



HAL
open science

Diversité de la 2'-O-ribose méthylation des ARN ribosomiques dans les cancers : développement d'outils et méthodes d'analyses

Hermes Paraqindes

► **To cite this version:**

Hermes Paraqindes. Diversité de la 2'-O-ribose méthylation des ARN ribosomiques dans les cancers : développement d'outils et méthodes d'analyses. Cancer. Université Claude Bernard - Lyon I, 2022. Français. NNT : 2022LYO10218 . tel-04812763

HAL Id: tel-04812763

<https://theses.hal.science/tel-04812763v1>

Submitted on 1 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE de DOCTORAT DE
L'UNIVERSITÉ CLAUDE BERNARD LYON 1**

**École Doctorale N° 340
Biologie Moléculaire Interactive et Cellulaire**

Discipline : Bioinformatique

Soutenue publiquement le 16/12/2022 par :

Hermes PARAQUINDES

**Diversité de la 2'-O-ribose méthylation des ARN
ribosomiques dans les cancers :
développement d'outils et méthodes d'analyses**

Devant le jury composé de :

Céline BROCHIER-ARMANET, Professeure des universités, UCBL1
Alexandre DAVID, DR, INSERM
Christine GASPIN, DR, INRAE Toulouse
Olivier NAMY, DR, CNRS
Virginie MARCEL, DR, INSERM

Présidente
Examineur
Rapporteur
Rapporteur
Directrice de thèse

Résumé

Thèse de DOCTORAT DE L'UNIVERSITÉ DE LYON

Diversité de la 2'-O-ribose méthylation des ARN ribosomiques dans les cancers: développement d'outils et méthodes d'analyses

par Hermes PARAQUINDES

Le ribosome est une macromolécule cellulaire composé de 80 protéines ribosomiques (RP) et de 4 Acide Ribonucléique ribosomique (ARNr). Le ribosome joue un rôle essentiel dans la cellule : il traduit les ARN messagers (ARNm) en protéines. Pendant longtemps, le ribosome a été vu comme une macromolécule non-spécialisée, homogène et jouant un rôle passif dans la synthèse des protéines. Cependant, des études récentes indiquent que les ribosomes montrent une hétérogénéité de composition en fonction des conditions physio-pathologiques avec des activités traductionnelles distinctes. Ainsi, les ribosomes semblent être des acteurs directs de la régulation traductionnelle. L'hétérogénéité de composition du ribosome a été observée pour les deux éléments qui le constitue : les RP et les ARNr. L'activité catalytique des ribosomes est essentiellement portée par les ARNr qui portent des modifications chimiques, incluant environ 110 méthylations de ribose (2'Ome) et 95 pseudouridilations. L'équipe « Ribosome, Traduction et Cancer » a démontré pour la première fois dans des lignées cellulaires que la 2'Ome des ARNr est altérée à certains sites seulement au cours de l'initiation tumorale mammaire et que cette altération site-spécifique affecte directement la traduction de certains ARNm codant pour des oncogènes. Par ailleurs, dans les tumeurs mammaires, l'expression de la fibrillarine (FBL), la méthyltransférase catalysant la 2'Ome des ARNr, est altérée et représente un marqueur indépendant de mauvais pronostic. Ces données suggèrent que la 2'Ome des ARNr est altérée dans les cancers et induit une dérégulation de la traduction favorisant la tumorigenèse et la progression de la maladie. Cependant, ces observations reposent sur des preuves indirectes ou provenant principalement de modèles cellulaires. Aujourd'hui, il reste à établir que la 2'Ome des ARNr varie dans les cancers et à caractériser leur association avec des caractéristiques biologiques et cliniques particulières afin de comprendre leur rôle en cancérologie.

Nous avons utilisé la technologie innovante de RiboMethSeq afin d'analyser le taux de 2'Ome de l'ensemble des sites méthylés de l'ARNr en un seul séquençage dans deux modèles cancers : les cancers du sein et les gliomes. Grâce aux améliorations méthodologiques et bioinformatiques que nous avons apportées au RiboMethSeq, nous montrons pour la première fois que chez l'Homme et plus particulièrement dans les cancers, il existe une variabilité du niveau de 2'Ome au sein d'une tumeur (variabilité intra-tumorale : toutes les molécules d'ARNr ne sont pas méthylées) mais aussi entre les tumeurs (variabilité inter-patient). La variabilité de la 2'Ome n'est pas aléatoire, mais se fait sur un nombre restreint de sites avec seulement 40 % des sites qui présentent une forte variabilité inter-patients. Par ailleurs, nous avons identifié des sites dont le niveau de méthylation est significativement différent entre les caractéristiques biologiques et cliniques des cancers, comme le sous-type ou le grade tumoral. Parmi ces sites, certains sont communs à tous les types de cancer analysés (gliomes et sein) et d'autres sont spécifiques à un type de cancer.

Dans l'ensemble, nos résultats démontrent pour la première fois que la 2'Ome des ARNr varie dans les cancers. De plus, la 2'Ome des ARNr est une couche supplémentaire reflétant l'hétérogénéité des cancers. Ces résultats ouvrent la voie à l'utilisation du ribosome comme biomarqueur, utile pour le diagnostic et la classification moléculaire des cancers, ainsi que comme potentielle cible thérapeutique. Les outils méthodologiques et bioinformatiques dédiés au RiboMethSeq développés sous forme de pipeline nextflow et package R seront diffusés au sein de la communauté scientifique et faciliterons l'analyse de la 2'Ome des ARNr.

Mots clés : ARN ribosomique ; 2'-O-ribose méthylation ; RiboMethSeq ; bioinformatique ; hétérogénéité ; cancer

Remerciements

J'aimerais remercier tout d'abord les membres de mon jury de thèse : Céline Brochier-Armanet, Alexandre David, Christine Gaspin, Olivier Namy d'avoir accepté d'assister à ma soutenance de thèse et d'évaluer le travail effectué pendant ma thèse.

Merci en particulier à Madame Christine Gaspin et monsieur Olivier Namy d'avoir pris le temps de lire et évaluer ce manuscrit.

