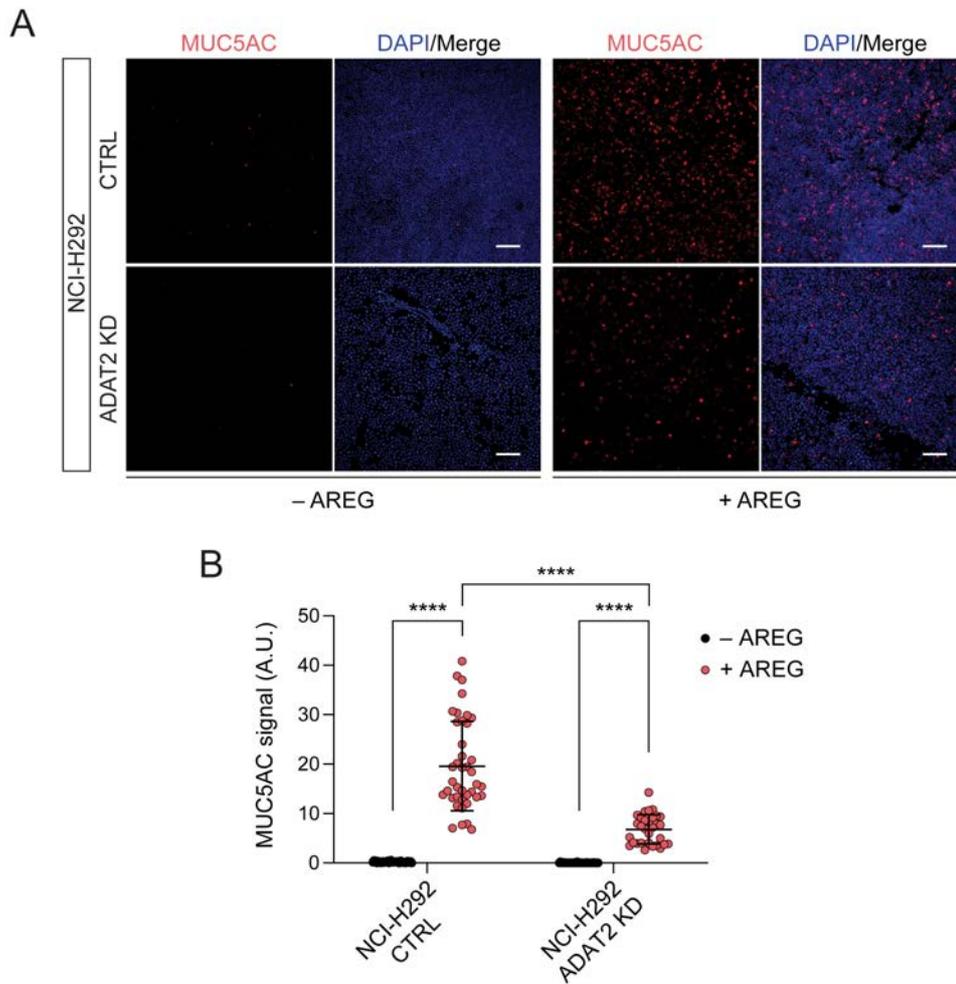


Figure 4.33. Immunostaining detection of MUC5AC in AREG-treated NCI-H292 cells. (A) Representative confocal microscopy images of MUC5AC immunostained NCI-H292 cells after 4 days in culture with or without 50 nM AREG. Scale bar corresponds to 10 μ m. DAPI was used to identify nuclei. (B) Quantification of MUC5AC signal intensity shown in (A), normalized to the DAPI area. A minimum of 31 different microscopy images were quantified per condition. Shown are the mean and associated SD. Statistical significance was obtained with a t-test. **** $P < 0.0001$.

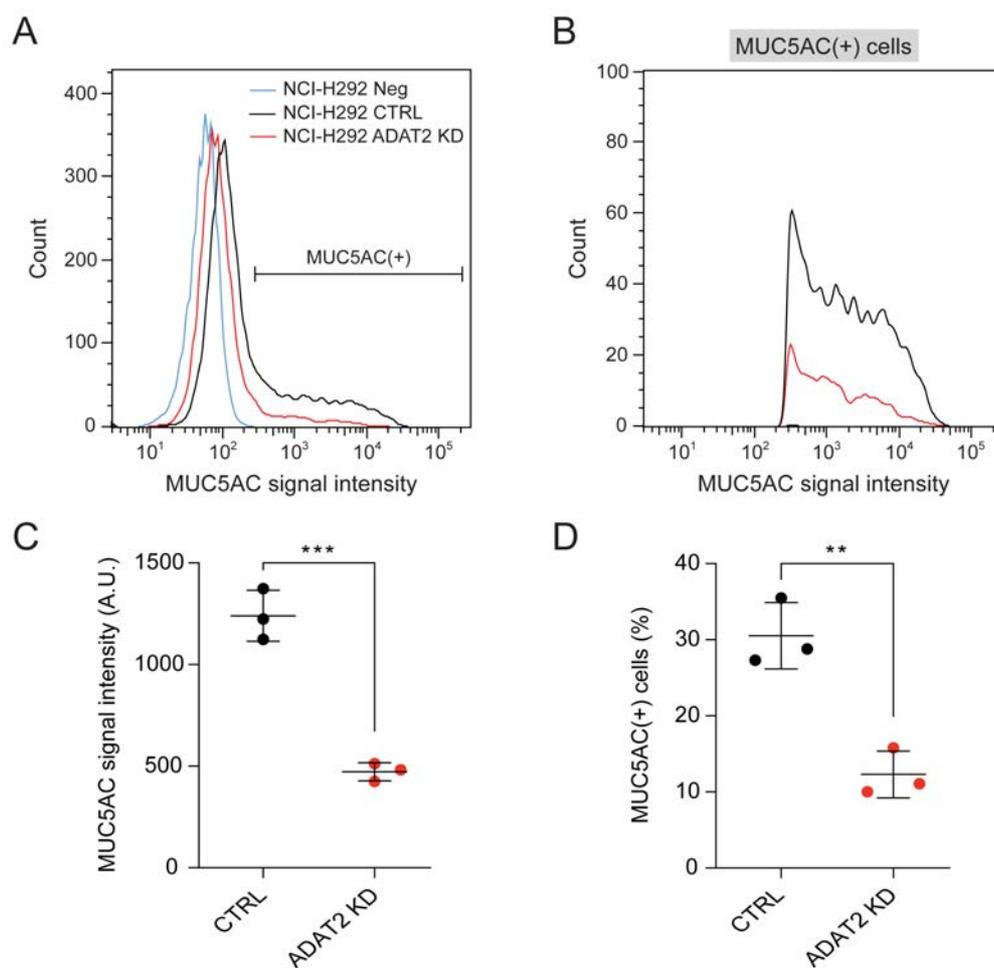


Figure 4.34. Flow cytometry analysis of MUC5AC expression in AREG-treated NCI-H292 cells.

(A) Representative histogram showing the number of cells and their MUC5AC signal intensity as detected by flow cytometry. Cells treated with AREG for 4 days were first stained with an anti-MUC5AC antibody and then with an AlexaFluor-488 anti-mouse IgG secondary antibody. Untreated CTRL cells stained only with secondary antibody (NCI-H292 Neg, blue line) were used as a negative control to set the gate for MUC5AC(+) cells. AREG-treated CTRL and ADAT2 KD cells are shown in black and red, respectively. (B) As in (A), but showing only cells that were determined positive for MUC5AC signal. (C) Quantification of MUC5AC signal intensity shown in (A) for three independent biological replicates. Shown are the mean and associated SD. Statistical significance was determined using a t-test. *** $P < 0.001$. (D) Quantification of the percentage of MUC5AC positive cells in (B) for three independent biological replicates. Statistical significance was computed with a t-test. ** $P < 0.01$.

Two possible explanations for this phenotype were: (1) ADAT2 KD cells do not properly differentiate into a mucous phenotype upon AREG treatment and, therefore, *MUC5AC* transcription is not upregulated, or (2) depletion of ADAT2 results in impaired translation of *MUC5AC* mRNA. We tested the first hypothesis through Western blot examination of the phosphorylation of EGFR and ERK that is induced by binding of AREG to EGFR (**Figure 4.35 A**). Comparable levels of phospho-EGFR and ERK were observed for control and KD cells, suggesting that AREG binds and activates EGFR equally in both conditions. However,

because defects in differentiation might occur downstream of ERK signaling, we also measured the transcript levels of goblet cell-associated genes (**Figure 4.35 B**). These included *MUC5AC*, *MUC5B*, the transcription factor *SPDEF* (Chen et al., 2009), the ER sensor *IRE1* known to play a role in the folding of mucins (Martino et al., 2009), the glycosyltransferase *GCNT3* (Chen et al., 2009), and the gene encoding *AREG*, as it has been reported that EGFR activation with amphiregulin creates a positive feedback loop of the AREG/EGFR/ERK pathway (Huang et al., 2017; Zuo et al., 2017). While there was a significant upregulation in the mRNA level of differentiation markers for both control and KD cells upon treatment with AREG, no differences in expression were observed between AREG-treated control and KD cells (**Figure 4.35 B**), indicating that both cell lines uniformly activate the transcriptional programs associated to mucous differentiation. Interestingly, we also noticed that the protein and transcript levels of ADAT2 were increased upon AREG treatment in both control and ADAT2 KD cells (**Figure 4.35 A, B**). While this could suggest the existence of a mechanism whereby I34 is upregulated according to the translational requirements of the cell, further research is required. Altogether, our results confirmed the dependence on I34-containing tRNAs for the efficient synthesis of transcript regions encoding ADAT-stretches such as those in *MUC5AC*.

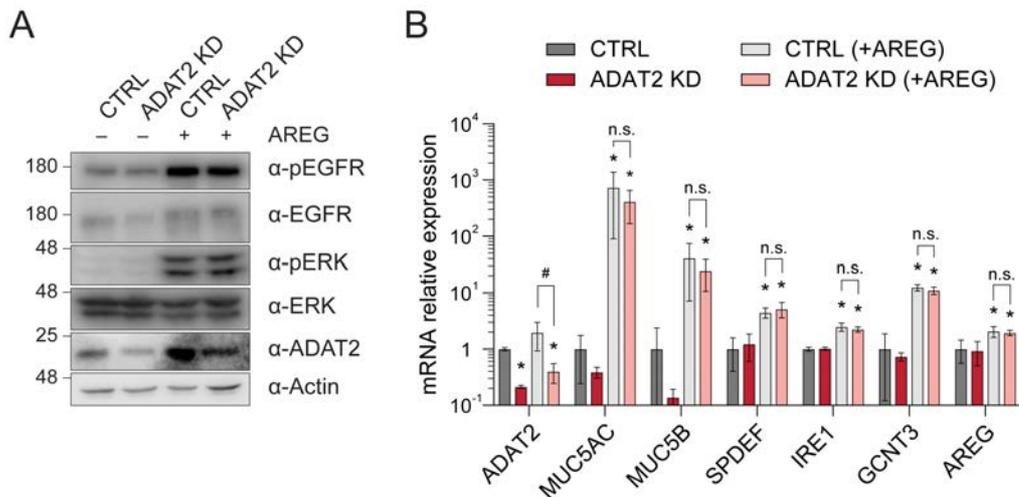


Figure 4.35. Expression of mucous differentiation markers and AREG activation pathway in NCI-H292 cells. (A) Western blot showing activation of the EGFR-induced signaling cascade upon AREG binding for CTRL and ADAT2 KD NCI-H292 cells. (B) Relative mRNA levels of mucous differentiation markers and *ADAT2* upon AREG treatment. Shown are the mean and associated SD for three independent biological replicates. Statistical significance was determined using the Kurskal-Wallis method. * $P < 0.05$ vs. “CTRL”; # $P < 0.05$ vs. “CTRL (+ AREG)”.

5. DISCUSSION

Due to their critical role in gene expression, it is not surprising that alterations in tRNA modifications and the enzymes that catalyze their synthesis are intimately associated to human diseases (de Crécy-Lagard et al., 2019; Hawer et al., 2018; Torres et al., 2014a). A large body of evidence, boosted by the application of next-generation sequencing in clinical settings, has uncovered a link between an exponentially increasing number of genetic variants of tRNA modification enzymes and neurological disorders that are often congenital and degenerative in nature (Ramos and Fu, 2019; Schaffer et al., 2019). However, the molecular bases for the pathogenicity of most mutations remain unclear. Similarly, the functional relationship between a valine to methionine substitution (V144M) in the ADAT3 subunit of human ADAT and the onset of intellectual disability to which it is connected is yet to be determined (Alazami et al., 2013; El-Hattab et al., 2016; Sharkia et al., 2018). Here, we applied a recently developed method for the detection of I34 formation in tRNAs (Wulff et al., 2017) to study the kinetics of human ADAT and to ascertain the impact of the V144M mutation. Furthermore, we also investigated the effect of the V144M substitution on the enzyme's quaternary structure and its cellular distribution.

In silico studies had previously predicted that the V144M mutation alters the 3D conformation of ADAT3 (Alazami et al., 2013). Consistent with this proposal, we found that the mutation partially destabilizes the tertiary structure of ADAT3. Size-exclusion chromatography and mass spectrometry experiments showed that the mutant ADAT3 subunit is cleaved by proteases that do not act upon the wild type subunit (**Figure 4.1**, **Figure 4.2**, and **Figure 4.3**). This cleavage is likely to occur during preparation of the *E. coli* crude extracts and does not affect the formation of the heterodimer, which, presumably, was previously attained. Likewise, the co-purification of tRNA with the mutant enzyme suggests that neither the mutation nor the cleavage of the ADAT3 subunit impair the binding affinity of ADAT* for tRNA (**Figure 4.2 B**).

In order to evaluate the activity of the wt and mutant enzymes, we used the RFLP method to determine their catalytic constants (**Figure 4.6** and **Table 4.1**). Our results indicated that the mutation causes a 10-fold decrease in catalytic efficiency, which is attributable to a 5-fold reduction in the k_{cat} value of the mutant enzyme. Consistent with the similarity in K_m values, and as suggested by our size-exclusion chromatography data, EMSA experiments showed that both wt ADAT and ADAT* bind cognate tRNAs with similar affinities (**Figure 4.4**). While it's true that our protein preparation for ADAT* contained a truncated form of ADAT3, we reckon that this factor cannot explain by itself the measured decrease in k_{cat} . Considering the caveats of expressing this protein in a bacterial organism such as *E. coli*, alternative heterologous systems could be used in the future to obtain an unperturbed protein with which to confirm the observed results.

The apparently moderate impact of the V144M mutation upon the enzyme's activity is consistent with the idea that partial function must be retained for viability of the organism. However, when considered in the context of its severe clinical phenotypes, this modest decrease in catalytic activity is intriguing. Three possible explanations for these contrasting biochemical and medical observations are: (1) human tissues, particularly neuronal tissues, might be extremely sensitive to variations in ADAT activity due to a requirement of a delicate fine-tuning of gene expression for normal development; (2) the effect of the V144M mutation is more severe upon other ADAT substrates untested in this study; (3) the V144M mutation may be affecting other aspects of ADAT's biology and therefore inducing toxicity through currently unknown mechanisms.

We hypothesized that one such aspect could be the subcellular location of the enzyme. The cellular distribution of human ADAT had never been directly determined before. Previous studies using animal cellular models had reported the presence of A34-to-I34 tRNA editing activity in the cytoplasm of *Xenopus laevis* oocytes (Fournier et al., 1983), in HeLa nuclear extracts (Achsel and Gross, 1993), and in anucleate rabbit reticulocyte lysates (Auxilien et al., 1996). Consistent with these observations, our results indicated that that human ADAT is predominantly a nuclear enzyme, with a small proportion of the heterodimer remaining in the cytoplasm (**Figure 4.7** and **Figure 4.8**). This is also in agreement with the presence of I34 in precursor tRNAs (Gogakos et al., 2017; Torres et al., 2015). Additionally, our experiments with fluorescent ADAT subunits suggested an ADAT2-mediated nuclear import mechanism of ADAT3. A scenario in which the two subunits are translated in the cytosol, where they interact forming an active heterodimer, and are subsequently translocated into the nucleus in an ADAT2-dependent manner seems plausible.