Merci aux membres de mon comité de suivi de thèse, Thierry Dubois, Franck Tirode et Philippe Veber, de m'avoir suivi et donné plein de bons conseils pendant ces trois dernières années.

Merci à la Ligue Contre le Cancer et à ses donateurs qui ont cru en mon projet de thèse et m'ont donné la possibilité de travailler sur ce sujet qui me passionne.

Un grand merci à ma super-chef : Virginie Marcel. Je sais que tu n'aimes pas que je t'appelle super-chef, mais je pense que tu le mérites. Tu incarnes le scientifique que j'aimerais être. Tu guides en donnant l'exemple et toujours avec le sourire. Je te serai toujours reconnaissant et j'espère t'avoir aidé à décortiquer quelques mystères du ribosome.

Pendant ces trois dernières années, j'ai eu la chance d'être accompagné par d'autres personnes sans qui cette thèse n'aurait pas été la même.

Merci à Janice Kielbassa qui a épaulé Virginie avec brio. Janice, tu m'as appris tellement de choses que je ne saurai pas par où commencer. Merci pour tous les conseils « on » et « off » le travail. Je te serai toujours reconnaissant.

Merci à Emilie Thomas qui a su nous guider dans ce projet. Emilie, je suis sûr que ce projet va grandir avec toi. Les biologistes du CRCL ont de la chance de pouvoir travailler avec toi. Merci.

J'aimerais remercier les deux chefs d'équipes, Jean-Jacques Diaz et Alain Viari, de m'avoir accueilli dans les meilleures équipes du CRCL. Jean-Jacques, merci pour ton écoute, ton temps et tes blagues. L'équipe RTC est à ton image. Alain, merci pour le temps que tu m'as consacré. J'ai toujours apprécié ces moments et tes explications de choses compliquées pour des personnes « simples » comme moi. Merci.

Un grand merci à l'équipe RTC (ou Diaz).

Vous être très nombreux et j'avoue que je risque d'oublier des gens. Donc je vous ferai part de mes remerciements en « live ». Sachez que je vous serais toujours reconnaissant de votre amitié et du temps passé ensemble. Difficile de trouver une meilleure équipe.

Merci à la plateforme de séquençage Gilles Thomas. J'ai appris beaucoup de choses grâce à vous. Hâte de vous retrouver autour d'un verre, d'un blind test, d'un badminton ou d'une pétanque raclette;).

Merci à la plateforme de génomique des cancers qui a fait un travail formidable et qui m'a toujours fourni des données de qualité.

J'ai souvent essayé de me rappeler que derrière les enchaînements des lettres A, C, T, G ou des chiffres 0 et 1 se cachent de vraies personnes, sans qui rien n'aurait de sens à mes yeux. Merci à vous.

Merci à mes amis, de loin ou de près. Merci d'être toujours là quand j'en ai besoin. Merci à mes amis biochimiste, Nico et Jorgo qui continuent à me supporter. À Alex et Eric qui comptent beaucoup. À Gestian qui va devoir m'aider à relâcher la pression peu après la fin de l'écriture de ce manuscrit. À Kostika avec qui j'ai hâte de chanter les sérénades de Korça. À Maxime et Anthony avec qui les jeux de sociétés ont beaucoup plus de sens.

Merci aux membres de EPEB qui m'ont beaucoup aidé à m'adapter à la vie en France et qui étaient toujours là quand j'avais besoin.

Merci à Ben et Jeanne, sans qui je ne serai jamais venu en France. Je ne compte même pas le nombre de fois où vous m'avez aidé. J'espère pouvoir vous rendre la pareille un jour.

À mon frère, ma sœur et mes parents qui me soutiennent malgré la distance et sont toujours là.

À Virginie, David, Sarah, Clémence et Brigitte qui font tout pour que je me sente chez moi.

Merci à la personne qui a toujours été à mes côtés ces trois dernières années, sans arrêt. À la personne avec qui j'ai hâte de continuer les prochains chapitres de ma vie avec plein de bons gâteaux et de bons moments. Ma vie est plus belle avec toi. Merci Caroline. Të dua. Cette thèse serait remplie de fautes et de figures très moches sans toi.

Merci à tous ceux que j'ai croisé pendant ces années.

Faleminderit.

Home wasn't a set house, or a
single town on a map. It was
wherever the people who loved
you were, whenever you were
together. Not a place, but a
moment, and then another,
building on each other like bricks
to create a solid shelter that you
take with you for your entire life,
wherever you may go.

What Happened to Goodbye
Sarah Dessen

...

Table des matières

Résumé	iii
Remerciements	v
Préambule	xxiii
1 Le(s) ribosome(s)	1
A La structure et la composition du ribosome	1
A.1 Introduction	1
A.2 Les composants du ribosome : les protéines ribosomiques et les ARNr	4
A.3 Les sous-unités et les centres fonctionnels du ribosome	7
A.4 La traduction	8
A.4.1 Recrutement de l'ARN messager par le ribo- some	10
B Hétérogénéité de composition du ribosome	16
B.1 Hétérogénéités des RP	17
B.1.1 Stœchiométrie des RP	18
B.1.2 Paralogie de RP	21
B.1.3 Les modifications post traductionnelles des RP	22
C Hétérogénéité de composition du ribosome liée aux ARNr . .	23
C.0.1 Hétérogénéité de séquences des ARNr	23
C.1 Les modifications post-transcriptionnelles des ARNr .	25
C.1.1 Les modifications de bases	26
C.1.2 Pseudourilation des ARNr	27
C.1.3 La 2'-O-ribose méthylation des ARNr	30
C.1.4 Revue : 2'O-Ribose methylation of rRNAs : Natural diversity in living organisms, biolo- gical processes and diseases	32

2	Ribosomes et Cancers	55
A	Le(s) cancer(s)	55
B	Biogenèse des ribosomes et cancer	59
B.1	Altération de la biogenèse des ribosomes dans les cancers	59
B.2	Hyperactivation de la biogenèse des ribosomes induite par des oncogènes et des suppresseurs de tumeurs clés	60
B.3	Ciblage thérapeutique de la biogenèse des ribosomes	61
C	Hétérogénéité de composition des ribosomes et cancer	62
C.1	Altération des RPs dans les cancers	62
C.1.1	Mutation dans les gènes codant les RPs, ribosomopathies et cancer	62
C.1.2	Altération de l'expression des RP dans le cancer	65
C.1.3	Conclusion	67
C.2	Altération des modifications chimiques des ARNr dans les cancers	68
C.2.1	Méthylation de bases des ARNr	68
C.2.2	Pseudouridilation des ARNr	70
C.2.3	La 2Ome des ARNr	73
3	Détection et quantification de la 2'Ome via les approches de séquençage haut-débit	77
A	Séquençage du transcriptome par RNA-seq	77
A.1	Matériel biologique	78
A.2	Séquençage RNA-seq Illumina	79
A.2.1	Préparation de la librairie de séquençage	80
A.2.2	Amplification	81
A.2.3	Séquençage	82
A.3	Planification expérimentale	82
A.4	Analyse des données de séquençage RNA-seq	85
A.4.1	Contrôle qualité	85
A.4.2	Alignement et quantification	86
A.4.3	Aspect analytique pour l'interprétation des données de RNAseq	87
B	Détection et quantification de la 2'-O-ribose méthylation des ARN	88
B.1	Techniques basées sur la propriété d'encombrement stérique induite par la présence d'une 2'Ome (propriété iv)	89

B.1.1	Primer extension (extension d'amorces)	89
B.1.2	2'Ome-seq	91
B.2	Techniques basées sur la propriété de résistance à l'oxydation et à la β -élimination induit par la présence d'une 2'Ome (propriété ii)	92
B.3	Technique basée sur la propriété de résistance à l'hydrolyse alcaline induite par la présence d'une 2'Ome (propriété i) : le RiboMethSeq	96
B.3.1	Principe du RiboMethSeq	96
B.3.2	RiboMethSeq et technologie Ion Torrent	97
B.3.3	RiboMethSeq et technologie Illumina	100
C	Méthodes d'analyses utilisées pour exploiter les données de RiboMethSeq issues d'échantillons humains	102
C.1	Analyses exploratoires des données	102
C.1.1	Visualisation des données quantitative avec boxplot	102
C.1.2	Analyse en composantes principales	103
C.1.3	Regroupement hiérarchique (hierarchical clustering)	106
C.2	Comparaison du taux de 2'Ome entre deux ou plusieurs groupes	108
4	Hypothèse et objectifs	111
5	Résultats	113
A	Riboscore : an R package for RiboMethSeq data analysis	113
A.1	Contexte	113
A.2	Résultats principaux	114
A.3	Ma contribution	114
A.4	Vignette du package R	115
B	Article I : snoDB 2.0 : An enhanced interactive database, specializing in human snoRNAs	131
B.1	Contexte	131
B.2	Résultats principaux	132
B.3	Ma participation	133
B.4	Article	134

C	Article II : Ribosomal RNA 2'O-methylation as a novel layer of inter-tumour heterogeneity in breast cancer	144
C.1	Contexte	144
C.2	Résultats principaux	145
C.3	Ma participation	146
C.4	Article	146
D	Article III : Ribosome biogenesis and rRNA 2'O-ribose methylation : Uncovering specific alterations in glioma subtypes . .	181
D.1	Contexte	181
D.2	Résultats principaux	182
D.3	Ma participation	183
D.4	Article	183
6	Discussion	221
A	Développement de méthodologies et d'outils dédiés au RiboMethSeq	221
A.1	Évolution de la technique de RiboMethSeq et de ses outils	221
A.2	Utilisation du RiboMethSeq dans d'autres espèces . . .	225
A.3	Utilisation du RiboMethSeq pour d'autres applications?	228
B	Hétérogénéité de la 2'Ome des ARNr dans les cancers	232
B.1	Démonstration de l'existence d'une hétérogénéité de la 2'Ome des ARNr dans les cancers	232
B.2	Quel rôle de la 2'Ome dans les fonctions biologiques?	234
B.3	Quel rôle de la 2'Ome dans la traduction?	237
7	Conclusion générale	239
A	Annexe	241
A	Liste des publications, communications orales, posters et collaborations	241
A.1	Publications acceptées	241
A.2	Publications en préparation	242
A.3	Communications orales	242
A.4	Posters	242
B	Article : Uncovering the Translational Regulatory Activity of the Tumor Suppressor BRCA1	244
B.1	Contexte	244
B.2	Résultats principaux	244

B.3 Ma contribution 245
B.4 Article 245

Bibliographie **298**

Table des figures

1.1	Structure du ribosome	2
1.2	Biogenèse des ribosomes	6
1.3	Le processus de la traduction	9
1.4	Régulation de la traduction	14
1.5	Hétérogénéité de composition du ribosome.	18
2.1	Incidence et mortalité des cancers dans le monde.	55
2.2	Les caractéristiques des cancers	58
2.3	snoARN dérégulés dans les cancers	73
3.1	Séquençage Ion Torrent	80
3.2	Principe du séquençage Illumina	83
3.3	Plan expérimental des analyses OMICS	84
3.4	Méthodes de détection et quantification de la 2'Ome.	95
3.5	Principe du RiboMethSeq	96
3.6	Calcul du taux de 2'Ome (Cscore)	99
3.7	Representation boxplot	104
3.8	Pourcentage de variance expliqué par chaque composante prin- cipale	106
3.9	Exemple de visualisation de l'ACP	106
3.10	Exemple de dendrogramme	108
3.11	Exemple de heatmap	109
6.1	Pipeline de traitement des données de séquençage de Ribo- MethSeq	224
6.2	Quantification de l'expression des snoARN à l'aide des don- nées de RiboMethSeq	231
6.3	Comparaison des sites les plus variants dans les cancers	234
6.4	Variabilité de la 2'Ome dans les cancers du sein TNBC	235
6.5	Clusterisation des échantillons TNBC sur la base des profils de 2'Ome	236