As opposed to what we had observed for wt ADAT3, we found that the V144M substitution causes delocalization of overexpressed ADAT3*, which is now present in both cytosol and nucleus (**Figure 4.10**). Thus, the V144M mutation induces a partial import of ADAT3* to the nucleus that is independent of ADAT2. Although the presence of ADAT3* in nucleus is not a result of proteolytic cleavage of the protein (**Figure 4.13**), the structural destabilization caused by the mutation may facilitate the transit of ADAT* through the nuclear pore (Lemke, 2016). Alternatively, it may allow ADAT3* to bind nuclear import factors or lead to the disruption of interactions that anchor ADAT3 to the cytosol.

In the presence of stoichiometric quantities of ADAT2, ADAT3* localizes to the nucleus to the same extent as wt ADAT3 (**Figure 4.10**). Thus, pathological effects of the altered intracellular distribution of ADAT3* may occur when ADAT2 and ADAT3 are not expressed at stoichiometric levels. In this sense, the Human Protein Atlas reports that ADAT2

expression levels are significantly lower than those of ADAT3 in brain tissues (www.proteinatlas.org (Uhlen et al., 2010, 2015)). In addition, the expression of ADAT2 at the tissue level can be dynamic. For example, mRNA levels of ADAT2 are increased more than two-fold in colon adenocarcinoma with respect to normal tissue, while ADAT3 mRNA levels remain unchanged (Oncomine: TCGA Colorectal datasets).

Considering these results, we cannot exclude the possibility that the phenotype caused by the V144M mutation might not exclusively arise from a decrease in I34 formation, but also affect uncharacterized alternative functions of human ADAT that could require the maintenance of a fine balance between its nuclear and cytoplasmic localization. In *T. brucei*, interaction of ADAT with the m³C methyltransferase TRM140a (human METTL6) leads to the formation of m³U at position 34 of tRNA^{Thr} and prevents widespread genomic DNA C-to-U deamination by ADAT while present in the nucleus (Rubio et al., 2017). Similarly, other tRNA modification enzymes have been shown to carry out alternative roles not linked to their canonical catalytic activity (Gutgsell et al., 2000; Johansson and Byström, 2002; Keffer-Wilkes et al., 2016; Persson et al., 1992). Additionally, we cannot rule out that the V144M mutation could also cause a gain-of-function of ADAT3 that results in pathogenic effects.

A clearer picture of the global cellular impact of the ADAT3 V144M mutation might be provided by studying its effects under physiological expression levels, as it happens in the cells of affected patients. However, despite our attempts at generating a cellular model in human cells through the introduction of the V144M substitution in the genome by CRISPR/Cas9, we were unable to select cells containing the mutation in homozygosis. We did obtain heterozygote cell lines carrying the mutation, but the mutant *ADAT3* allele quickly reverted to its wt form in cell culture (data not shown). Thus, while the V144M mutation in homozygosis is viable in humans, our results indicate that it is not tolerated by cultured cells. It should be noted that the V144M mutation has been described in a small number of families with a limited geographical distribution (a total of 40 individuals from 20 Middle Eastern families) (Alazami et al., 2013; El-Hattab et al., 2016; Sharkia et al., 2018). It is possible that the genetic background where *ADAT** has been reported provides compensatory genetic variants that reduce the impact of the mutation in most tissues.

Taken together, our results are compatible with a pleiotropic cellular effect of the V144M mutation caused by its impact upon ADAT's structure and activity, and its effect on the enzyme's cellular distribution under conditions where ADAT2 and ADAT3 levels are non-stoichiometric. As stated above, it remains to be determined whether the modest reduction in enzymatic activity is enough to have an impact on protein synthesis, which could contribute to the pathogenic phenotype associated to the ADAT3 V144M mutation. Mature tRNAs that

are substrate of ADAT are fully modified with I34 in both *S. cerevisiae* and *H. sapiens*, and as opposed to other tRNA modifications, reports of fluctuations in the extent of the I34 modification are rare (Arimbasseri et al., 2015; Chan et al., 2018; Gogakos et al., 2017; Heiss et al., 2017; Huber et al., 2019; Torres et al., 2015). Together with the essential nature of ADAT, this suggests that I34 is a constitutive feature required for proper cellular functionality.

However, the precise function of I34 in translation elongation has not been established so far. Previous reports suggest that I34 may improve the accuracy of codon-anticodon recognition, as A34 was shown to be unable to discriminate between the four bases (Agris et al., 2018; Borén et al., 1993; Lim and Curran, 2001), as well as to promote the stabilization of the contiguous A-site codon-anticodon interaction when I34 is located at the P-site (Lim, 1995), thereby affecting translation efficiency. In support of a role for I34 in translational efficiency is also the observation that emergence of heterodimeric ADAT in eukaryotes correlates with an increase in the number of genes that encode proteins enriched in TAPSLIVR residues, which have a codon usage biased towards triplets read by I34-modified tRNAs (Rafels-Ybern et al., 2015, 2018). In particular, genes exhibiting the highest preference for ADAT-sensitive codons were found to contain extremely biased stretches of the amino acids TAPS (Rafels-Ybern et al., 2015), and to encode a significant number of protein components of the extracellular matrix (ECM) (**Figure 1.16**). Thus, we hypothesized that the translation of these proteins could be dependent on the inosine modification.

To better understand the contributions of I34 to gene translation, we generated cell lines silenced for the ADAT2 catalytic subunit (**Figure 4.14**), which results in decreased I34 levels (Torres et al., 2015). Depletion of ADAT2 lead to impaired growth in several cell lines (**Figure 4.15**), a phenotype that may be related to the global delay in cell cycle progression observed in these cells. Interestingly, ADAT2 KD cells also displayed weaker attachment to culture surfaces than control cells (data not shown). While it has been suggested that I34 may play a role in translation of cell cycle-regulated genes (Frenkel-Morgenstern et al., 2012), it is also known that defects in the synthesis of ECM proteins can have a negative impact on cellular proliferation (Bonnans et al., 2014; Hynes, 2009). Indeed, we observed alterations in the synthesis of the ECM components dystroglycan (**Figure 4.28**) and syndecan-3 (**Figure 4.29**), whose protein levels were reduced in ADAT2 depleted cells compared to control cells, but their transcript levels remained unaltered between conditions. A more complicated scenario occurred for MUC1 and MUC5AC in differentiated colorectal cancer cells, where variations in their protein levels were accompanied by a differentiation-independent upregulation of their transcript levels in ADAT2 KD cells (**Figure 4.27**). It is possible that these transcriptional changes might constitute a compensation mechanism for defective protein synthesis (El-Brolosy and Stainier, 2017).

Despite the impact of the silencing of ADAT2 in the synthesis of these ECM constituents, we did not detect activation of the unfolded protein response (**Figure 4.16**), which is initiated upon sensing of unfolded or misfolded proteins in the ER that can arise from altered translation. Likewise, ADAT2 KD cells showed patterns of insoluble proteins similar to those of control cells (**Figure 4.27**), suggesting that the changes in protein levels may be either due to inefficient translation or to variations in protein turnover. Because many of the proteins enriched in ADAT-stretches are post-translationally glycosylated, and changes in glycosylation have been shown to alter protein stability (Gavrilov et al., 2015), we analyzed the O-glycan composition of purified mucins from ADAT2 KD and control cells. No major changes in glycosylation patterns were observed between the two conditions, with only minor differences in glycan sialylation being noted (**Figure 4.31**), which suggests that the depletion of ADAT2 does not affect the post-translational modification processing of proteins enriched in ADAT-stretches. If, alternatively, the translational efficiency of these proteins is altered by a reduction of I34 levels, the effects might be presumably weak, as cells are able to tolerate them relatively well without activating stress responses.

The fact that no significant differences in decoding rates were observed between control and ADAT2 KD cells by ribosome profiling analysis (**Figure 4.21 – Figure 4.26**) may further support this idea. However, this could also be related to a combination of factors, both technical and physiological. On the one hand, a sample size of only two biological replicates per condition, sequencing depth that might not have been high enough to provide good coverage of transcripts, and the use of cycloheximide to avoid ribosome run-off during sample preparation might have been amongst the technical features that could limit the sensitivity of the method to detect variations in ribosome occupancy at the sub-codon level, as it has been later described (Artieri and Fraser, 2014; Diamant and Tuller, 2016; Hussmann et al., 2015; O'Connor et al., 2016; Santos et al., 2019; Weinberg et al., 2016). Additionally, a major factor that also needs to be considered is that our knockdown cellular model leads to a reduction of the pool of tRNAs bearing I34, but it does not completely abrogate it. It is therefore plausible that the remaining modified pool might be sufficient to carry out translation elongation without significant consequences for the maintenance of global translation.

Two alternative strategies could be adopted in order to overcome this last constraint. One could either further decrease the levels of ADAT2 leading to a smaller supply of I34-modified tRNAs, or increase the cellular demand for these tRNAs in such a way that it widens the imbalance between supply and demand. While the knockout of ADAT is lethal in a range of organisms including humans (Gerber and Keller, 1999; Liu et al., 2016; Rubio et al., 2007; Torres et al., 2015; Tsutsumi et al., 2007; Zhou et al., 2014), and therefore its generation is not possible, recently developed methods that allow rapid and targeted degradation of

endogenous proteins in mammalian cells could be applied to this context (Röth et al., 2019). Due to their fast action, these methods minimize the risk that phenotypes arising from the disruption of target proteins are masked by compensatory mechanisms, and that secondary non-specific defects accumulate over time as a result of off-target effects (Damke et al., 1995; El-Brolosy and Stainier, 2017; Rossi et al., 2015). However, most targeted proteolytic systems require either fusing the protein of interest to a polypeptide tag, which can compromise the structural integrity and biological activity of the targeted protein itself, or having available a highly specific antibody against that same protein. Because of the drawbacks inherent to these procedures and their time-consuming development, we opted for the second option and sought a cellular scenario that would yield a rapid and large demand for I34-modified tRNAs.

MUC5AC is the human gene with the longest stretch of ADAT-sensitive codons. Therefore, we hypothesized that a situation that entailed expression of large amounts of this protein was likely to be affected by the silencing of ADAT2. A common pathway by which many stimuli, including oxidative stress, cigarette smoke, and several cytokines, induce MUC5AC expression involves the activation of epithelial growth factor receptor (EGFR), which triggers a signal transduction cascade that eventually culminates with enhanced transcription of *MUC5AC* (Burgel and Nadel, 2008; Lemjabbar-Alaoui et al., 2011; Lemjabbar et al., 2003). This is particularly relevant in asthmatic patients, who tend to present increased levels of EGFR expression in their airways that correlate with disease severity (Fedorov et al., 2005; Polosa et al., 2002; Puddicombe et al., 2000). Amphiregulin is an EGFR activator that is also found at high levels in sputum samples of asthmatic individuals (Enomoto et al., 2009; Kim et al., 2009). Using an *in vitro* cellular model of airway remodeling based on amphiregulin treatment of NCI-H292 lung cells, we have shown that I34 is required for the efficient translation of MUC5AC. Measurement of MUC5AC protein expression by immunofluorescence (**Figure 4.33**) and by flow cytometry-based assays (**Figure 4.34**) revealed a consistent reduction of ~ 3-fold in ADAT2 KD cells relative to control cells. This decrease in protein output was not due to transcriptional variations of *MUC5AC* mRNA, as both control and ADAT2 KD cells showed similar levels of transcript not only for *MUC5AC* but also for several mucous differentiation markers (**Figure 4.35**). Our results indicate that ADAT2 silenced cells properly activate the amphiregulin-induced transcriptional program, but are unable to efficiently synthesize MUC5AC polypeptides.

Of note, we detected an increase in transcript and protein levels of ADAT2 in differentiated cells that had been treated with AREG (**Figure 4.35**). This observation contrasts with a recent report which claimed that differentiation of human embryonic stem cells (hESC) is accompanied by a reduction in the levels of ADAT2 protein and mRNA,

together with a decrease in the I34 modification (Bornelöv et al., 2019). The authors performed ribosome profiling and argued that the depletion of ADAT2 explained the higher A-site ribosome occupancy of NNC and NNT ADAT-sensitive codons shown for differentiated cells versus self-proliferating cells, and concluded that the activity of ADAT contributes to optimized codon usage in self-renewing hESC (Bornelöv et al., 2019). However, the study presents a few flaws. On the one hand, the claim that I34 levels are reduced in differentiated hESC cannot be sustained by the tRNA-seq data, as the largest differences in modification status were observed for tRNA genes with sequencing coverage as low as < 12 reads/gene. And on the other hand, the ribosome occupancy of NNG codons specifying ADAT amino acids, which should be insensitive to the presence or absence of I34, is not compared to that of NNC/T codons. It is therefore unclear whether the activity of ADAT does indeed play a role in maintaining cellular pluripotency.

Our observation of the increased ADAT2 levels in AREG-treated cells could suggest the opposite, i.e. that the activity of ADAT is higher in differentiated cells, or it could indicate that cells are able to modulate their translational machinery in order to meet transcriptional demands. Regulation of mucin genes at the mRNA level, either through transcriptional upregulation or by means of mRNA stabilization, has been widely reported (Rose and Voinow, 2006). However, to our knowledge, no mechanisms for the translational regulation of mucins have been described to date. Here, we propose that changes in ADAT activity and, consequently, in the levels of I34 can modulate the synthesis of, at least, MUC5AC. It is possible that this effect may be extended to other proteins encoding ADAT-stretches, such as MUC5B. Overexpression of MUC5AC and MUC5B in the airway epithelia is associated with asthma and chronic obstructive pulmonary disease (Bonser et al., 2016; Kesimer et al., 2017; Kirkham et al., 2008). Our results could therefore have therapeutic implications, as fine-tuning of ADAT activity might be useful to inhibit MUC5AC and MUC5B hyperproduction.

In conclusion, our results suggest that a reduction of the I34-modified tRNA pool does not seriously impact the translation of human transcripts in a global manner, but it does lead to impaired protein synthesis of ECM genes enriched in stretches of ADAT-sensitive codons. We cannot rule out at this point that depletion of I34-tRNAs might cause minor defects in general translation which we cannot detect with currently available methods due to high background noise. Because some of the genes that we found to be affected by the lack of I34, such as *SDC3* or *DAG1*, play important roles in neurological development, it would be of great interest to examine their expression in patients carrying the ADAT3 V144M mutation. Furthermore, the use of animal models carrying the ADAT3 mutation could help us better understand the role of ADAT in neurodevelopment.

At the end of this dissertation, many questions remain to be answered with regards to the organization of the genetic code, the intricate evolutionary relationship between the code and the translation machinery, and the mechanisms that different organisms have developed to carry out protein synthesis. In the particular case of I34, some are: Why did eukaryotes adopt a G34-sparing strategy and expanded their A34 substrate repertoire instead of following the same tRNAome structure of bacteria? Did the emergence of heterodimeric ADAT and its role in the synthesis of ECM components contribute to the evolution of multicellularity? Does human ADAT have additional functions other than its tRNA deamination activity?