6.6 Analyse de GSEA 236

Liste des abréviations

2'Ome	Méthylation des riboses en position 2'
5-FU	5-Fluorouracil
Ψ	Pseudouridylation
A	le site aminoacyl
ACP	Analyse en Composantes Principales
ac4C	N4-acetylcytidine
ADN	Acide DéoxyriboNucléique
ARN	Acide ribonucléique
ARNm	ARN m essenger
ARNnc	ARN n on codant
ARNr	ARN ribosomique
ARNt	ARN de t ransfert
BRCA1	Breast Cancer 1
cryo-EM	Cryo-microscopie électronique
DBA	L'anémie de D iamond- B lackfan
DC	Centre de D écodage
DKC1	Enzyme D yskérine
dNTP	Déoxyribonucléotide triphosphate
E	le site de sortie
eIF	Facteurs d'initiation
ENCODE	Encyclopedia of DNA Elements
ER	Œstrogènes
ES	Segments d'expansion
FBL	Methyltransférase Fibrilarine
FFPE	extraits de tissus fixés au formol et inclus en paraffine
GBM	Glioblastome multiforme
GDP	Guanosine diphosphate
GTP	Guanosine triphosphate
HC	Regroupement hiérarchique

HER2	Human Epidermal Growth Factor Receptor-2
IDH	Isocitrate déshydrogénase
iMet-ARNt	ARNt initiateur
IQR	Écart interquartile
IRES	Internal ribosome entry site
LAL	Leucémie aiguë lymphoblastique
LSU	Large SubUnit
MAPK	Mitogen-Activated Protein Kinase
METTL5	Methyltransferase-like 5
mTOR	mammalian Target of Rapamycin
MDa	millions de daltons
NAR	Nucleic Acids Research
NSUN5	Putative methyltransferase NSUN5
Nop56	Protéine nucléolaire 56
Nop58	Protéine nucléolaire 58
P le site peptidyl	
PABP	PolyA Binding Protein
PC	Composante principale
PCR	Polymerase Chain Reaction
PR	Progestérone
PTC	Centre peptidyl-transférase
PTM	Modifications post-traductionnelles
PolI	ARN polymérase I
PolII	ARN polymérase II
PolIII	ARN polymérase III
Q1	Premier quartile
Q3	Troisième quartile
QC	Contrôle qualité
RIN	RNA Integrity Number
RIP-seq	Séquençage d'arn par immunoprecipitation
RNA-seq	Séquençage du transcriptome
RP	Protéines ribosomiques
RPL	Protéines ribosomiques de la grande sous-unité
RPS	Protéines ribosomiques de la petite sous-unité
RT	Transcriptase inverse
RT-qPCR	quantitative reverse transcription PCR

S	unité de Svedberg
SD	Écart type
SDS	Syndrome de Shwachman-Diamond
snoARN	petits ARN non codant
SNORA	petits ARN non codant à boîte H/ACA
SNORD	petits ARN non codant à boîte C/D
SSU	Petite sous-unité
STC	Syndrome de Treacher Collins
TCGA	Cancer Genome Atlas
TNBC	Triple-negative subtype
UV	Ultraviolet
X-DC	Dyskératose congénitale liée à l'X

*Somewhere, something incredible
is waiting to be known*

Carl Sagan

Préambule

La crise sanitaire due à la COVID-19 a fait entrer un nouveau terme scientifique dans le langage courant : l'ARN messenger. Le terme « ribosome » reste quant à lui peu connu du grand public et même de la communauté scientifique. Or, le ribosome est l'acteur principal de la traduction des ARNm en protéines et donc essentiel pour comprendre le fonctionnement de nos fameux vaccins à ARNm. En effet, chaque protéine qui est la brique même de l'ensemble des cellules vivantes de toutes les espèces sur Terre est produite par le ribosome lui-même et une cellule peut contenir jusqu'à 10 millions de ribosomes. Ainsi, le ribosome a la lourde responsabilité de produire correctement et efficacement toutes les protéines de la cellule à partir de ces fameux ARNm. Il est difficile d'imaginer comment un composant aussi important pour la vie d'une cellule puisse être modifié ou altéré sans conséquence majeure. C'est pour cette raison que le ribosome a été considéré dès sa découverte comme une structure stable et homogène, ayant un simple rôle d'effecteur de la traduction. Cependant, au cours de ces dernières années, plusieurs études font état de caractéristiques inattendues du ribosome. Sa composition serait hétérogène au sein d'un organisme, d'un tissu, d'un type cellulaire ou même d'une pathologie. Le ribosome serait un acteur direct de l'efficacité, la spécificité et la fidélité de traduction.

C'est dans ce contexte que mon projet de thèse s'est inscrit avec pour objectif de mieux comprendre une des sources de l'hétérogénéité de composition du ribosome : la 2'-O-ribose méthylation des ARN ribosomiques. Plusieurs questions seront abordées dans ce manuscrit. Comment mesurer la 2'-O-ribose méthylation en utilisant des technologies de séquençage haut-débit ? Est-ce que la 2'-O-ribose méthylation varie chez l'Homme et plus particulièrement dans le contexte pathologique particulier qu'est le cancer ? Est-ce que cette variabilité est aléatoire ou au contraire reflète des caractéristiques biologiques et cliniques des cancers ?

Ma thèse est organisée en trois sections principales.

L'introduction se focalise tout d'abord (Chapitre 1) sur le ribosome, notamment sur sa structure et son fonctionnement tel qu'on le connaît aujourd'hui et dans un deuxième temps, sur le concept émergent de l'hétérogénéité de composition du ribosome et son impact sur la régulation de la traduction.

Un effort particulier sera fait sur la 2'-O-ribose méthylation des ARN ribosomiques et son rôle dans l'hétérogénéité de composition du ribosome. Ensuite (Chapitre 2), les liens connus aujourd'hui entre ribosome et cancer seront décrits, en traitant d'un côté les altérations quantitatives du ribosome, liées au fait que les cellules cancéreuses ont besoin de produire rapidement et en grande quantité les protéines qui permettent leur prolifération; et de l'autre les altérations qualitatives, c'est-à-dire, l'hétérogénéité de composition du ribosome dans les cancers. Enfin (Chapitre 3), la technologie de séquençage haut-débit qui permet d'identifier et de quantifier les altérations de la 2'-O-ribose méthylation et qui est au cœur de ma thèse, le RiboMethSeq, sera décrite ainsi qu'une courte description des méthodes d'analyse de données utilisées dans cette thèse.