Without direct experimental evidence that would shed light into these matters, we can only hypothesize what the answers might be. The development of optimal cytoplasmic decoding solutions for a universal genetic code has been shown to be diverse among different organisms (Grosjean et al., 2010; Marck and Grosjean, 2002). One major divergence lays in the mechanisms co-opted by Bacteria and Eukarya, which developed tRNA anticodon A34- and G34-sparing strategies, respectively (Maraia and Arimbasseri, 2017; Marck and Grosjean, 2002). This has been linked to the appearance of tRNA modification enzymes uridine methyltransferases in Bacteria and heterodimeric ADAT in Eukarya, which are thought to have driven the genomic enrichment of kingdom-specific new tRNAs that improved translation (Novoa et al., 2012). The evolutionary selection of these strategies would reflect their efficiency within their particular contexts. However, despite providing some advantageous innovations to the translation machinery, they may also present limitations. One such limitation could be a reduced ability of most bacterial organisms to synthesize proteins displaying certain amino acid biases. This is yet to be tested, but the fact that the abundance of genes encoding stretches of TAPSLIVR residues is significantly higher in eukaryotic genomes than in bacterial ones, and that their codon composition is biased towards the use of ADAT-sensitive codons, whereas such bias is absent in bacteria, suggests that the I34 modification provided a functional adaptation to efficiently synthesize proteins enriched in TAPSLIVR stretches (Rafels-Ybern et al., 2018, 2019).

Two important traits of proteins that encode ADAT-stretches, specially of the mucin-like domain overrepresented among them, are: (1) their enrichment in intrinsically disordered protein regions (**Figure 5.1**) (van der Lee et al., 2014) and (2) their significant contribution to the ECM proteome. Intrinsically disordered protein regions have been shown to be necessary for the formation of mucinous gel-like networks and, together with the post-translational modifications that occur at these locations, they are also often involved in protein interactions that affect cell signaling (van der Lee et al., 2014). Intriguingly, the number of intrinsically disordered proteins increases with organism complexity (Ward et al., 2004; Yruela et al., 2017). An average of 2% of archaeal, 4% of bacterial, and 33% of eukaryotic proteins have

been predicted to contain disordered regions of more than 30 residues in length (Ward et al., 2004). It is therefore possible that, rather than promoting certain protein folds, decoding with the I34 modification somehow provided a mechanism to allow just the opposite, i.e. the formation of intrinsically disordered protein regions.

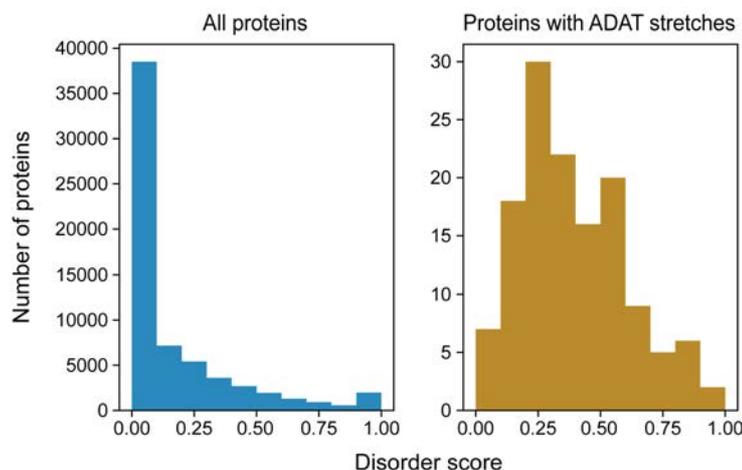


Figure 5.1. Distribution of the human proteins according to intrinsic disorder score. The distributions of the human proteome (left panel) and of those human proteins identified to encode stretches of ADAT-sensitive codons (right panel) are plotted. Data was obtained from the MobiDB database (Piovesan et al., 2018).

Furthermore, the enrichment of proteins containing ADAT-stretches in ECM components could point towards a relationship between the emergence of heterodimeric ADAT, the expansion of the I34 substrate repertoire, and the establishment of multicellularity. Multicellularity evolved in many eukaryotic lineages allowing the appearance of organisms of larger size, as well as providing possibilities for cellular specialization, formation of differentiated tissues and separation between germ and somatic cells (Rokas, 2008). A critical step for the transition to multicellularity was the evolution of mechanisms for stable cell adhesion and the emergence of the ECM on the metazoan stem clade. Because many of the domains found in ECM proteins have also been identified in unicellular eukaryotic organisms, it has been proposed that it was the rearrangement of preexistent domains what lead to the evolution of metazoan ECM proteins (King et al., 2008; Lang et al., 2016). Interestingly, Metazoans are the organisms with the longest ADAT-stretches and the highest enrichment in ADAT-sensitive codons (Rafels-Ybern et al., 2018). While correlations cannot be taken as evidence of causation, and further experimental work is required, these observations might suggest that the emergence of heterodimeric ADAT allowed the synthesis of longer proteins containing domains rich in TAPSLIVR residues, thus contributing to the evolution of multicellularity.

Regarding potential alternative functions of human ADAT, there is currently little evidence that would point in that direction, besides the above-mentioned role of *T. brucei* ADAT in DNA C-to-U editing. However, recent research may provide novel insight into that matter. Several reports have identified A-to-I editing events in nuclear encoded transcripts of filamentous fungi that occur during sexual reproduction (Liu et al., 2016, 2017; Teichert et al., 2017). This observation would not be particularly surprising if it were not because fungi, like plants, do not encode homologs of the metazoan enzymes Adenosine Deaminases Acting on mRNA (ADARs) (Wang et al., 2016). In contrast to metazoan mRNA A-to-I editing, which shows a preference for double-stranded RNA regions (Bass, 2002), fungal editing tends to target adenosines located in structurally stable hairpin loops (Liu et al., 2016). Additionally, analysis of the sequences flanking A-to-I editing sites in fungi revealed a strong preference for U at the 5' -1 site and for A and G at the 3' +1 site (Liu et al., 2016), which differs from the sequence specificity of human ADARs (Eggington et al., 2011), but resembles that of heterodimeric ADATs (Auxilien et al., 1996; Haumont et al., 1984; Saint-Léger et al., 2016). Together with the recently described A-to-I editing activity of bacterial TadA in mRNA hairpin loops (Bar-Yaacov et al., 2017), these findings raise the question whether this enzymatic activity may be conserved in eukaryotic ADATs. Intriguingly, the expression of *TAD2* and *TAD3* orthologs was found to be upregulated during sexual development in *Neurospora crassa* (Liu et al., 2017). If an mRNA editing activity for heterodimeric ADAT was experimentally confirmed, it could offer a new perspective into the pathogenic role of the ADAT3 V144M mutation, as many of the A-to-I mRNA editing events have been reported to occur in the nervous system (Rosenthal, 2015).

In summary, here we set out to investigate the role of the I34 tRNA modification in protein synthesis by examining the effect of a pathological mutation in the ADAT3 subunit, and by studying how the depletion of the modification impacts the synthesis of proteins encoded by stretches of ADAT-sensitive codons. We found that the ADAT3 V144M mutation leads to reduced catalytic activity of the enzyme. This decrease is likely influenced by the destabilizing effect of the mutation on the tertiary structure of ADAT3, and by the altered subcellular localization of mutant ADAT3. The fact that multiple genetic alterations have been recently described for ADAT3 in association with neuropathies suggests that the I34 modification might play an important role in the translation of genes required for a normal development. In this regard, we investigated the impact of the I34 modification on protein synthesis by silencing ADAT2 in several cell lines. We showed that depletion of I34 results in slower cellular growth without affecting general protein synthesis or causing ER stress. Importantly, we revealed that silencing of ADAT2 impairs the translation of proteins of the extracellular matrix that contain ADAT-stretches. Of the ECM components found to be

affected by the depletion of I34, syndecan-3 and dystroglycan are known to have relevant roles in neurological development, and the expression of mucins have been linked to several respiratory and oncologic disorders. Thus, it would be of great interest to elucidate whether a modulation of the levels of ADAT could have a positive impact in the output of these pathologies.

6. CONCLUSIONS

The ADAT3 V144M mutation destabilizes the tertiary structure of ADAT3 and, in *E. coli*, it induces proteolytic cleavage of ADAT3. However, it does not affect the formation of the ADAT heterodimer nor its affinity for tRNA.

Mutant ADAT exhibits a 10-fold reduction in catalytic efficiency compared to the wild type enzyme.

Human ADAT is predominantly a nuclear enzyme, and ADAT3 translocates to nucleus in an ADAT2-dependent manner.

The V144M mutation causes subcellular delocalization of ADAT3 *in vitro* by induction of a partial nuclear import that is independent of ADAT2, but it does not alter its interaction with ADAT2.

ADAT2 silencing impairs cellular proliferation in several human cell lines.

ADAT2 deficiency does not lead to protein aggregation, nor does it activate the unfolded protein response.

ADAT2 depletion does not result in variations of codon decoding rates detectable by ribosome profiling.

Deficiency of ADAT2 leads to lower protein expression of the ECM components syndecan-3 and dystroglycan.

ADAT2 silencing alters the expression of mucins MUC1 and MUC5AC at the transcriptional and translational levels in colorectal cancer cells, but it does not affect mucin glycosylation.

ADAT2 is required for the efficient translation of MUC5AC in an *in vitro* cellular model of airway remodeling.

7. REFERENCES

- Abedin, M., and King, N. (2010). Diverse evolutionary paths to cell adhesion. *Trends Cell Biol.* 20, 734–742.
- Achsel, T., and Gross, H.J. (1993). Identity determinants of human tRNA(Ser): sequence elements necessary for serylation and maturation of a tRNA with a long extra arm. *EMBO J.* 12, 3333–3338.
- Afratis, N.A., Nikitovic, D., Multhaupt, H.A.B., Theocharis, A.D., Couchman, J.R., and Karamanos, N.K. (2017). Syndecans – key regulators of cell signaling and biological functions. *FEBS J.* 284, 27–41.
- Agris, P.F. (1991). Wobble position modified nucleosides evolved to select transfer RNA codon recognition: a modified-wobble hypothesis. *Biochimie* 73, 1345–1349.
- Agris, P.F., Eruysal, E.R., Narendran, A., Väre, V.Y.P., Vangaveti, S., and Ranganathan, S. V. (2018). Celebrating wobble decoding: Half a century and still much is new. *RNA Biol.* 15, 537–553.
- Akashi, H. (1994). Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* 136, 927–935.
- Akhavan, A., Crivelli, S.N., Singh, M., Lingappa, V.R., and Muschler, J.L. (2007). SEA domain proteolysis determines the functional composition of dystroglycan. *FASEB J.* 22, 612–621.
- Alazami, A.M., Hijazi, H., Al-Dosari, M.S., Shaheen, R., Hashem, A., Aldahmesh, M.A., Mohamed, J.Y., Kentab, A., Salih, M.A., Awaji, A., et al. (2013). Mutation in ADAT3 , encoding adenosine deaminase acting on transfer RNA, causes intellectual disability and strabismus. *J. Med. Genet.* 50, 425–430.
- Alexandrov, A., Chernyakov, I., Gu, W., Hiley, S.L., Hughes, T.R., Grayhack, E.J., and Phizicky, E.M. (2006). Rapid tRNA decay can result from lack of nonessential modifications. *Mol. Cell* 21, 87–96.
- Algire, M.A., Maag, D., and Lorsch, J.R. (2005). Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Mol. Cell* 20, 251–262.
- Alkalaeva, E.Z., Pisarev, A. V, Frolova, L.Y., Kisselev, L.L., and Pestova, T. V (2006). In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. *Cell* 125, 1125–1136.
- Anger, A.M., Armache, J.P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D.N., and Beckmann, R. (2013). Structures of the human and *Drosophila* 80S ribosome. *Nature* 497, 80–85.
- Archer, E.J., Simpson, M.A., Watts, N.J., O’Kane, R., Wang, B., Erie, D.A., McPherson, A., and Weeks, K.M. (2013). Long-range architecture in a viral RNA genome. *Biochemistry* 52, 3182–3190.
- Arimbasseri, A.G., Blewett, N.H., Iben, J.R., Lamichhane, T.N., Cherkasova, V., Hafner, M., and Maraia, R.J. (2015). RNA Polymerase III Output Is Functionally Linked to tRNA Dimethyl-G26 Modification. *PLOS Genet.* 11, e1005671.
- Arthur, L.L., Pavlovic-Djuranovic, S., Koutmou, K.S., Green, R., Szczesny, P., and Djuranovic, S. (2015). Translational control by lysine-encoding A-rich sequences. *Sci. Adv.* 1, e1500154.
- Artieri, C.G., and Fraser, H.B. (2014). Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. *Genome Res.* 24, 2011–2021.
- Arts, G.J., Kuersten, S., Romby, P., Ehresmann, B., and Mattaj, I.W. (1998). The role of exportin-t in selective nuclear export of mature tRNAs. *EMBO J.* 17, 7430–7441.

- Atkins, J.F., Loughran, G., Bhatt, P.R., Firth, A.E., and Baranov, P. V. (2016). Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* *44*, 7007–7078.
- Atkinson, G.C., Baldauf, S.L., and Hauryliuk, V. (2008). Evolution of nonstop, no-go and nonsense-mediated mRNA decay and their termination factor-derived components. *BMC Evol. Biol.* *8*, 1–18.
- Auxilien, S., Crain, P.F., Trewyn, R.W., and Grosjean, H. (1996). Mechanism, specificity and general properties of the yeast enzyme catalysing the formation of inosine 34 in the anticodon of transfer RNA. *J. Mol. Biol.* *262*, 437–458.
- Bansil, R., and Turner, B.S. (2006). Mucin structure, aggregation, physiological functions and biomedical applications. *Curr. Opin. Colloid Interface Sci.* *11*, 164–170.
- Bansil, R., and Turner, B.S. (2018). The biology of mucus: Composition, synthesis and organization. *Adv. Drug Deliv. Rev.* *124*, 3–15.
- Bar-Yaacov, D., Mordret, E., Towers, R., Biniashvili, T., Soyris, C., Schwartz, S., Dahan, O., and Pilpel, Y. (2017). RNA editing in bacteria recodes multiple proteins and regulates an evolutionarily conserved toxin-antitoxin system. *Genome Res.* *27*, 1696–1703.
- Barresi, R., and Campbell, K.P. (2006). Dystroglycan: from biosynthesis to pathogenesis of human disease. *J. Cell Sci.* *119*, 199–207.
- Bass, B.L. (2002). RNA Editing by Adenosine Deaminases That Act on RNA. *Annu. Rev. Biochem.* *71*, 817–846.
- Beaudoin, J.-D., Novoa, E.M., Vejnar, C.E., Yartseva, V., Takacs, C.M., Kellis, M., and Giraldez, A.J. (2018). Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat. Struct. Mol. Biol.* *25*, 677–686.
- Begley, U., Dyavaiah, M., Patil, A., Rooney, J.P., DiRenzo, D., Young, C.M., Conklin, D.S., Zitomer, R.S., and Begley, T.J. (2007). Trm9-Catalyzed tRNA Modifications Link Translation to the DNA Damage Response. *Mol. Cell* *28*, 860–870.
- Behrmann, E., Loerke, J., Budkevich, T. V., Yamamoto, K., Schmidt, A., Penczek, P.A., Vos, M.R., Bürger, J., Mielke, T., Scheerer, P., et al. (2015). Structural snapshots of actively translating human ribosomes. *Cell* *161*, 845–857.
- Beringer, M., and Rodnina, M. V. (2007). The ribosomal peptidyl transferase. *Mol. Cell* *26*, 311–321.
- Berrow, N.S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D.I., and Owens, R.J. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* *35*, e45–e45.
- Bertram, G., Bell, H.A., Ritchie, D.W., Fullerton, G., and Stansfield, I. (2000). Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon recognition. *RNA* *6*, 1236–1247.
- Bespalov, M.M., Sidorova, Y.A., Tumova, S., Ahonen-Bishopp, A., Magalhães, A.C., Kuleskiy, E., Paveliev, M., Rivera, C., Rauvala, H., and Saarma, M. (2011). Heparan sulfate proteoglycan syndecan-3 is a novel receptor for GDNF, neurturin, and artemin. *J. Cell Biol.* *192*, 153–169.
- Bonnans, C., Chou, J., and Werb, Z. (2014). Remodelling the extracellular matrix in development and disease. *Nat. Rev. Mol. Cell Biol.* *15*, 786–801.
- Bonser, L.R., Zlock, L., Finkbeiner, W., and Erle, D.J. (2016). Epithelial tethering of MUC5AC-rich mucus impairs mucociliary transport in asthma. *J. Clin. Invest.* *126*, 2367–2371.
- Borén, T., Elias, P., Samuelsson, T., Claesson, C., Barciszewska, M., Gehrke, C.W., Kuo,

- K.C., and Lustig, F. (1993). Undiscriminating Codon Reading with Adenosine in the Wobble Position. *J. Mol. Biol.* *230*, 739–749.
- Bornelöv, S., Selmi, T., Flad, S., Dietmann, S., and Frye, M. (2019). Codon usage optimization in pluripotent embryonic stem cells. *Genome Biol.* *20*, 119.
- Braat, A.K., Yan, N., Am, E., Harrison, D., and Macdonald, P.M. (2004). Localization-dependent Oskar protein accumulation control after the initiation of translation. *Dev. Cell* *7*, 125–131.
- Brenner, S., Jacob, F., and Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* *190*, 576–581.
- Briggs, D.C., Yoshida-Moriguchi, T., Zheng, T., Venzke, D., Anderson, M.E., Strazzulli, A., Moracci, M., Yu, L., Hohenester, E., and Campbell, K.P. (2016). Structural basis of laminin binding to the LARGE glycans on dystroglycan. *Nat. Chem. Biol.* *12*, 810–814.
- Brown, A., Shao, S., Murray, J., Hegde, R.S., and Ramakrishnan, V. (2015). Structural basis for stop codon recognition in eukaryotes. *Nature* *524*, 493–496.
- Brule, C.E., and Grayhack, E.J. (2017). Synonymous codons: Choose wisely for expression. *Trends Genet.* *33*, 283–297.
- Buchan, J.R., Aucott, L.S., and Stansfield, I. (2006). tRNA properties help shape codon pair preferences in open reading frames. *Nucleic Acids Res.* *34*, 1015–1027.
- Budkevich, T., Giesebrecht, J., Altman, R.B., Munro, J.B., Mielke, T., Nierhaus, K.H., Blanchard, S.C., and Spahn, C.M.T. (2011). Structure and dynamics of the mammalian ribosomal pretranslocation complex. *Mol. Cell* *44*, 214–224.
- Buhr, F., Jha, S., Thommen, M., Mittelstaet, J., Kutz, F., Schwalbe, H., Rodnina, M. V., and Komar, A.A. (2016). Synonymous codons direct cotranslational folding toward different protein conformations. *Mol. Cell* *61*, 341–351.
- Burgel, P.-R., and Nadel, J.A. (2008). Epidermal growth factor receptor-mediated innate immune responses and their roles in airway diseases. *Eur. Respir. J.* *32*, 1068–1081.
- Burkhardt, D.H., Rouskin, S., Zhang, Y., Li, G.-W., Weissman, J.S., and Gross, C.A. (2017). Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. *Elife* *6*, 1–23.
- Del Campo, C., Bartholomäus, A., Fedyunin, I., and Ignatova, Z. (2015). Secondary structure across the bacterial transcriptome reveals versatile roles in mRNA regulation and function. *PLoS Genet.* *11*, 1–23.
- Cannarrozzi, G., Schraudolph, N.N., Faty, M., von Rohr, P., Friberg, M.T., Roth, A.C., Gonnet, P., Gonnet, G., and Barral, Y. (2010). A role for codon order in translation dynamics. *Cell* *141*, 355–367.
- Cantara, W.A., Crain, P.F., Rozenski, J., McCloskey, J.A., Harris, K.A., Zhang, X., Vendeix, F.A.P., Fabris, D., and Agris, P.F. (2011). The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.* *39*, 195–201.
- Chakravarti, R., and Adams, J.C. (2006). Comparative genomics of the syndecans defines an ancestral genomic context associated with matrilins in vertebrates. *BMC Genomics* *7*, 1–16.
- Chan, P.P., and Lowe, T.M. (2016). GtRNADB 2.0: An expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* *44*, D184–D189.
- Chan, C., Pham, P., Dedon, P.C., and Begley, T.J. (2018). Lifestyle modifications: coordinating the tRNA epitranscriptome with codon bias to adapt translation during stress responses. *Genome Biol.* *19*, 228.
- Chan, C.T.Y., Pang, Y.L.J., Deng, W., Babu, I.R., Dyavaiah, M., Begley, T.J., and Dedon,

- P.C. (2012). Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nat. Commun.* 3, 937.
- Chaney, J.L., and Clark, P.L. (2015). Roles for synonymous codon usage in protein biogenesis. *Annu. Rev. Biophys.* 44, 143–166.
- Chaney, J.L., Steele, A., Carmichael, R., Rodriguez, A., Specht, A.T., Ngo, K., Li, J., Emrich, S., and Clark, P.L. (2017). Widespread position-specific conservation of synonymous rare codons within coding sequences. *PLoS Comput. Biol.* 13, 1–19.
- Charneski, C.A., and Hurst, L.D. (2013). Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol.* 11, e1001508.
- Chatterjee, K., Nostramo, R.T., Wan, Y., and Hopper, A.K. (2018). tRNA dynamics between the nucleus, cytoplasm and mitochondrial surface: Location, location, location. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1861, 373–386.
- Chen, G., Korfhagen, T.R., Xu, Y., Kitzmiller, J., Wert, S.E., Maeda, Y., Gregorieff, A., Clevers, H., and Whitsett, J.A. (2009). SPDEF is required for mouse pulmonary goblet cell differentiation and regulates a network of genes associated with mucus production. *J. Clin. Invest.* 119, 2914–2924.
- Chou, H.-J., Donnard, E., Gustafsson, H.T., Garber, M., and Rando, O.J. (2017). Transcriptome-wide analysis of roles for tRNA modifications in translational regulation. *Mol. Cell* 68, 978–992.e4.
- Chu, D., Kazana, E., Bellanger, N., Singh, T., Tuite, M.F., and von der Haar, T. (2014). Translation elongation can control translation initiation on eukaryotic mRNAs. *EMBO J.* 33, 21–34.
- Chugh, S., Gnanapragassam, V.S., Jain, M., Rachagani, S., Ponnusamy, M.P., and Batra, S.K. (2015). Pathobiological implications of mucin glycans in cancer: Sweet poison and novel targets. *Biochim. Biophys. Acta - Rev. Cancer* 1856, 211–225.
- Clarke, T.F., and Clark, P.L. (2008). Rare codons cluster. *PLoS One* 3, e3412.
- Corfield, A.P. (2015). Mucins: A biologically relevant glycan barrier in mucosal protection. *Biochim. Biophys. Acta - Gen. Subj.* 1850, 236–252.
- Cormand, B., Pihko, H., Bayes, M., Valanne, L., Santavuori, P., Talim, B., Gershoni-Baruch, R., Ahmad, A., van Bokhoven, H., Brunner, H.G., et al. (2001). Clinical and genetic distinction between Walker-Warburg syndrome and muscle-eye-brain disease. *Neurology* 56, 1059–1069.
- Couchman, J.R., Gopal, S., Lim, H.C., Nørgaard, S., and Multhaupt, H.A.B. (2015). Syndecans: From peripheral coreceptors to mainstream regulators of cell behaviour. *Int. J. Exp. Pathol.* 96, 1–10.
- de Crécy-Lagard, V., Boccaletto, P., Mangleburg, C.G., Sharma, P., Lowe, T.M., Leidel, S.A., and Bujnicki, J.M. (2019). Matching tRNA modifications in humans to their known and predicted enzymes. *Nucleic Acids Res.* 47, 2143–2159.
- Crick, F.H.C. (1955). On degenerate templates and the adaptor hypothesis: A note for the RNA tie club. *Wellcome Libr. Hist. Underst. Med.*
- Crick, F.H.C. (1956). Ideas on protein synthesis (Oct. 1956). *Wellcome Libr. Hist. Underst. Med.*
- Crick, F.H.C. (1958). On protein synthesis. *Symp. Soc. Exp. Biol.* 12, 138–163.
- Crick, F.H.C. (1966). Codon—anticodon pairing: The wobble hypothesis. *J. Mol. Biol.* 19, 548–555.
- Crick, F.H.C. (1968). The origin of the genetic code. *J. Mol. Biol.* 38, 367–379.

- Crick, F.H.C. (1970). Central dogma of molecular biology. *Nature* 227, 561–563.
- Crick, F.H.C., Barnett, L., Brenner, S., and Watts-tobin, R.J. (1961). General nature of the genetic code for proteins. *Nature* 192, 1227–1232.
- Crombie, T., Swaffield, J.C., and Brown, A.J.P. (1992). Protein folding within the cell is influenced by controlled rates of polypeptide elongation. *J. Mol. Biol.* 228, 7–12.
- Damke, H., Baba, T., Van Der Blik, A.M., and Schmid, S.L. (1995). Clathrin-independent pinocytosis is induced in cells overexpressing a temperature-sensitive mutant of dynamin. *J. Cell Biol.* 131, 69–80.
- Dana, A., and Tuller, T. (2012). Determinants of translation elongation speed and ribosomal profiling biases in mouse embryonic stem cells. *PLoS Comput. Biol.* 8, e1002755.
- Dana, A., and Tuller, T. (2014). The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res.* 42, 9171–9181.
- Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y.S., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 146, 247–261.
- Davies, J.R., Wickström, C., and Thornton, D.J. (2012). Gel-forming and cell-associated mucins: preparation for structural and functional studies. *Methods Mol. Biol.* 842, 27–47.
- Dekker, J., Rossen, J.W.A., Büller, H.A., and Einerhand, A.W.C. (2002). The MUC family: An obituary. *Trends Biochem. Sci.* 27, 126–131.
- Delannoy, E., Le Ret, M., Faivre-Nitschke, E., Estavillo, G.M., Bergdoll, M., Taylor, N.L., Pogson, B.J., Small, I., Imbault, P., and Gualberto, J.M. (2009). Arabidopsis tRNA adenosine deaminase arginine edits the wobble nucleotide of chloroplast tRNA^{Arg}(ACG) and is essential for efficient chloroplast translation. *Plant Cell* 21, 2058–2071.
- Delker, R.K., Zhou, Y., Strikoudis, A., Stebbins, C.E., and Papavasiliou, F.N. (2013). Solubility-based genetic screen identifies RING finger protein 126 as an E3 ligase for activation-induced cytidine deaminase. *Proc. Natl. Acad. Sci.* 110, 1029–1034.
- Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M., and Yusupova, G. (2012). A new understanding of the decoding principle on the ribosome. *Nature* 484, 256–259.
- Deng, H., Cheema, J., Zhang, H., Woolfenden, H., Norris, M., Liu, Z., Liu, Q., Yang, X., Yang, M., Deng, X., et al. (2018). Rice in vivo RNA structurome reveals RNA secondary structure conservation and divergence in plants. *Mol. Plant* 11, 607–622.
- Deng, W., Babu, I.R., Su, D., Yin, S., Begley, T.J., and Dedon, P.C. (2015). Trm9-catalyzed tRNA modifications regulate global protein expression by codon-biased translation. *PLoS Genet.* 11, e1005706.
- Dever, T.E., Dinman, J.D., and Green, R. (2018). Translation elongation and recoding in eukaryotes. *Cold Spring Harb. Perspect. Biol.* 10, 1–20.
- Diament, A., and Tuller, T. (2016). Estimation of ribosome profiling performance and reproducibility at various levels of resolution. *Biol. Direct* 11, 1–12.
- Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., and Assmann, S.M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505, 696–700.
- Dittmar, K.A., Sørensen, M.A., Elf, J., Ehrenberg, M., and Pan, T. (2005). Selective charging of tRNA isoacceptors induced by amino-acid starvation. *EMBO Rep.* 6, 151–157.
- Dittmar, K.A., Goodenbour, J.M., and Pan, T. (2006). Tissue-specific differences in human transfer RNA expression. *PLoS Genet.* 2, 2107–2115.
- Dix, D.B., and Thompson, R.C. (2006). Codon choice and gene expression: synonymous

- codons differ in translational accuracy. *Proc. Natl. Acad. Sci.* **86**, 6888–6892.
- Doerfel, L.K., Wohlgemuth, I., Kothe, C., Peske, F., Urlaub, H., and Rodnina, M. V (2013). EF-P Is essential for rapid synthesis. *Science* (80-.). **339**, 85–88.
- Dong, H., Nilsson, L., and Kurland, C.G. (1996). Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.* **260**, 649–663.
- Draper, G.W., Shoemark, D.K., and Adams, J.C. (2019). Modelling the early evolution of extracellular matrix from modern Ctenophores and Sponges. *Essays Biochem.* **63**, 389–405.
- Drummond, D.A., and Wilke, C.O. (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341–352.
- Durbeej, M., Henry, M.D., Ferletta, M., Campbell, K.P., and Ekblom, P. (1998). Distribution of dystroglycan in normal adult mouse tissues. *J. Histochem. Cytochem.* **46**, 449–457.
- Duret, L., and Mouchiroud, D. (1999). Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci.* **96**, 4482–4487.
- Eggington, J.M., Greene, T., and Bass, B.L. (2011). Predicting sites of ADAR editing in double-stranded RNA. *Nat. Commun.* **2**, 319.
- El-Brolosy, M.A., and Stainier, D.Y.R. (2017). Genetic compensation: A phenomenon in search of mechanisms. *PLoS Genet.* **13**, 1–17.
- El-Hattab, A.W., Saleh, M.A., Hashem, A., Al-Owain, M., Asmari, A. Al, Rabei, H., Abdelraouf, H., Hashem, M., Alazami, A.M., Patel, N., et al. (2016). ADAT3-related intellectual disability: Further delineation of the phenotype. *Am. J. Med. Genet. Part A* **170**, 1142–1147.
- Elf, J., Nilsson, D., Tenson, T., and Ehrenberg, M. (2003). Selective charging of tRNA isoacceptors explains patterns of codon usage. *Science* (80-.). **300**, 1718–1722.
- Elias, Y., and Huang, R.H. (2005). Biochemical and structural studies of A-to-I editing by tRNA:A34 deaminases at the wobble position of transfer RNA. *Biochemistry* **44**, 12057–12065.
- Enomoto, Y., Orihara, K., Takamasu, T., Matsuda, A., Gon, Y., Saito, H., Ra, C., and Okayama, Y. (2009). Tissue remodeling induced by hypersecreted epidermal growth factor and amphiregulin in the airway after an acute asthma attack. *J. Allergy Clin. Immunol.* **124**, 913-920.e7.
- Farabaugh, P.J. (1996). Programmed translational frameshifting. *Annu. Rev. Genet.* **30**, 507–528.
- Faure, G., Ogurtsov, A.Y., Shabalina, S.A., and Koonin, E. V. (2016). Role of mRNA structure in the control of protein folding. *Nucleic Acids Res.* **44**, 10898–10911.
- Faure, G., Ogurtsov, A.Y., Shabalina, S.A., and Koonin, E. V. (2017). Adaptation of mRNA structure to control protein folding. *RNA Biol.* **14**, 1649–1654.
- Fedorov, I.A., Wilson, S.J., Davies, D.E., and Holgate, S.T. (2005). Epithelial stress and structural remodelling in childhood asthma. *Thorax* **60**, 389–394.
- Ferguson, A., Wang, L., Altman, R.B., Terry, D.S., Juette, M.F., Burnett, B.J., Alejo, J.L., Dass, R.A., Parks, M.M., Vincent, C.T., et al. (2015). Functional dynamics within the human ribosome regulate the rate of active protein synthesis. *Mol. Cell* **60**, 475–486.
- Flis, J., Holm, M., Rundlet, E.J., Loerke, J., Hilal, T., Dabrowski, M., Bürger, J., Mielke, T., Blanchard, S.C., Spahn, C.M.T., et al. (2018). tRNA translocation by the eukaryotic 80S ribosome and the impact of GTP hydrolysis. *Cell Rep.* **25**, 2676-2688.e7.