Les résultats sont présentés sous forme d'article scientifique publié ou en cours de préparation et sont scindés en deux parties. La première partie est consacrée aux développements des méthodes et des outils dédiés à l'analyse de la 2'-O-ribose méthylation des ARNr. La deuxième partie se consacre à l'analyse de la 2'-O-ribose méthylation des ARNr dans les cancers du sein et les gliomes.

Enfin, les résultats de ma thèse seront discutés et replacés dans le contexte plus large qu'est le concept émergeant de l'hétérogénéité de composition du ribosome.

1 Le(s) ribosome(s)

A La structure et la composition du ribosome

A.1 Introduction

Le ribosome est la micro-machinerie cellulaire qui synthétise les protéines à partir des acides aminés portés par les ARN de transfert (ARNt) en décodant l'ARN messager (ARNm). Sa découverte remonte à la Seconde Guerre Mondiale. En 1941, Albert Claude a isolé des granules cytoplasmiques enrichis en ARN à partir de cellules animales (CLAUDE, 1941) qu'il appellera « microsomes » (CLAUDE, 1943) des mots grecs μικρός mikros et σώμα soma qui signifient respectivement « petit » et « corps ». Au même moment, les études de Caspersson (CASPERSSON, 1939, 1941) et Brachet (BRACHET, 1941) ont montré qu'il existe une forte corrélation entre la localisation de l'ARN dans le cytoplasme et l'intensité de la synthèse protéique. C'était le premier lien indirect entre les microsomes et la synthèse protéique. Plus tard, Schachman *et al* ont montré que ces particules sont composées d'ARN et de protéines (SCHACHMAN, PARDEE et STANIER, 1952). L'amélioration des techniques de microscopie électronique, permettant notamment d'étudier directement le contenu des cellules, a permis à Palade de publier les premières images des microsomes. Il a ainsi montré qu'ils sont très abondants dans une cellule (PALADE, 1955). De plus, ces images ont permis de constater que les microsomes étaient soit libres dans le cytoplasme, soit attachés au réticulum endoplasmique, un compartiment subcellulaire impliqué dans la sécrétion protéique. Les expérimentations de Keller, Zamecnick et Littlefield ont démontré pour la première fois le lien direct entre les microsomes et l'incorporation d'acides aminés radioactifs dans les protéines nouvellement synthétisées (KELLER et J W LITTLEFIELD, 1957; John W LITTLEFIELD et al., 1955). Ce sera finalement en 1958, lors du premier congrès consacré aux microsomes,

que Roberts propose le nom de « ribosome » pour décrire ces particules ribonucléoprotéiques (SOCIETY. , BIOPHYSICAL et ROBERTS, 1958). Depuis, le ribosome a été largement étudié pour son rôle crucial permettant la synthèse des protéines au sein de la cellule.

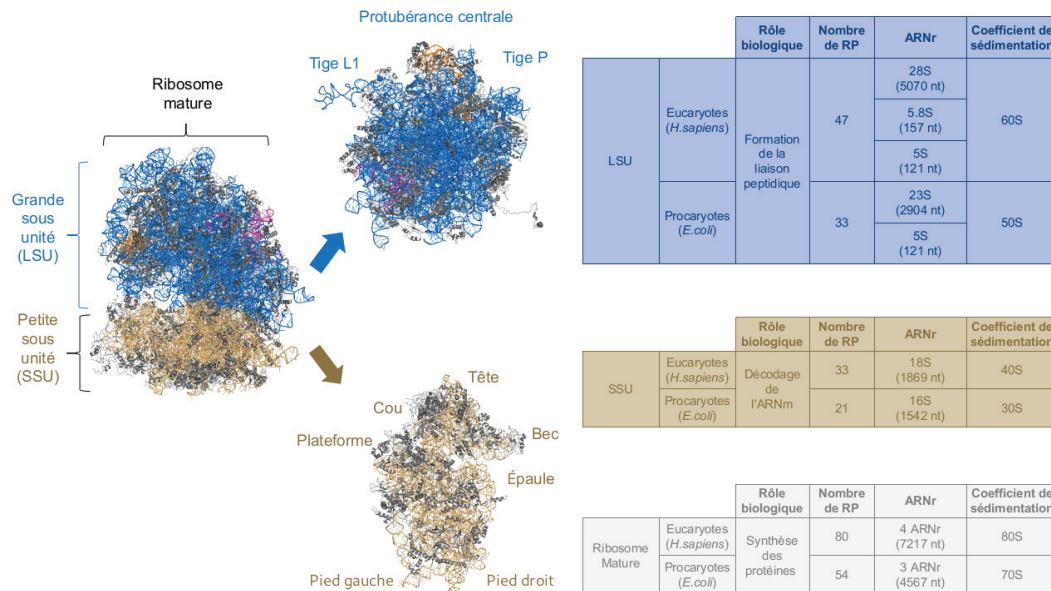


FIGURE 1.1 – Structure du ribosome. À gauche la structure 3D du ribosome humain (PDB :4UG0) composé de deux sous-unités : la grande sous-unité (LSU) et la petite sous-unité (SSU). Les protéines sont colorées en noir, l'ARNr 28S en bleu, l'ARNr 5S en orange, l'ARNr 5.8S en rose et l'ARNr 18S en jaune. Au milieu la morphologie des sous-unités et à droite un tableau récapitulatif des sous-unités chez l'Homme et la bactérie *E.coli*.