- Fournier, M., Haumont, E., Henau, S. De, Gangloff, J., and Grosjean, H. (1983). Post-transcriptional modification of the wobble nucleotide in anticodon-substituted yeast tRNA^{II} Arg after microinjection into xenopus laevis oocytes. *Nucleic Acids Res.* 11, 707–718.
- Frantz, C., Stewart, K.M., and Weaver, V.M. (2010). The extracellular matrix at a glance. *J. Cell Sci.* 123, 4195–4200.
- Fraser, C.S., Berry, K.E., Hershey, J.W.B., and Doudna, J.A. (2007). eIF3j is located in the decoding center of the human 40S ribosomal subunit. *Mol. Cell* 26, 811–819.
- French, B.T., and Trewyn, R.W. (1990). Modification of the anticodon wobble position of tRNA(Ala) in vitro does not require 5' or 3' processing. *Gene* 96, 301–304.
- Frenkel-Morgenstern, M., Danon, T., Christian, T., Igarashi, T., Cohen, L., Hou, Y.-M., and Jensen, L.J. (2012). Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. *Mol. Syst. Biol.* 8, 1–10.
- Frolova, L., Le Goff, X., Rasmussen, H.H., Cheperegin, S., Drugeon, G., Kress, M., Arman, I., Haenni, A.L., Celis, J.E., Philippe, M., et al. (1994). A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. *Nature* 372, 701–703.
- Frolova, L., Le Goff, X., Zhouravleva, G., Davydova, E., Philippe, M., and Kisselev, L. (1996). Eukaryotic polypeptide chain release factor eRF3 is an eRF1- and ribosome-dependent guanosine triphosphatase. *RNA* 2, 334–341.
- Frolova, L.Y., Tsivkovskii, R.Y., Sivolobova, G.F., Oparina, N.Y., Serpinsky, O.I., Blinov, V.M., Tatkov, S.I., and Kisselev, L.L. (1999). Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA* 5, 1014–1020.
- Gamble, C.E., Brule, C.E., Dean, K.M., Fields, S., and Grayhack, E.J. (2016). Adjacent codons act in concert to modulate translation efficiency in yeast. *Cell* 166, 679–690.
- Gardin, J., Yeasmin, R., Yurovsky, A., Cai, Y., Skiena, S., and Futcher, B. (2014). Measurement of average decoding rates of the 61 sense codons in vivo. *Elife* 3, 1–20.
- Garel, J.-P. (1974). Functional adaptation of tRNA population. *J. Theor. Biol.* 43, 211–225.
- Garel, J.P., Hentzen, D., and Daillie, J. (1974). Codon responses of tRNA-Ala, tRNA-Gly and tRNA-Ser from the posterior part of the silk gland of Bombyx mori L. *FEBS Lett.* 39, 359–363.
- Gaston, K.W., Rubio, M.A.T., Spears, J.L., Pastar, I., Papavasiliou, F.N., and Alfonzo, J.D. (2007). C to U editing at position 32 of the anticodon loop precedes tRNA 5' leader removal in trypanosomatids. *Nucleic Acids Res.* 35, 6740–6749.
- Gavrilov, Y., Shental-Bechor, D., Greenblatt, H.M., and Levy, Y. (2015). Glycosylation May Reduce Protein Thermodynamic Stability by Inducing a Conformational Distortion. *J. Phys. Chem. Lett.* 6, 3572–3577.
- Des Georges, A., Dhote, V., Kuhn, L., Hellen, C.U.T., Pestova, T. V., Frank, J., and Hashem, Y. (2015). Structure of mammalian eIF3 in the context of the 43S preinitiation complex. *Nature* 525, 491–495.
- Gerber, A.P., and Keller, W. (1999). An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* (80-.). 286, 1146–1149.
- Geslain, R., Aeby, E., Guitart, T., Jones, T.E., De Moura, M.C., Charriè, F., Schneider, A., and De Poupplana, L.R. (2006). Trypanosoma seryl-tRNA synthetase is a metazoan-like enzyme with high affinity for tRNA^{Sec}. *J. Biol. Chem.* 281, 38217–38225.
- Ghoneim, D.H., Zhang, X., Brule, C.E., Mathews, D.H., and Grayhack, E.J. (2019). Conservation of location of several specific inhibitory codon pairs in the Saccharomyces

- sensu stricto yeasts reveals translational selection. *Nucleic Acids Res.* *47*, 1164–1177.
- Giegé, R., Jühling, F., Pütz, J., Stadler, P., Sauter, C., and Florentz, C. (2012). Structure of transfer RNAs: Similarity and variability. *Wiley Interdiscip. Rev. RNA* *3*, 37–61.
- Gingold, H., and Pilpel, Y. (2011). Determinants of translation efficiency and accuracy. *Mol. Syst. Biol.* *7*, 1–13.
- Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sørensen, K.D., et al. (2014). A dual program for translation regulation in cellular proliferation and differentiation. *Cell* *158*, 1281–1292.
- Godfrey, C., Foley, A.R., Clement, E., and Muntoni, F. (2011). Dystroglycanopathies: Coming into focus. *Curr. Opin. Genet. Dev.* *21*, 278–285.
- Gogakos, T., Brown, M., Garzia, A., Meyer, C., Hafner, M., and Tuschl, T. (2017). Characterizing expression and processing of precursor and mature human tRNAs by Hydro-tRNAseq and PAR-CLIP. *Cell Rep.* *20*, 1463–1475.
- Gondelaud, F., and Ricard-Blum, S. (2019). Structures and interactions of syndecans. *FEBS J.* 1–14.
- Goodarzi, H., Nguyen, H.C.B., Zhang, S., Dill, B.D., Molina, H., and Tavazoie, S.F. (2016). Modulated expression of specific tRNAs drives gene expression and cancer progression. *Cell* *165*, 1416–1427.
- Gott, J.M., and Emeson, R.B. (2000). Functions and mechanisms of RNA editing. *Annu. Rev. Genet.* *34*, 499–531.
- Gould, S.E., Upholt, W.B., and Kosher, R.A. (1992). Syndecan 3: a member of the syndecan family of membrane-intercalated proteoglycans that is expressed in high amounts at the onset of chicken limb cartilage differentiation. *Proc. Natl. Acad. Sci.* *89*, 3271–3275.
- Gouy, M., and Gautier, C. (1982). Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* *10*, 7055–7074.
- Gray, N.K., and Hentze, M.W. (1994). Regulation of protein synthesis by mRNA structure. *Mol. Biol. Rep.* *19*, 195–200.
- Gromadski, K.B., Daviter, T., and Rodnina, M. V. (2006). A uniform response to mismatches in codon-anticodon complexes ensures ribosomal fidelity. *Mol. Cell* *21*, 369–377.
- Grosjean, H., and Fiers, W. (1982). Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene* *18*, 199–209.
- Grosjean, H., Auxilien, S., Constantinesco, F., Simon, C., Corda, Y., Becker, H.F., Foiret, D., Morin, A., Jin, Y.X., Fournier, M., et al. (1996). Enzymatic conversion of adenosine to inosine and to N1-methylinosine in transfer RNAs: A review. *Biochimie* *78*, 488–501.
- Grosjean, H., de Crécy-Lagard, V., and Marck, C. (2010). Deciphering synonymous codons in the three domains of life: Co-evolution with specific tRNA modification enzymes. *FEBS Lett.* *584*, 252–264.
- Gutgsell, N., Englund, N., Niu, L., Kaya, Y., Lane, B.G., and Ofengand, J. (2000). Deletion of the *Escherichia coli* pseudouridine synthase gene *truB* blocks formation of pseudouridine 55 in tRNA in vivo, does not affect exponential growth, but confers a strong selective disadvantage in competition with wild-type cells. *RNA* *6*, 1870–1881.
- Gutierrez, E., Shin, B.-S., Woolstenhulme, C.J., Kim, J.-R., Saini, P., Buskirk, A.R., and Dever, T.E. (2013). eIF5A promotes translation of polyproline motifs. *Mol. Cell* *51*, 35–45.
- Gutman, G.A., and Hatfield, G.W. (1989). Nonrandom utilization of codon pairs in

- Escherichia coli*. *Proc. Natl. Acad. Sci.* **86**, 3699–3703.
- Guy, M.P., and Phizicky, E.M. (2014). Two-subunit enzymes involved in eukaryotic post-transcriptional tRNA modification. *RNA Biol.* **11**, 1608–1618.
- Gysi, S., Rhiner, C., Flibotte, S., Moerman, D.G., and Hengartner, M.O. (2013). A network of HSPG core proteins and HS modifying enzymes regulates netrin-dependent guidance of D-type motor neurons in *Caenorhabditis elegans*. *PLoS One* **8**, 1–15.
- Halvorsen, M., Martin, J.S., Broadaway, S., and Laederach, A. (2010). Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genet.* **6**, 1–11.
- Hanson, G., and Collier, J. (2017). Codon optimality, bias and usage in translation and mRNA decay. *Nat. Rev. Mol. Cell Biol.* **19**, 20–30.
- Hatrup, C.L., and Gendler, S.J. (2008). Structure and function of the cell surface (tethered) mucins. *Annu. Rev. Physiol.* **70**, 431–457.
- Haumont, E., Fournier, M., de Henau, S., and Grosjean, H. (1984). Enzymatic conversion of adenosine to inosine in the wobble position of yeast tRNA^{Asp}: the dependence on the anticodon sequence. *Nucleic Acids Res.* **12**, 2705–2715.
- Hawer, H., Hammermeister, A., Ravichandran, K.E., Glatt, S., Schaffrath, R., and Klassen, R. (2018). Roles of elongator dependent tRNA modification pathways in neurodegeneration and cancer. *Genes (Basel)*. **10**, 1–23.
- Heiss, M., Reichle, V.F., and Kellner, S. (2017). Observing the fate of tRNA and its modifications by nucleic acid isotope labeling mass spectrometry: NAIL-MS. *RNA Biol.* **14**, 1260–1268.
- Hellen, C.U.T. (2018). Translation termination and ribosome recycling in eukaryotes. *Cold Spring Harb. Perspect. Biol.* **10**, a032656.
- Hennebicq-Reig, S., Lesuffleur, T., Capon, C., De Bolos, C., Kim, I., Moreau, O., Richet, C., Hémon, B., Recchi, M.A., Maës, E., et al. (1998). Permanent exposure of mucin-secreting HT-29 cells to benzyl-N-acetyl-alpha-D-galactosaminide induces abnormal O-glycosylation of mucins and inhibits constitutive and stimulated MUC5AC secretion. *Biochem. J.* **334** (Pt 1), 283–295.
- Hershberg, R., and Petrov, D.A. (2008). Selection on codon bias. *Annu. Rev. Genet.* **42**, 287–299.
- Hienola, A., Tumova, S., Kuleskiy, E., and Rauvala, H. (2006). N-syndecan deficiency impairs neural migration in brain. *J. Cell Biol.* **174**, 569–580.
- Hinnebusch, A.G. (2011). Molecular mechanism of scanning and start codon selection in Eukaryotes. *Microbiol. Mol. Biol. Rev.* **75**, 434–467.
- Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R., and Zamir, A. (1965). Structure of a ribonucleic acid. *Science* (80-.). **147**, 1462–1465.
- Hollingsworth, M.A., and Swanson, B.J. (2004). Mucins in cancer: Protection and control of the cell surface. *Nat. Rev. Cancer* **4**, 45–60.
- Holt, K.H., Crosbie, R.H., Venzke, D.P., and Campbell, K.P. (2000). Biosynthesis of dystroglycan: Processing of a precursor propeptide. *FEBS Lett.* **468**, 79–83.
- Hopper, A.K. (2013). Transfer RNA post-transcriptional processing, turnover, and subcellular dynamics in the yeast *Saccharomyces cerevisiae*. *Genetics* **194**, 43–67.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57.

- Huang, L., Pike, D., Sleat, D.E., Nanda, V., and Lobel, P. (2014). Potential pitfalls and solutions for use of fluorescent fusion proteins to study the lysosome. *PLoS One* 9, e88893.
- Huang, L., Pu, J., He, F., Liao, B., Hao, B., Hong, W., Ye, X., Chen, J., Zhao, J., Liu, S., et al. (2017). Positive feedback of the amphiregulin-EGFR-ERK pathway mediates PM2.5 from wood smoke-induced MUC5AC expression in epithelial cells. *Sci. Rep.* 7, 11084.
- Huang, X., Poy, F., Zhang, R., Joachimiak, A., Sudol, M., and Eck, M.J. (2000). Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan. *Nat. Struct. Biol.* 7, 634–638.
- Huang, Y., Koonin, E. V., Lipman, D.J., and Przytycka, T.M. (2009b). Selection for minimization of translational frameshifting errors as a factor in the evolution of codon usage. *Nucleic Acids Res.* 37, 6799–6810.
- Huber, S., Leonardi, A., Dedon, P., and Begley, T. (2019). The Versatile Roles of the tRNA Epitranscriptome during Cellular Responses to Toxic Exposures and Environmental Stress. *Toxics* 7, 17.
- Huet, G., Kim, I., de Bolos, C., Lo-Guidice, J.M., Moreau, O., Hemon, B., Richet, C., Delannoy, P., Real, F.X., and Degand, P. (1995). Characterization of mucins and proteoglycans synthesized by a mucin-secreting HT-29 cell subpopulation. *J. Cell Sci.* 108 (Pt 3), 1275–1285.
- Huh, W., Falvo, J. V, Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O’Shea, E.K. (2003). Global analysis of protein localization in budding yeast. *Nature* 425, 686–691.
- Hussmann, J.A., and Press, W.H. (2014). Local correlations in codon preferences do not support a model of tRNA recycling. *Cell Rep.* 8, 1624–1629.
- Hussmann, J.A., Patchett, S., Johnson, A., Sawyer, S., and Press, W.H. (2015). Understanding biases in ribosome profiling experiments reveals signatures of translation dynamics in yeast. *PLoS Genet.* 11, 1–25.
- Hynes, R.O. (2009). The Extracellular Matrix: Not Just Pretty Fibrils. *Science* (80-.). 326, 1216–1219.
- Hynes, R.O., and Naba, A. (2012). Overview of the matrisome—An inventory of extracellular matrix constituents and functions. *Cold Spring Harb. Perspect. Biol.* 4.
- Iben, J.R., and Maraia, R.J. (2014). tRNA gene copy number variation in humans. *Gene* 536, 376–384.
- Ibraghimov-Beskrovnaya, O., Ervasti, J.M., Leveille, C.J., Slaughter, C.A., Sernett, S.W., and Campbell, K.P. (1992). Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix. *Nature* 355, 696–702.
- Ikemura, T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* 151, 389–409.
- Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* 2, 13–34.
- Incamato, D., Neri, F., Anselmi, F., and Oliviero, S. (2014). Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol.* 15, 491.
- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* (80-.). 324, 218–223.

- Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802.
- Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* 7, 1534–1550.
- Ingolia, N.T., Hussmann, J.A., and Weissman, J.S. (2019). Ribosome profiling: global views of translation. *Cold Spring Harb. Perspect. Biol.* 11, a032698.
- Iozzo, R. V., and Gubbiotti, M.A. (2018). Extracellular matrix: The driving force of mammalian diseases. *Matrix Biol.* 71–72, 1–9.
- Ishimura, R., Nagy, G., Dotu, I., Zhou, H., Yang, X.-L., Schimmel, P., Senju, S., Nishimura, Y., Chuang, J.H., and Ackerman, S.L. (2014). Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. *Science* (80-). 345, 455–459.
- Ishimura, R., Nagy, G., Dotu, I., Chuang, J.H., and Ackerman, S.L. (2016). Activation of GCN2 kinase by ribosome stalling links translation elongation with translation initiation. *Elife* 5, 1–22.
- Ito, K., and Chiba, S. (2013). Arrest Peptides: Cis -Acting Modulators of Translation . *Annu. Rev. Biochem.* 82, 171–202.
- Jackson, R.J., Hellen, C.U.T., and Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11, 113–127.
- Jackson, R.J., Hellen, C.U.T., and Pestova, T. V. (2012). Termination and post-termination events in eukaryotic translation (Elsevier Inc.).
- Johansson, M.J.O., and Byström, A.S. (2002). Dual function of the tRNA(m(5)U54)methyltransferase in tRNA maturation. *RNA* 8, 324–335.
- Kanaya, S., Yamada, Y., Kudo, Y., and Ikemura, T. (1999). Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: Gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* 238, 143–155.
- Karagöz, G.E., Acosta-Alvear, D., and Walter, P. (2019). The Unfolded Protein Response: Detecting and Responding to Fluctuations in the Protein-Folding Capacity of the Endoplasmic Reticulum. *Cold Spring Harb. Perspect. Biol.* a033886.
- Karamanos, N.K., Theocharis, A.D., Neill, T., and Iozzo, R. V. (2019). Matrix modeling and remodeling: A biological interplay regulating tissue homeostasis and diseases. *Matrix Biol.* 75–76, 1–11.
- Karcher, D., and Bock, R. (2009). Identification of the chloroplast adenosine-to-inosine tRNA editing enzyme. *RNA* 15, 1251–1257.
- Keffer-Wilkes, L.C., Veerareddygar, G.R., and Kothe, U. (2016). RNA modification enzyme TruB is a tRNA chaperone. *Proc. Natl. Acad. Sci.* 113, 14306–14311.
- Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., Chang, H.Y., and Segal, E. (2010). Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467, 103–107.
- Kesimer, M., Ford, A.A., Ceppe, A., Radicioni, G., Cao, R., Davis, C.W., Doerschuk, C.M., Alexis, N.E., Anderson, W.H., Henderson, A.G., et al. (2017). Airway Mucin Concentration as a Marker of Chronic Bronchitis. *N. Engl. J. Med.* 377, 911–922.
- Khatter, H., Myasnikov, A.G., Natchiar, S.K., and Klaholz, B.P. (2015). Structure of the human 80S ribosome. *Nature* 520, 640–645.
- Kim, J., Malashkevich, V., Roday, S., Lisbin, M., Schramm, V.L., and Almo, S.C. (2006).

- Structural and Kinetic Characterization of *Escherichia coli* TadA, the Wobble-Specific tRNA Deaminase †. *Biochemistry* 45, 6407–6416.
- Kim, K.W., Jee, H.M., Park, Y.H., Choi, B.S., Sohn, M.H., and Kim, K.-E. (2009). Relationship Between Amphiregulin and Airway Inflammation in Children With Asthma and Eosinophilic Bronchitis. *Chest* 136, 805–810.
- Kim, S.H., Suddath, F.L., Quigley, G.J., McPherson, A., Sussman, J.L., Wang, A.H., Seeman, N.C., and Rich, A. (1974). Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* (80-). 185, 435–440.
- Kimchi-Sarfaty, C., Oh, J.M., Kim, I.W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S. V, and Gottesman, M.M. (2007). A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science* (80-). 315, 525–528.
- King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., et al. (2008). The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451, 783–788.
- Kinnunen, T., Raulo, E., Nolo, R., Maccarana, M., Lindahl, U., and Rauvala, H. (1996). Neurite outgrowth in brain neurons induced by heparin-binding growth-associated molecule (HB-GAM) depends on the specific interaction of HB-GAM with heparan sulfate at the cell surface. *J. Biol. Chem.* 271, 2243–2248.
- Kinnunen, T., Kaksonen, M., Saarinen, J., Kalkkinen, N., Peng, H.B., and Rauvala, H. (1998). Cortactin-Src kinase signaling pathway is involved in N-syndecan- dependent neurite outgrowth. *J. Biol. Chem.* 273, 10702–10708.
- Kirkham, S., Kolsum, U., Rousseau, K., Singh, D., Vestbo, J., and Thornton, D.J. (2008). MUC5B is the major mucin in the gel phase of sputum in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 178, 1033–1039.
- Komar, A.A., Lesnik, T., and Reiss, C. (1999). Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett.* 462, 387–391.
- Koonin, E. V., and Novozhilov, A.S. (2017). Origin and evolution of the universal genetic code. *Annu. Rev. Genet.* 51, 45–62.
- Koutmou, K.S., Schuller, A.P., Brunelle, J.L., Radhakrishnan, A., Djuranovic, S., and Green, R. (2015). Ribosomes slide on lysine-encoding homopolymeric A stretches. *Elife* 4, 1–18.
- Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* 361, 13–37.
- Kramer, E.B., Vallabhaneni, H., Mayer, L.M., and Farabaugh, P.J. (2010). A comprehensive analysis of translational missense errors in the yeast *Saccharomyces cerevisiae*. *RNA* 16, 1797–1808.
- Kretz, K.A., Trewyn, R.W., Keith, G., and Grosjean, H. (1990). Site directed replacement of nucleotides in the anti-codon loop of tRNA: Application to the study of inosine biosynthesis in yeast tRNA Ala. *J. Chromatogr. Libr.* 45, B143–B171.
- Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-sequence determinants of expression in *Escherichia coli*. *Science* (80-). 324, 255–258.
- Kufe, D.W. (2009). Mucins in cancer: Function, prognosis and therapy. *Nat. Rev. Cancer* 9, 874–885.
- Kumar, P., Hellen, C.U.T., and Pestova, T. V (2016). Toward the mechanism of eIF4F-mediated ribosomal attachment to mammalian capped mRNAs. *Genes Dev.* 30, 1573–1588.
- Kwok, C.K., Tang, Y., Assmann, S.M., and Bevilacqua, P.C. (2015). The RNA structurome:

- Transcriptome-wide structure probing with next-generation sequencing. *Trends Biochem. Sci.* *40*, 221–232.
- Lang, T., Hansson, G.C., and Samuelsson, T. (2007). Gel-forming mucins appeared early in metazoan evolution. *Proc. Natl. Acad. Sci.* *104*, 16209–16214.
- Lang, T., Klasson, S., Larsson, E., Johansson, M.E.V., Hansson, G.C., and Samuelsson, T. (2016). Searching the Evolutionary Origin of Epithelial Mucus Protein Components - Mucins and FCGBP. *Mol. Biol. Evol.* *33*, 1921–1936.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* *10*, R25.
- Lareau, L.F., Hite, D.H., Hogan, G.J., and Brown, P.O. (2014). Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *Elife* *3*, 1–16.
- van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., et al. (2014). Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* *114*, 6589–6631.
- Lemjabbar-Alaoui, H., Sidhu, S.S., Mengistab, A., Gallup, M., and Basbaum, C. (2011). TACE/ADAM-17 Phosphorylation by PKC-Epsilon Mediates Premalignant Changes in Tobacco Smoke-Exposed Lung Cells. *PLoS One* *6*, e17489.
- Lemjabbar, H., Li, D., Gallup, M., Sidhu, S., Drori, E., and Basbaum, C. (2003). Tobacco Smoke-induced Lung Cell Proliferation Mediated by Tumor Necrosis Factor α -converting Enzyme and Amphiregulin. *J. Biol. Chem.* *278*, 26202–26207.
- Lemke, E.A. (2016). The Multiple Faces of Disordered Nucleoporins. *J. Mol. Biol.* *428*, 2011–2024.
- Leonova, E.I., and Galzitskaya, O. V. (2015). Cell communication using intrinsically disordered proteins: What can syndecans say? *J. Biomol. Struct. Dyn.* *33*, 1037–1050.
- Lesuffleur, T., Barbat, A., Dussaulx, E., and Zweibaum, A. (1990). Growth Adaptation to Methotrexate of HT-29 Human Colon Carcinoma Cells Is Associated with Their Ability to Differentiate into Columnar Absorptive and Mucus-secreting Cells. *Cancer Res.* *50*, 6334–6343.
- Lesuffleur, T., Porchet, N., Aubert, J.P., Swallow, D., Gum, J.R., Kim, Y.S., Real, F.X., and Zweibaum, A. (1993). Differential expression of the human mucin genes MUC1 to MUC5 in relation to growth and differentiation of different mucus-secreting HT-29 cell subpopulations. *J. Cell Sci.* *106* (Pt 3, 771–783.
- Li, G.W., Oh, E., and Weissman, J.S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* *484*, 538–541.
- Li, G.W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* *157*, 624–635.
- Lim, V.I. (1995). Analysis of action of the wobble adenine on codon reading within the ribosome. *J. Mol. Biol.* *252*, 277–282.
- Lim, V.I., and Curran, J.F. (2001). Analysis of codon:anticodon interactions within the ribosome provides new insights into codon reading and the genetic code structure. *RNA* *7*, 942–957.
- Lithwick, G., and Margalit, H. (2005). Relative predicted protein levels of functionally associated proteins are conserved across organisms. *Nucleic Acids Res.* *33*, 1051–1057.

- Liu, H., Wang, Q., He, Y., Chen, L., Hao, C., Jiang, C., Li, Y., Dai, Y., Kang, Z., and Xu, J.-R. (2016). Genome-wide A-to-I RNA editing in fungi independent of ADAR enzymes. *Genome Res.* 26, 499–509.
- Liu, H., Li, Y., Chen, D., Qi, Z., Wang, Q., Wang, J., Jiang, C., and Xu, J.-R. (2017). A-to-I RNA editing is developmentally regulated and generally adaptive for sexual reproduction in *Neurospora crassa*. *Proc. Natl. Acad. Sci.* 114, E7756–E7765.
- Lomakin, I.B., Shirokikh, N.E., Yusupov, M.M., Hellen, C.U.T., and Pestova, T. V. (2006). The fidelity of translation initiation: reciprocal activities of eIF1, IF3 and YciH. *EMBO J.* 25, 196–210.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- Loveland, A.B., Demo, G., Grigorieff, N., and Korostelev, A.A. (2017). Ensemble cryo-EM elucidates the mechanism of translation fidelity. *Nature* 546, 113–117.
- Lu, J., and Deutsch, C. (2008). Electrostatics in the ribosomal tunnel modulate chain elongation rates. *J. Mol. Biol.* 384, 73–86.
- Lund, E., and Dahlberg, J.E. (1998). Proofreading and aminoacylation of tRNAs before export from the nucleus. *Science* (80-.). 282, 2082–2085.
- Ma, J., Rubin, B.K., and Voynow, J.A. (2018). Mucins, Mucus, and Goblet Cells. *Chest* 154, 169–176.
- Machnicka, M.A., Olchowik, A., Grosjean, H., and Bujnicki, J.M. (2014). Distribution and frequencies of post-transcriptional modifications in tRNAs. *RNA Biol.* 11, 1619–1629.
- Manon-Jensen, T., Itoh, Y., and Couchman, J.R. (2010). Proteoglycans in health and disease: The multiple roles of syndecan shedding. *FEBS J.* 277, 3876–3889.
- Mantsyzov, A.B., Ivanova, E. V., Birdsall, B., Alkalaeva, E.Z., Kryuchkova, P.N., Kelly, G., Frolova, L.Y., and Polshakov, V.I. (2010). NMR solution structure and function of the C-terminal domain of eukaryotic class 1 polypeptide chain release factor. *FEBS J.* 277, 2611–2627.
- Manzo, N.D., Foster, W.M., and Stripp, B.R. (2012). Amphiregulin-Dependent Mucous Cell Metaplasia in a Model of Nonallergic Lung Injury. *Am. J. Respir. Cell Mol. Biol.* 47, 349–357.
- Maraia, R., and Arimbasseri, A. (2017). Factors That Shape Eukaryotic tRNAomes: Processing, Modification and Anticodon–Codon Use. *Biomolecules* 7, 26.
- Maraia, R.J., and Iben, J.R. (2014). Different types of secondary information in the genetic code. *RNA* 20, 977–984.
- Marck, C., and Grosjean, H. (2002). tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* 8, 1189–1232.
- Martino, M.E.B., Olsen, J.C., Fulcher, N.B., Wolfgang, M.C., O’Neal, W.K., and Ribeiro, C.M.P. (2009). Airway Epithelial Inflammation-induced Endoplasmic Reticulum Ca²⁺ Store Expansion Is Mediated by X-box Binding Protein-1. *J. Biol. Chem.* 284, 14904–14913.
- McKenney, K.M. (2018). Investigating the basis of tRNA editing and modification enzyme coactivation in *Trypanosoma brucei*. The Ohio State University.
- Merrick, W.C., and Pavitt, G.D. (2018). Protein synthesis initiation in eukaryotic cells. *Cold Spring Harb. Perspect. Biol.* 10, a033092.
- Moazed, D., and Noller, H.F. (1989). Intermediate states in the movement of transfer RNA in the ribosome. *Nature* 342, 142–148.

- Monies, D., Abouelhoda, M., AlSayed, M., Alhassnan, Z., Alotaibi, M., Kayyali, H., Al-Owain, M., Shah, A., Rahbeeni, Z., Al-Muhaizea, M.A., et al. (2017). The landscape of genetic diseases in Saudi Arabia based on the first 1000 diagnostic panels and exomes. *Hum. Genet.* *136*, 921–939.
- Monies, D., Vågbø, C.B., Al-Owain, M., Alhomaiddi, S., and Alkuraya, F.S. (2019a). Recessive truncating mutations in *ALKBH8* cause intellectual disability and severe impairment of wobble uridine modification. *Am. J. Hum. Genet.* 1–8.
- Monies, D., Abouelhoda, M., Assoum, M., Moghrabi, N., Rafiullah, R., Almontashiri, N., Alowain, M., Alzaidan, H., Alsayed, M., Subhani, S., et al. (2019b). Lessons learned from large-scale, first-tier clinical exome sequencing in a highly consanguineous population. *Am. J. Hum. Genet.* 1–20.
- Morisaki, T., Lyon, K., DeLuca, K.F., DeLuca, J.G., English, B.P., Zhang, Z., Lavis, L.D., Grimm, J.B., Viswanathan, S., Looger, L.L., et al. (2016). Real-time quantification of single RNA translation dynamics in living cells. *Science* (80-). *352*, 1425–1429.
- Moriyama, E.N., and Powell, J.R. (1997). Codon usage bias and tRNA abundance in *Drosophila*. *J. Mol. Evol.* *45*, 514–523.
- Morscher, R.J., Ducker, G.S., Li, S.H.-J., Mayer, J.A., Gitai, Z., Sperl, W., and Rabinowitz, J.D. (2018). Mitochondrial translation requires folate-dependent tRNA methylation. *Nature* *554*, 128–132.
- Mortimer, S.A., Kidwell, M.A., and Doudna, J.A. (2014). Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* *15*, 469–479.
- Muto, H., and Ito, K. (2008). Peptidyl-prolyl-tRNA at the ribosomal P-site reacts poorly with puromycin. *Biochem. Biophys. Res. Commun.* *366*, 1043–1047.
- Najafabadi, H.S., Goodarzi, H., and Salavati, R. (2009). Universal function-specificity of codon usage. *Nucleic Acids Res.* *37*, 7014–7023.
- Nakamura, Y., Gojobori, T., and Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* *28*, 292.
- Nedialkova, D.D., and Leidel, S.A. (2015). Optimization of codon translation rates via tRNA modifications maintains proteome integrity. *Cell* *161*, 1606–1618.
- Nickolls, A.R., and Bönnemann, C.G. (2018). The roles of dystroglycan in the nervous system: insights from animal models of muscular dystrophy. *Dis. Model. Mech.* *11*, dmm035931.
- Nirenberg, M.W., and Matthaei, J.H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci.* *47*, 1588–1602.
- Nottrott, S., Simard, M.J., and Richter, J.D. (2006). Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat. Struct. Mol. Biol.* *13*, 1108–1114.
- Novoa, E.M., and Ribas de Pouplana, L. (2012). Speeding with control: Codon usage, tRNAs, and ribosomes. *Trends Genet.* *28*, 574–581.
- Novoa, E.M., Pavon-Eternod, M., Pan, T., and Ribas de Pouplana, L. (2012). A role for tRNA modifications in genome structure and codon usage. *Cell* *149*, 202–213.
- O'Brien, E.P., Vendruscolo, M., and Dobson, C.M. (2014). Kinetic modelling indicates that fast-translating codons can coordinate cotranslational protein folding by avoiding misfolded intermediates. *Nat. Commun.* *5*, 2988.
- O'Connor, P.B.F., Andreev, D.E., and Baranov, P. V. (2016). Comparative survey of the relative impact of mRNA features on local ribosome profiling read density. *Nat. Commun.*