Les analyses de biochimie et de microscopie structurale ont permis la comparaison des ribosomes issus de différents organismes provenant de tous les domaines du vivant. On a pu ainsi constater une forte conservation de sa composition et de sa structure (Fig : 1.1). Quelle que soit son origine, chaque ribosome est ainsi composé de deux sous-unités de tailles différentes : la petite sous-unité (SSU : Small SubUnit) et la grande sous-unité (LSU : Large SubUnit). Chaque sous-unité est composée de protéines ribosomiques (RP : Ribosomal Proteins) et d'ARN ribosomiques (ARNr). Dans la nature, à travers les différents domaines du vivant, on trouve principalement deux types de ribosomes : le ribosome procaryote et le ribosome eucaryote, qui présentent des tailles différentes. La taille des ribosomes, des sous-unités ribosomiques et des ARNr est généralement désignée par leur coefficient de sédimentation exprimé en unité de Svedberg (S). Ainsi le ribosome entier est appelé

70S chez les procaryotes et 80S chez les eucaryotes. Même s'il existe une structure globale très similaire, les deux types de ribosomes diffèrent dans leurs compositions en RP et ARNr ainsi que dans leurs fonctions. Le ribosome des eucaryotes est environ 40 % plus grand que celui des procaryotes, qui comprend par exemple les bactéries. Cette différence est due à la présence d'éléments supplémentaires d'ARNr appelés segments d'expansion (ES) mais aussi de protéines ribosomiques, que ce soit des fragments d'acide aminé supplémentaires ou de nouvelles protéines ribosomiques (SPAHN et al., 2001). Par conséquent, le poids moléculaire varie entre 2,3 millions de daltons (MDa) chez les bactéries jusqu'à 4,3 MDa chez l'Homme (Fig : 1.1). Chez les procaryotes, le ribosome possède deux fois plus d'ARNr que de protéines ribosomiques en termes de masse moléculaire. Ce rapport ARNr/RP est différent en fonction des espèces eucaryotes où l'on observe que les ribosomes ont une masse moléculaire en RP plus importante qu'en ARNr.

Au niveau structural, alors que les segments d'expansion et certaines RP spécifiques aux eucaryotes se situent sur la couche externe du ribosome, le centre du ribosome correspond à une région fortement conservée au cours de l'évolution. En effet, cette région contient les centres catalytiques du ribosome répartis dans les deux sous-unités. De plus, le ribosome possède trois sites qui permettent la liaison aux ARNt : le site aminoacyl (A), dont l'ARNt porte un aminoacyl ; le site peptidyl (P), dont l'ARNt porte le peptide néosynthétisé ; et le site de sortie (E pour exit), dont l'ARNt est dépourvu d'aminoacyl ou du peptide en cours de synthèse. Ces trois sites sont localisés à l'interface des deux sous-unités du ribosome. Deux de ces sites possèdent un centre catalytique : la SSU possède le centre de décodage (DC pour Decoding Center) où a lieu le décodage de l'ARNm et la LSU contient le centre peptidyl-transférase (PTC pour Peptidyl-Transferase Center) où est catalysée la liaison peptidique entre un nouvel acide aminé et le néo-peptide.

Comme identifié dès sa découverte, la plupart des ribosomes se retrouvent dans le cytoplasme sous forme libre ou attaché au réticulum endoplasmique. De plus, dans une cellule eucaryote, on retrouve différentes populations de ribosome, dans le cytoplasme, dans les mitochondries et dans les chloroplastes chez les plantes. Ces ribosomes diffèrent des ribosomes retrouvés dans le cytoplasme par leur taille, généralement réduite, mais également par leur rôle biologique.

Tout au long de cette thèse, je vais me focaliser sur le ribosome eucaryote cytoplasmique et en particulier humain, sauf indication contraire.

A.2 Les composants du ribosome : les protéines ribosomiques et les ARNr

Au début des années 1970, il était admis que les RP portaient l'activité catalytique du ribosome et que les ARNr étaient uniquement un support de structure pour assurer le bon repliement des RP. Ce point de vue a changé dans les 1980, quand plusieurs études ont montré que certains ARNr possèdent des activités catalytiques et donc peuvent jouer le rôle d'enzymes (GUERRIER-TAKADA et al., 1983; KRUGER et al., 1982). Barta *et al* ont montré que l'ARNr de la grande sous-unité était directement impliqué dans la formation de la liaison peptidique au cours du processus de la traduction (BARTA et al., 1984). De plus, les études structurales du ribosome ont montré que les centres catalytiques étaient entourés principalement d'ARNr et que les RP étaient localisées à la surface du ribosome. Ces résultats ont permis de confirmer que le ribosome est un ribozyme, ce qui signifie que ce sont les ARNr qui possèdent l'activité catalytique du ribosome. Cependant, les RP sont nécessaires pour l'assemblage et le bon fonctionnement du ribosome. La résolution de la structure à haute définition du ribosome procaryote et de la fonction des sous-unités a valu le prix Nobel de chimie à trois scientifiques en 2009 (T. Steitz, V. Ramakrishnan, A. Yonath)¹. Ce n'est qu'en 2011 que Ben-Shem *et al* ont pu obtenir la structure cristallographique du ribosome eucaryote (chez la levure) à une résolution de 3 angströms (BEN-SHEM et al., 2011). En 2013, Anger *et al* ont publié au même moment les structures du ribosome de drosophile (*Drosophila melanogaster*) et humain à une résolution de 6 angströms (ANGER et al., 2013). Des données structurales d'une résolution de 3 angströms par cryo-microscopie électronique (**cryo-EM**) ont ensuite été obtenues par une équipe de l'Institut de Génétique et de Biologie Moléculaire à Strasbourg (Heena KHATTER et al., 2015). Ces résultats ont permis d'avoir un aperçu à haute résolution des ARNr et des protéines.

Les ribosomes eucaryotes contiennent 33 RP dans la SSU et 47 dans la LSU (Fig : 1.1). Les gènes codant pour les RP sont transcrits en ARNm par l'ARN polymérase II dans le noyau. Les ARNm avant leur maturation contiennent

1. <https://www.nobelprize.org/prizes/chemistry/2009/summary/>

des introns qui sont éliminés lors de l'épissage. Certains de ces introns contiennent de petits ARN non codant appelés snoARNs, impliqués dans la modification chimique des ARNr (cf. C.1.2, C.1.3). Les ARNm matures des RP sont traduits en protéines dans le cytoplasme comme les autres ARNm, par les ribosomes. Les RP peuvent subir à leur tour des modifications post-traductionnelles comme l'addition de méthylation ou de phosphorylation (ODINTSOVA et al., 2003; Y. YU et al., 2009). Les RP matures sont ensuite importées dans le noyau, au niveau du nucléole où a lieu le processus d'assemblage des sous-unités du ribosome (Fig : 1.2). Chez l'Homme, les RP ont une masse moléculaire moyenne de 18,877 Da et une longueur de 167 acides aminés. Elles ont une charge globale positive ce qui leur permet d'interagir facilement avec les ARN qui sont chargés négativement.