- Ogle, J.M., Brodersen, D.E., Clemons, W.M., Tarry, M.J., Carter, A.P., and Ramakrishnan, V. (2001). Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* (80-.). 292, 897–902.
- Otsubo, Y., Matsuo, T., Nishimura, A., Yamamoto, M., and Yamashita, A. (2018). tRNA production links nutrient conditions to the onset of sexual differentiation through the TORC1 pathway. *EMBO Rep.* 19, e44867.
- Ozbek, S., Balasubramanian, P.G., Chiquet-Ehrismann, R., Tucker, R.P., and Adams, J.C. (2010). The evolution of extracellular matrix. *Mol. Biol. Cell* 21, 4300–4305.
- Pan, T. (2018). Modifications and functional genomics of human transfer RNA. *Cell Res.* 28, 395–404.
- Parisien, M., Wang, X., and Pan, T. (2013). Diversity of human tRNA genes from the 1000-genomes project. *RNA Biol.* 10, 1853–1867.
- Park, E.-H., Walker, S.E., Lee, J.M., Rothenburg, S., Lorsch, J.R., and Hinnebusch, A.G. (2011). Multiple elements in the eIF4G1 N-terminus promote assembly of eIF4G1•PABP mRNPs in vivo. *EMBO J.* 30, 302–316.
- Parmley, J.L., and Huynen, M.A. (2009). Clustering of codons with rare cognate tRNAs in human genes suggests an extra level of expression regulation. *PLoS Genet.* 5, 19–21.
- Patil, A., Chan, C.T.Y., Dyavaiah, M., Rooney, J.P., Dedon, P.C., and Begley, T.J. (2012). Translational infidelity-induced protein stress results from a deficiency in Trm9-catalyzed tRNA modifications. *RNA Biol.* 9, 990–1001.
- Pavlov, M.Y., Watts, R.E., Tan, Z., Cornish, V.W., Ehrenberg, M., and Forster, A.C. (2009). Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proc. Natl. Acad. Sci.* 106, 50–54.
- Pavon-Eternod, M., Gomes, S., Geslain, R., Dai, Q., Rosner, M.R., and Pan, T. (2009). tRNA over-expression in breast cancer and functional consequences. *Nucleic Acids Res.* 37, 7268–7280.
- Pechmann, S., and Frydman, J. (2013). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat. Struct. Mol. Biol.* 20, 237–243.
- Pechmann, S., Chartron, J.W., and Frydman, J. (2014). Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. *Nat. Struct. Mol. Biol.* 21, 1100–1105.
- Percudani, R., Pavesi, A., and Ottonello, S. (1997). Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. *J. Mol. Biol.* 268, 322–330.
- Pereira, M., Francisco, S., Varanda, A.S., Santos, M., Santos, M.A.S., and Soares, A.R. (2018). Impact of tRNA modifications and tRNA-modifying enzymes on proteostasis and human disease. *Int. J. Mol. Sci.* 19.
- Persson, B.C., Gustafsson, C., Berg, D.E., and Bjork, G.R. (1992). The gene for a tRNA modifying enzyme, m5U54-methyltransferase, is essential for viability in *Escherichia coli*. *Proc. Natl. Acad. Sci.* 89, 3995–3998.
- Pestova, T. V, Lorsch, J.R., and Hellen, C.U.T. (2007). The mechanism of translation initiation in eukaryotes. In *Translational Control in Biology and Medicine*, M.B.M. Mathews, and N. Sonenberg, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 87–128.
- Petersen, C.P., Bordeleau, M.E., Pelletier, J., and Sharp, P.A. (2006). Short RNAs repress translation after initiation in mammalian cells. *Mol. Cell* 21, 533–542.
- Peysnelon, F., Xue, B., Uversky, V.N., and Ricard-Blum, S. (2011). Intrinsic disorder of the

- extracellular matrix. *Mol. Biosyst.* **7**, 3353–3365.
- Phizicky, E.M., and Hopper, A.K. (2010). tRNA biology charges to the front. *Genes Dev.* **24**, 1832–1860.
- Piovesan, D., Tabaro, F., Paladin, L., Necci, M., Mičetić, I., Camilloni, C., Davey, N., Dosztányi, Z., Mészáros, B., Monzon, A.M., et al. (2018). MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. *Nucleic Acids Res.* **46**, D471–D476.
- Pisarev, A. V., Hellen, C.U.T., and Pestova, T. V. (2007). Recycling of eukaryotic posttermination ribosomal complexes. *Cell* **131**, 286–299.
- Pisarev, A. V., Skabkin, M.A., Pisareva, V.P., Skabkina, O. V., Rakotondrafara, A.M., Hentze, M.W., Hellen, C.U.T., and Pestova, T. V. (2010). The role of ABCE1 in eukaryotic posttermination ribosomal recycling. *Mol. Cell* **37**, 196–210.
- Plotkin, J.B., and Kudla, G. (2011). Synonymous but not the same: The causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42.
- Plotkin, J.B., Robins, H., and Levine, A.J. (2004). Tissue-specific codon usage and the expression of human genes. *Proc. Natl. Acad. Sci.* **101**, 12588–12591.
- Polosa, R., Puddicombe, S.M., Krishna, M.T., Tuck, A.B., Howarth, P.H., Holgate, S.T., and Davies, D.E. (2002). Expression of c-erbB receptors and ligands in the bronchial epithelium of asthmatic subjects. *J. Allergy Clin. Immunol.* **109**, 75–81.
- Pop, C., Rouskin, S., Ingolia, N.T., Han, L., Phizicky, E.M., Weissman, J.S., and Koller, D. (2014). Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol. Syst. Biol.* **10**, 770–770.
- Preis, A., Heuer, A., Barrio-Garcia, C., Hauser, A., Eyler, D.E., Berninghausen, O., Green, R., Becker, T., and Beckmann, R. (2014). Cryoelectron microscopic structures of eukaryotic translation termination complexes containing eRF1-eRF3 or eRF1-ABCE1. *Cell Rep.* **8**, 59–65.
- Presnyak, V., Alhusaini, N., Chen, Y.-H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., et al. (2015). Codon optimality is a major determinant of mRNA stability. *Cell* **160**, 1111–1124.
- Puddicombe, S.M., Polosa, R., Richter, A., Krishna, M.T., Howarth, P.H., Holgate, S.T., and Davies, D.E. (2000). Involvement of the epidermal growth factor receptor in epithelial repair in asthma. *FASEB J.* **14**, 1362–1374.
- Purvis, I.J., Bettany, A.J.E., Santiago, T.C., Coggins, J.R., Duncan, K., Eason, R., and Brown, A.J.P. (1987). The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis. *J. Mol. Biol.* **193**, 413–417.
- Qian, W., Yang, J.R., Pearson, N.M., Maclean, C., and Zhang, J. (2012). Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.* **8**.
- Quax, T.E.F., Claassens, N.J., Söll, D., and van der Oost, J. (2015). Codon bias as a means to fine-tune gene expression. *Mol. Cell* **59**, 149–161.
- Rafels-Ybern, À., Stephan-Otto Attolini, C., and de Pouplana, L. (2015). Distribution of ADAT-dependent codons in the human transcriptome. *Int. J. Mol. Sci.* **16**, 17303–17314.
- Rafels-Ybern, À., Torres, A.G., Grau-Bove, X., Ruiz-Trillo, I., and Ribas de Pouplana, L. (2018). Codon adaptation to tRNAs with Inosine modification at position 34 is widespread among Eukaryotes and present in two Bacterial phyla. *RNA Biol.* **15**, 500–507.
- Rafels-Ybern, À., Torres, A.G., Camacho, N., Herencia-Roperó, A., Roura Frigolé, H., Wulff, T.F., Raboteq, M., Bordons, A., Grau-Bove, X., Ruiz-Trillo, I., et al. (2019). The expansion of inosine at the wobble position of tRNAs, and its role in the evolution of

- proteomes. *Mol. Biol. Evol.* 36, 650–662.
- Ragone, F.L., Spears, J.L., Wohlgamuth-Benedum, J.M., Kreef, N., Papavasiliou, F.N., and Alfonzo, J.D. (2011). The C-terminal end of the *Trypanosoma brucei* editing deaminase plays a critical role in tRNA binding. *RNA* 17, 1296–1306.
- Rak, R., Dahan, O., and Pilpel, Y. (2018). Repertoires of tRNAs: The couplers of genomics and proteomics. *Annu. Rev. Cell Dev. Biol.* 34, 239–264.
- Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell* 108, 557–572.
- Ramos, J., and Fu, D. (2019). The emerging impact of tRNA modifications in the brain and nervous system. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1862, 412–428.
- Rapino, F., Delaunay, S., Rambow, F., Zhou, Z., Tharun, L., De Tullio, P., Sin, O., Shostak, K., Schmitz, S., Piepers, J., et al. (2018). Codon-specific translation reprogramming promotes resistance to targeted therapy. *Nature* 558, 605–609.
- Rich, A., and RajBhandary, U.L. (1976). Transfer RNA: Molecular structure, sequence, and properties. *Annu. Rev. Biochem.* 45, 805–860.
- Richter, J.D., and Collier, J. (2015). Pausing on polyribosomes: Make way for elongation in translational control. *Cell* 163, 292–300.
- Rodnina, M. V. (2016). The ribosome in action: Tuning of translational efficiency and protein folding. *Protein Sci.* 25, 1390–1406.
- Rokas, A. (2008). The Origins of Multicellularity and the Early History of the Genetic Toolkit For Animal Development. *Annu. Rev. Genet.* 42, 235–251.
- Rolfe, D.F., and Brown, G.C. (1997). Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiol. Rev.* 77, 731–758.
- Rose, M.C., and Voynow, J.A. (2006). Respiratory Tract Mucin Genes and Mucin Glycoproteins in Health and Disease. *Physiol. Rev.* 86, 245–278.
- Rosenthal, J.J.C. (2015). The emerging role of RNA editing in plasticity. *J. Exp. Biol.* 218, 1812–1821.
- Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölper, S., Krüger, M., and Stainier, D.Y.R. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* 524, 230–233.
- Röth, S., Fulcher, L.J., and Sapkota, G.P. (2019). Advances in targeted degradation of endogenous proteins. *Cell. Mol. Life Sci.* 76, 2761–2777.
- Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505, 701–705.
- Roy, M.G., Livraghi-Butrico, A., Fletcher, A.A., McElwee, M.M., Evans, S.E., Boerner, R.M., Alexander, S.N., Bellinghausen, L.K., Song, A.S., Petrova, Y.M., et al. (2014). Muc5b is required for airway defence. *Nature* 505, 412–416.
- Rozov, A., Demeshkina, N., Westhof, E., Yusupov, M., and Yusupova, G. (2015). Structural insights into the translational infidelity mechanism. *Nat. Commun.* 6, 1–9.
- Rubio, M.A.T., Pastar, I., Gaston, K.W., Ragone, F.L., Janzen, C.J., Cross, G.A.M., Papavasiliou, F.N., and Alfonzo, J.D. (2007). An adenosine-to-inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA. *Proc. Natl. Acad. Sci.* 104, 7821–7826.
- Rubio, M.A.T., Gaston, K.W., McKenney, K.M., Fleming, I.M.C., Paris, Z., Limbach, P.A., and Alfonzo, J.D. (2017). Editing and methylation at a single site by functionally

- interdependent activities. *Nature* 542, 494–497.
- Sagi, D., Rak, R., Gingold, H., Adir, I., Maayan, G., Dahan, O., Broday, L., Pilpel, Y., and Rechavi, O. (2016). Tissue-and Time-Specific Expression of Otherwise Identical tRNA Genes. *PLOS Genet.* 12, e1006264.
- Saikia, M., Wang, X., Mao, Y., Wan, J., Pan, T., and Qian, S.-B. (2016). Codon optimality controls differential mRNA translation during amino acid starvation. *RNA* 22, 1719–1727.
- Saint-Léger, A., Bello, C., Dans, P.D., Torres, A.G., Novoa, E.M., Camacho, N., Orozco, M., Kondrashov, F.A., and Ribas de Pouplana, L. (2016). Saturation of recognition elements blocks evolution of new tRNA identities. *Sci. Adv.* 2, e1501860.
- Salehi Chaleshtori, A.R., Miyake, N., Ahmadvand, M., Bashti, O., Matsumoto, N., and Noruzinia, M. (2018). A novel 8-bp duplication in ADAT3 causes mild intellectual disability. *Hum. Genome Var.* 5, 7.
- Sampson, J.R., and Uhlenbeck, O.C. (1988). Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed in vitro. *Proc. Natl. Acad. Sci.* 85, 1033–1037.
- Santos, D.A., Shi, L., Tu, B.P., and Weissman, J.S. (2019). Cycloheximide can distort measurements of mRNA levels and translation efficiency. *Nucleic Acids Res.* 1–12.
- Satpati, P., and Åqvist, J. (2014). Why base tautomerization does not cause errors in mRNA decoding on the ribosome. *Nucleic Acids Res.* 42, 12876–12884.
- Schaffer, A.E., Pinkard, O., and Collier, J.M. (2019). tRNA metabolism and neurodevelopmental disorders. *Annu. Rev. Genomics Hum. Genet.* 20, annurev-genom-083118-015334.
- Schaub, M., and Keller, W. (2002). RNA editing by adenosine deaminases generates RNA and protein diversity. *Biochimie* 84, 791–803.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682.
- Schmitt, B.M., Rudolph, K.L.M., Karagianni, P., Fonseca, N.A., White, R.J., Talianidis, I., Odom, D.T., Marioni, J.C., and Kutter, C. (2014). High-resolution mapping of transcriptional dynamics across tissue development reveals a stable mRNA–tRNA interface. *Genome Res.* 24, 1797–1807.
- Schuller, A.P., and Green, R. (2018). Roadblocks and resolutions in eukaryotic translation. *Nat. Rev. Mol. Cell Biol.* 19, 526–541.
- Shabalina, S.A., Ogurtsov, A.Y., and Spiridonov, N.A. (2006). A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.* 34, 2428–2437.
- Shaner, N.C., Steinbach, P.A., and Tsien, R.Y. (2005). A guide to choosing fluorescent proteins. *Nat. Methods* 2, 905–909.
- Shao, S., Murray, J., Brown, A., Taunton, J., Ramakrishnan, V., and Hegde, R.S. (2016). Decoding mammalian ribosome-mRNA states by translational GTPase complexes. *Cell* 167, 1229–1240.e15.
- Sharkia, R., Zalan, A., Jabareen-Masri, A., Zahalka, H., and Mahajnah, M. (2018). A new case confirming and expanding the phenotype spectrum of ADAT3-related intellectual disability syndrome. *Eur. J. Med. Genet.* 0–1.
- Sharma, A.K., Sormanni, P., Ahmed, N., Ciryam, P., Friedrich, U.A., Kramer, G., and O'Brien, E.P. (2019). A chemical kinetic basis for measuring translation initiation and elongation rates from ribosome profiling data. *PLOS Comput. Biol.* 15, e1007070.
- Sharp, P.M., Tuohy, T.M.F., and Mosurski, K.R. (1986). Codon usage in yeast: cluster

- analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* *14*, 5125–5143.
- Sharp, P.M., Averof, M., Lloyd, A.T., Matassi, G., and Peden, J.F. (1995). DNA sequence evolution: the sounds of silence. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.* *349*, 241–247.
- Shi, H., and Moore, P.B. (2000). The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA* *6*, 1091–1105.
- Shoemaker, C.J., and Green, R. (2011). Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *Proc. Natl. Acad. Sci.* *108*, E1392–E1398.
- Shoyab, M., Plowman, G., McDonald, V., Bradley, J., and Todaro, G. (1989). Structure and function of human amphiregulin: a member of the epidermal growth factor family. *Science* (80-.). *243*, 1074–1076.
- Siller, E., DeZwaan, D.C., Anderson, J.F., Freeman, B.C., and Barral, J.M. (2010). Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. *J. Mol. Biol.* *396*, 1310–1318.
- Skabkin, M.A., Skabkina, O. V., Dhote, V., Komar, A.A., Hellen, C.U.T., and Pestova, T. V. (2010). Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. *Genes Dev.* *24*, 1787–1801.
- Van der Sluis, M., De Koning, B.A.E., De Bruijn, A.C.J.M., Velcich, A., Meijerink, J.P.P., Van Goudoever, J.B., Büller, H.A., Dekker, J., Van Seuning, I., Renes, I.B., et al. (2006). Muc2-Deficient Mice Spontaneously Develop Colitis, Indicating That MUC2 Is Critical for Colonic Protection. *Gastroenterology* *131*, 117–129.
- Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in Eukaryotes: Mechanisms and biological targets. *Cell* *136*, 731–745.
- Song, H., Mugnier, P., Das, A.K., Webb, H.M., Evans, D.R., Tuite, M.F., Hemmings, B.A., and Barford, D. (2000). The crystal structure of human eukaryotic release factor eRF1—mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* *100*, 311–321.
- Sørensen, M.A., Kurland, C.G., and Pedersen, S. (1989). Codon usage determines translation rate in *Escherichia coli*. *J. Mol. Biol.* *207*, 365–377.
- Spahn, C.M.T., Gomez-Lorenzo, M.G., Grassucci, R.A., Jørgensen, R., Andersen, G.R., Beckmann, R., Penczek, P.A., Ballesta, J.P.G., and Frank, J. (2004). Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J.* *23*, 1008–1019.
- Spears, J.L., Rubio, M.A.T., Gaston, K.W., Wywiał, E., Strikoudis, A., Bujnicki, J.M., Papavasiliou, F.N., and Alfonzo, J.D. (2011). A Single Zinc Ion Is Sufficient for an Active Trypanosoma brucei tRNA Editing Deaminase. *J. Biol. Chem.* *286*, 20366–20374.
- Spencer, P.S., Siller, E., Anderson, J.F., and Barral, J.M. (2012). Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J. Mol. Biol.* *422*, 328–335.
- Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.-W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T., et al. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* *519*, 486–490.
- Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* *26*, 148–153.
- Stadler, M., and Fire, A. (2011). Wobble base-pairing slows in vivo translation elongation in metazoans. *RNA* *17*, 2063–2073.