En plus des RP, un ribosome eucaryote contient quatre molécules d'ARNr distinctes, dont une dans la SSU (18S) et trois dans la LSU (28S, 5.8S et 5S) (Fig : 1.1). Ces ARN sont codés par des gènes qui présentent une organisation particulière. Le génome humain contient en effet entre 200 et 400 copies de gènes d'ADN ribosomique (ADNr) qui sont répétées en tandem sur 5 chromosomes (13, 14, 15, 21 et 22) (STULTS et al., 2007). Chaque ADNr code pour un pré-ARNr 47S qui est par la suite clivé, d'abord en pré-ARN 45S et ensuite en 3 ARNr : 18S, 5.8S et 28S (Fig : 1.2). La quatrième molécule d'ARNr, le 5S, est codé par un gène distinct, aussi retrouvé en plusieurs copies dans le génome, mais principalement dans le chromosome 1 (GIBBONS et al., 2015). La transcription de ces 4 ARNr peut compter pour plus de 60 % de l'activité énergétique totale de la cellule (MOSS et al., 2006; Jonathan R WARNER, 1999). L'ARNr 5S est transcrit par l'ARN polymérase III, alors que le pré-ARNr 47S par l'ARN polymérase I (Fig : 1.2). De plus, le pré-ARNr 47S est transcrit et mûri dans le nucléole, alors que l'ARN 5S est transcrit en dehors du nucléole. Les ARNr subissent des modifications chimiques, qui peuvent être ajoutées de manière co-transcriptionnelle ou post-transcriptionnelle. Parmi les modifications chimiques on peut mentionner les modifications de base comme la méthylation, l'acétylation, l'aminocarboxypropylation qui sont catalysées par des enzymes spécifiques, ou des modifications ajoutées par des complexes ribonucléoprotéiques contenant des snoARN comme les pseudouridylations (cf. C.1.2) et les 2'-O-ribose méthylations (cf C.1.3).

La biogenèse des ribosomes commence de manière co-transcriptionnelle avec les ARNr (CRUZ, KARBSTEIN et WOOLFORD, 2015). En effet, l'assemblage

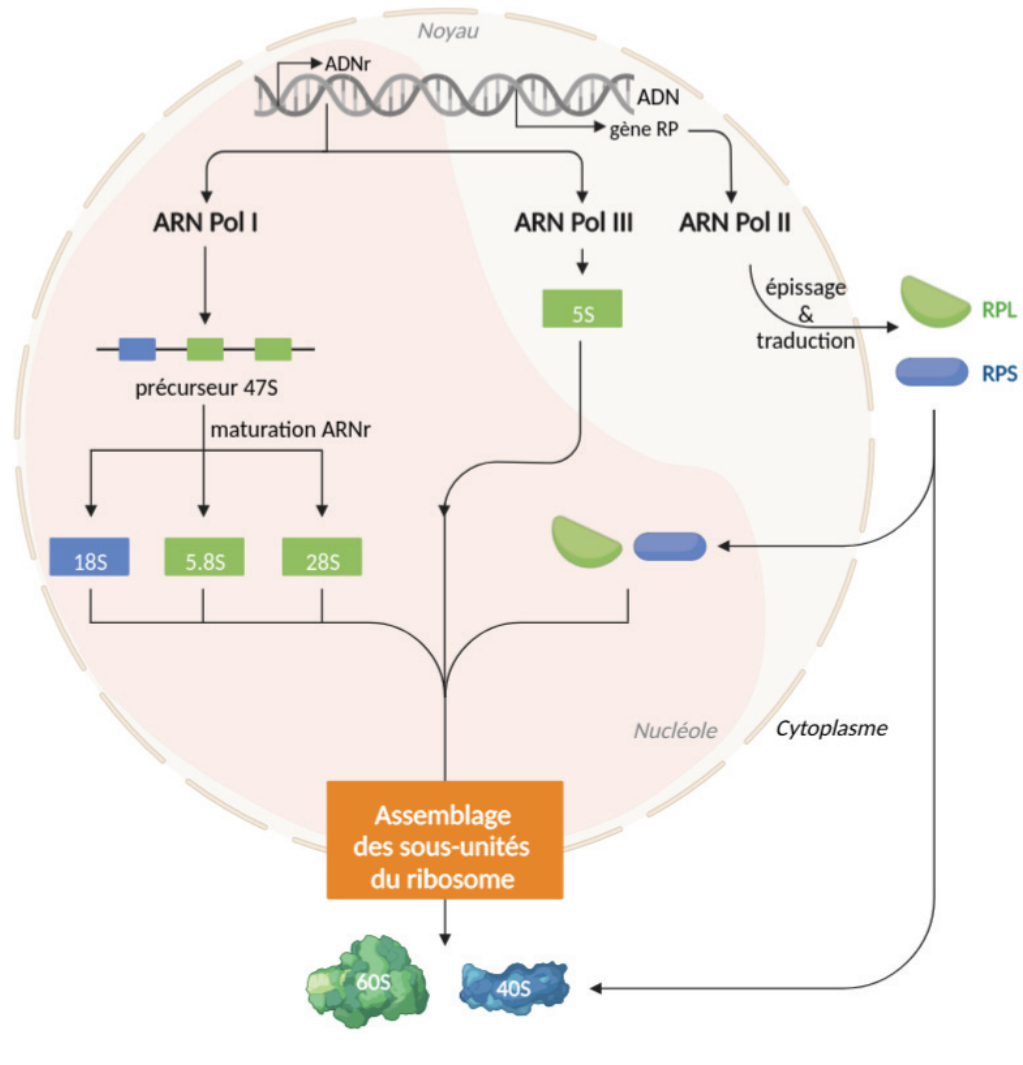


FIGURE 1.2 – Biogenèse des ribosomes. La biogenèse du ribosome requiert les 3 ARN polymérase, et plus de 300 facteurs. Elle dépend principalement de la disponibilité des ARNr. Les gènes de l'ADNr sont répartis sur les cinq chromosomes acrocentriques et s'agrègent au moment de la transcription pour former le nucléole. Le précurseur 47S pré-ARNr est synthétisé par l'ARN polymérase I, puis maturé pour former les ARNr 5.8S, 18S et 28S. L'ARN 5S, lui, est synthétisé par l'ARN polymérase III, et les protéines ribosomiques par l'ARN polymérase II. Le processus commence dans le nucléole et noyau et fini dans le cytoplasme. (Adapté de (MORIN et al., 2022))