- Steigemann, P., Molitor, A., Fellert, S., Jäckle, H., and Vorbrüggen, G. (2004). Heparan Sulfate Proteoglycan Syndecan Promotes Axonal and Myotube Guidance by Slit/Robo Signaling. *Curr. Biol.* *14*, 225–230.
- Stenico, M., Lloyd, A.T., and Sharp, P.M. (1994). Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. *Nucleic Acids Res.* *22*, 2437–2446.
- Stoletzki, N., and Eyre-Walker, A. (2006). Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Mol. Biol. Evol.* *24*, 374–381.
- Strobel, E.J., Yu, A.M., and Lucks, J.B. (2018). High-throughput determination of RNA structures. *Nat. Rev. Genet.* *19*, 615–634.
- Strous, G.J., and Dekker, J. (1992). Mucin-Type Glycoproteins. *Crit. Rev. Biochem. Mol. Biol.* *27*, 57–92.
- Supek, F. (2016). The Code of Silence: Widespread Associations Between Synonymous Codon Biases and Gene Function. *J. Mol. Evol.* *82*, 65–73.
- Tahmasebi, S., Khoutorsky, A., Mathews, M.B., and Sonenberg, N. (2018). Translation deregulation in human disease. *Nat. Rev. Mol. Cell Biol.* *19*, 791–807.
- Takyar, S., Hickerson, R.P., and Noller, H.F. (2005). mRNA helicase activity of the ribosome. *Cell* *120*, 49–58.
- Taniguchi, T., Miyauchi, K., Sakaguchi, Y., Yamashita, S., Soma, A., Tomita, K., and Suzuki, T. (2018). Acetate-dependent tRNA acetylation required for decoding fidelity in protein synthesis. *Nat. Chem. Biol.* *14*, 1010–1020.
- Tats, A., Tenson, T., and Remm, M. (2008). Preferred and avoided codon pairs in three domains of life. *BMC Genomics* *9*, 463.
- Taylor, D.J., Nilsson, J., Merrill, A.R., Andersen, G.R., Nissen, P., and Frank, J. (2007). Structures of modified eEF2·80S ribosome complexes reveal the role of GTP hydrolysis in translocation. *EMBO J.* *26*, 2421–2431.
- Teichert, I., Dahlmann, T.A., Kück, U., and Nowrousian, M. (2017). RNA editing during sexual development occurs in distantly related filamentous ascomycetes. *Genome Biol. Evol.* *9*, 855–868.
- Thai, P., Loukoianov, A., Wachi, S., and Wu, R. (2008). Regulation of Airway Mucin Gene Expression. *Annu. Rev. Physiol.* *70*, 405–429.
- Thanaraj, T.A., and Argos, P. (1996). Protein secondary structural types are differentially coded on messenger RNA. *Protein Sci.* *5*, 1973–1983.
- Theocharis, A.D., Manou, D., and Karamanos, N.K. (2019). The extracellular matrix as a multitasking player in disease. *FEBS J.* febs.14818.
- Thommen, M., Holtkamp, W., and Rodnina, M. V. (2017). Co-translational protein folding: progress and methods. *Curr. Opin. Struct. Biol.* *42*, 83–89.
- Thornlow, B.P., Hough, J., Roger, J.M., Gong, H., Lowe, T.M., and Corbett-Detig, R.B. (2018). Transfer RNA genes experience exceptionally elevated mutation rates. *Proc. Natl. Acad. Sci. U. S. A.* *115*, 8996–9001.
- Thornton, D.J., Rousseau, K., and McGuckin, M.A. (2008). Structure and function of the polymeric mucins in airways mucus. *Annu. Rev. Physiol.* *70*, 459–486.
- Torres, A.G. (2019). Enjoy the Silence: Nearly Half of Human tRNA Genes Are Silent. *Bioinform. Biol. Insights* *13*, 117793221986845.
- Torres, A.G., Batlle, E., and Ribas de Pouplana, L. (2014a). Role of tRNA modifications in human diseases. *Trends Mol. Med.* *20*, 306–314.

- Torres, A.G., Piñeyro, D., Filonava, L., Stracker, T.H., Batlle, E., and Ribas de Pouplana, L. (2014b). A-to-I editing on tRNAs: biochemical, biological and evolutionary implications. *FEBS Lett.* *588*, 4279–4286.
- Torres, A.G., Piñeyro, D., Rodríguez-Escribà, M., Camacho, N., Reina, O., Saint-Léger, A., Filonava, L., Batlle, E., and Ribas de Pouplana, L. (2015). Inosine modifications in human tRNAs are incorporated at the precursor tRNA level. *Nucleic Acids Res.* *43*, 5145–5157.
- Torres, A.G., Reina, O., Stephan-Otto Attolini, C., and Ribas de Pouplana, L. (2019). Differential expression of human tRNA genes drives the abundance of tRNA-derived fragments. *Proc. Natl. Acad. Sci.* *116*, 8451–8456.
- Tsutsumi, S., Sugiura, R., Ma, Y., Tokuoka, H., Ohta, K., Ohte, R., Noma, A., Suzuki, T., and Kuno, T. (2007). Wobble inosine tRNA modification is essential to cell cycle progression in *G 1 / S* and *G 2 / M* transitions in fission yeast. *J. Biol. Chem.* *282*, 33459–33465.
- Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010a). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* *141*, 344–354.
- Tuller, T., Waldman, Y.Y., Kupiec, M., and Ruppin, E. (2010b). Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci.* *107*, 3645–3650.
- Tytgat, K.M.A.J., Büller, H.A., Opdam, F.J.M., Kim, Y.S., Einerhand, A.W.C., and Dekker, J. (1994). Biosynthesis of human colonic mucin: Muc2 is the prominent secretory mucin. *Gastroenterology* *107*, 1352–1363.
- Ude, S., Lassak, J., Starosta, A.L., Kraxenberger, T., Wilson, D.N., and Jung, K. (2013). Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science* (80-.). *339*, 82–85.
- Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., et al. (2010). Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* *28*, 1248–1250.
- Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., et al. (2015). Tissue-based map of the human proteome. *Science* (80-.). *347*, 1260419–1260419.
- UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* *47*, D506–D515.
- Urrutia, A.O., and Hurst, L.D. (2001). Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. *Genetics* *159*, 1191–1199.
- Varenne, S., Buc, J., Lloubes, R., and Lazdunski, C. (1984). Translation is a non-uniform process. *J. Mol. Biol.* *180*, 549–576.
- Varki, A., Cummings, R.D., Aebi, M., Packer, N.H., Seeberger, P.H., Esko, J.D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., et al. (2015). Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* *25*, 1323–1324.
- Velcich, A., Yang, W.C., Heyer, J., Fragale, A., Nicholas, C., Viani, S., Kucherlapati, R., Lipkin, M., Yang, K., and Augenlicht, L. (2002). Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science* (80-.). *295*, 1726–1729.
- Vestsigian, K., Woese, C., and Goldenfeld, N. (2006). Collective evolution and the genetic code. *Proc. Natl. Acad. Sci.* *103*, 10696–10701.
- Villa, N., Do, A., Hershey, J.W.B., and Fraser, C.S. (2013). Human Eukaryotic Initiation Factor 4G (eIF4G) Protein Binds to eIF3c, -d, and -e to Promote mRNA Recruitment to the Ribosome. *J. Biol. Chem.* *288*, 32932–32940.

- Wagner, C.E., Wheeler, K.M., and Ribbeck, K. (2018). Mucins and their role in shaping the functions of mucus barriers. *Annu. Rev. Cell Dev. Biol.* *34*, 189–215.
- Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E., et al. (2014). Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* *505*, 706–709.
- Wang, C., Xu, J.R., and Liu, H. (2016). A-to-I RNA editing independent of ADARs in filamentous fungi. *RNA Biol.* *13*, 940–945.
- Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., and Brown, P.O. (2002). Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci.* *99*, 5860–5865.
- Wang, Y.J., Vaidyanathan, P.P., Rojas-Duran, M.F., Udeshi, N.D., Bartoli, K.M., Carr, S.A., and Gilbert, W. V. (2018). Lso2 is a conserved ribosome-bound protein required for translational recovery in yeast. *PLOS Biol.* *16*, e2005903.
- Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* *337*, 635–645.
- Warnecke, T., and Hurst, L.D. (2010). GroEL dependency affects codon usage—support for a critical role of misfolding in gene evolution. *Mol. Syst. Biol.* *6*, 1–11.
- Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess Jr, J.W., Swanstrom, R., Burch, C.L., and Weeks, K.M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* *460*, 711–716.
- Weinberg, D.E., Shah, P., Eichhorn, S.W., Hussmann, J.A., Plotkin, J.B., and Bartel, D.P. (2016). Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep.* *14*, 1787–1799.
- Wen, J.-D., Lancaster, L., Hodges, C., Zeri, A.-C., Yoshimura, S.H., Noller, H.F., Bustamante, C., and Tinoco, I. (2008). Following translation by single ribosomes one codon at a time. *Nature* *452*, 598–603.
- Wilson, D.N., Arenz, S., and Beckmann, R. (2016). Translation regulation via nascent polypeptide-mediated ribosome stalling. *Curr. Opin. Struct. Biol.* *37*, 123–133.
- Woese, C.R., Dugre, D.H., Dugre, S.A., Kondo, M., and Saxinger, W.C. (1966). On the fundamental nature and evolution of the genetic code. *Cold Spring Harb. Symp. Quant. Biol.* *31*, 723–736.
- Wohlgemuth, I., Brenner, S., Beringer, M., and Rodnina, M. V. (2008). Modulation of the rate of peptidyl transfer on the ribosome by the nature of substrates. *J. Biol. Chem.* *283*, 32229–32235.
- Wolf, J., Gerber, A.P., and Keller, W. (2002). TadA, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. *EMBO J.* *21*, 3841–3851.
- Woolstenhulme, C.J., Guydosh, N.R., Green, R., and Buskirk, A.R. (2015). High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. *Cell Rep.* *11*, 13–21.
- Wu, B., Eliscovich, C., Yoon, Y.J., and Singer, R.H. (2016). Translation dynamics of single mRNAs in live cells and neurons. *Science* (80-.). *352*, 1430–1435.
- Wulff, T.F., Argüello, R.J., Molina Jordàn, M., Roura Frigolé, H., Hauquier, G., Filonava, L., Camacho, N., Gatti, E., Pierre, P., Ribas de Pouplana, L., et al. (2017). Detection of a subset of posttranscriptional transfer RNA modifications in vivo with a restriction fragment length polymorphism-based method. *Biochemistry* *56*, 4029–4038.
- Xiao, Z., Zou, Q., Liu, Y., and Yang, X. (2016). Genome-wide assessment of differential

- translations with ribosome profiling data. *Nat. Commun.* 7, 1–11.
- El Yacoubi, B., Bailly, M., and de Crécy-Lagard, V. (2012). Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu. Rev. Genet.* 46, 69–95.
- Yan, X., Hoek, T.A., Vale, R.D., and Tanenbaum, M.E. (2016). Dynamics of translation of single mRNA molecules in vivo. *Cell* 165, 976–989.
- Yankaskas, J.R., Haizlip, J.E., Conrad, M., Koval, D., Lazarowski, E., Paradiso, A.M., Rinehart, C.A., Sarkadi, B., Schlegel, R., and Boucher, R.C. (1993). Papilloma virus immortalized tracheal epithelial cells retain a well-differentiated phenotype. *Am. J. Physiol. Physiol.* 264, C1219–C1230.
- Yoshihisa, T. (2014). Handling tRNA introns, archaeal way and eukaryotic way. *Front. Genet.* 5, 1–16.
- Yruela, I., Oldfield, C.J., Niklas, K.J., and Dunker, A.K. (2017). Evidence for a strong correlation between transcription factor protein disorder and organismic complexity. *Genome Biol. Evol.* 9, 1248–1265.
- Yu, C.-H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M.S., and Liu, Y. (2015). Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Mol. Cell* 59, 744–754.
- Yusupova, G., and Yusupov, M. (2014). High-resolution structure of the eukaryotic 80S ribosome. *Annu. Rev. Biochem.* 83, 467–486.
- Zhang, G., and Ignatova, Z. (2009). Generic algorithm to predict the speed of translational elongation: Implications for protein biogenesis. *PLoS One* 4, e5036.
- Zhang, G., Hubalewska, M., and Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* 16, 274–280.
- Zhou, M., Wang, T., Fu, J., Xiao, G., and Liu, Y. (2015). Nonoptimal codon usage influences protein structure in intrinsically disordered regions. *Mol. Microbiol.* 97, 974–987.
- Zhou, T., Weems, M., and Wilke, C.O. (2009). Translationally optimal codons associate with structurally sensitive sites in proteins. *Mol. Biol. Evol.* 26, 1571–1580.
- Zhou, W., Karcher, D., and Bock, R. (2014). Identification of enzymes for adenosine-to-inosine editing and discovery of cytidine-to-uridine editing in nucleus-encoded transfer RNAs of Arabidopsis. *Plant Physiol.* 166, 1985–1997.
- Zinshteyn, B., and Gilbert, W. V. (2013). Loss of a conserved tRNA anticodon modification perturbs cellular signaling. *PLoS Genet.* 9, e1003675.
- Zinshteyn, B., and Nishikura, K. (2009). Adenosine-to-inosine RNA editing. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 1, 202–209.
- Zuo, W.-L., Yang, J., Gomi, K., Chao, I., Crystal, R.G., and Shaykhiev, R. (2017). EGF-Amphiregulin Interplay in Airway Stem/Progenitor Cells Links the Pathogenesis of Smoking-Induced Lesions in the Human Airway Epithelium. *Stem Cells* 35, 824–837.