des ARNr et des protéines commence dans le nucléole pour finir dans le cytoplasme (Fig : 1.2). Ce processus très complexe fait intervenir plus de 200

facteurs d'assemblage. Brièvement, les RP et l'ARNr 5S sont importés du cytoplasme vers le nucléole afin de commencer l'assemblage des particules pré-ribosomiques de la petite sous-unité (pré-40S) et de la grande sous-unité (pré-60S). Les particules pré-ribosomiques sont ensuite exportées dans le noyau puis dans le cytoplasme pour finir leur assemblage et leur maturation jusqu'à obtenir des ribosomes matures fonctionnels (KLINGE et WOOLFORD, 2018; THOMSON, FERREIRA-CERCA et HURT, 2013).

A.3 Les sous-unités et les centres fonctionnels du ribosome

Dans une cellule eucaryote, la petite sous-unité 40S du ribosome est composée de 33 RP (nommée **RPS**) et d'un ARNr d'une longueur de 1869 nucléotides chez l'Homme : l'ARNr 18S. Cette sous-unité est responsable du recrutement du ribosome sur l'ARNm et du décodage de l'information génétique portée par l'ARNm. Structuellement, la SSU est divisée en différentes régions qui comprennent la tête, le cou, l'épaule, le corps, la plateforme et les pieds gauche et droit du ribosome (Fig : 1.1). L'organisation structurale est principalement définie par le repliement de l'ARNr 18S. Les sites de liaison de l'ARNm et de l'ARNt sont situés à la surface de la SSU. La tête et l'épaule forment un tunnel pour l'entrée de l'ARNm et la sortie se fait entre la tête et la plateforme (YUSUPOVA et al., 2001).

Lors de l'étape de décodage, le codon de l'ARNm et l'anticodon de l'ARNt sont appariés dans le centre de décodage (**DC**). Le DC est situé dans la SSU au niveau de l'interface avec la LSU et à proximité du site A. Il a pour rôle de vérifier le bon appariement codon-anticodon. En effet, une interaction tripartite se met ici en place, impliquant les ARNm : ARNt : ARNr. L'ARNr a donc un rôle crucial dans le décodage. De plus, les modifications chimiques de l'ARNr jouent un rôle essentiel dans le décodage en maintenant la structure très particulière de l'ARNr dans le DC. L'interaction ARNm : ARNt : ARNr induit un changement de conformation de l'aminoacyl-ARNt dans le site A qui positionne l'acide aminé à proximité du centre PTC dans la LSU.

La grande sous-unité 60S du ribosome est composée de 47 RP et de 3 ARNr : 5S, 5.8S et 28S d'une longueur de 121, 158 et 5070 nucléotides respectivement. La grande sous-unité est divisée en plusieurs régions qui sont la protubérance centrale, la tige L1, la tige P et le tunnel de sortie (exit tunel) (Fig : 1.2). Les trois sites de liaison à l'ARNt (A, P et E) et le PTC se situent du côté de

l'interface avec la petite sous-unité. Le centre PTC est situé à proximité des sites de liaison à l'ARNt A et P et est constitué principalement de l'ARNr 28S (BASHAN et al., 2003 ; RIVAS et FOX, 2020).

Le rapprochement des ARNt localisés dans les sites A et P de la LSU permet aux ARNr situés dans le PTC de catalyser la liaison peptidique entre le peptidyl-ARNt du site P et l'aminoacyl-ARNt du site A, d'où le qualificatif de ribozyme. Pendant cette réaction, l'ARNt du site P est dé-acétylé et la chaîne polypeptidique se retrouve alors liée à l'ARNt du site A. Il est à noter que cette région est riche en modifications chimiques de l'ARNr 28S. Par ailleurs, le centre PTC est à proximité du tunnel de sortie. Pendant la traduction, la chaîne peptidique en cours de synthèse sort du ribosome via le tunnel de sortie avant d'être libéré pour continuer son processus de maturation et de repliement. L'organisation structurale et les capacités de ribozyme de l'ARNr sont donc essentielles pour assurer la fonction du ribosome : synthétiser les protéines à partir des ARNm.

A.4 La traduction

Si le génome porte l'information génétique, c'est le ribosome qui assure la synthèse des éléments qui donnent l'identité et la fonction d'une cellule : les protéines. En effet, les ARNm matures sont pris en charge par le ribosome afin d'être traduit en protéines. Ce processus complexe et très énergivore pour la cellule est appelé traduction. La traduction comprend trois étapes qui sont l'initiation (recrutement du ribosome sur l'ARNm), l'élongation (synthèse protéique) et la terminaison (arrêt de la synthèse et recyclage du ribosome) (Fig : 1.3). Pendant l'initiation, le ribosome avec un ARNt initiateur (iMet-ARNt) dans le site P est recruté sur l'ARNm à l'aide des facteurs d'initiation de la traduction. Ainsi pendant l'initiation, il y a l'association du ribosome sur l'ARNm, la sélection de l'ARNt initiateur et la sélection du codon initiateur. À la fin de l'étape d'initiation, le ribosome est prêt pour accepter le prochain ARNt et former la première liaison peptidique, ce qui marque aussi le début de l'étape de l'élongation. L'élongation étant le processus de synthèse de la chaîne peptidique par le ribosome et donc du décodage de l'ARNm. Ce processus s'arrête lors de l'apparition d'un codon stop. Après la reconnaissance de ce codon stop de l'ARNm par le ribosome, la chaîne peptidique formée est libérée.

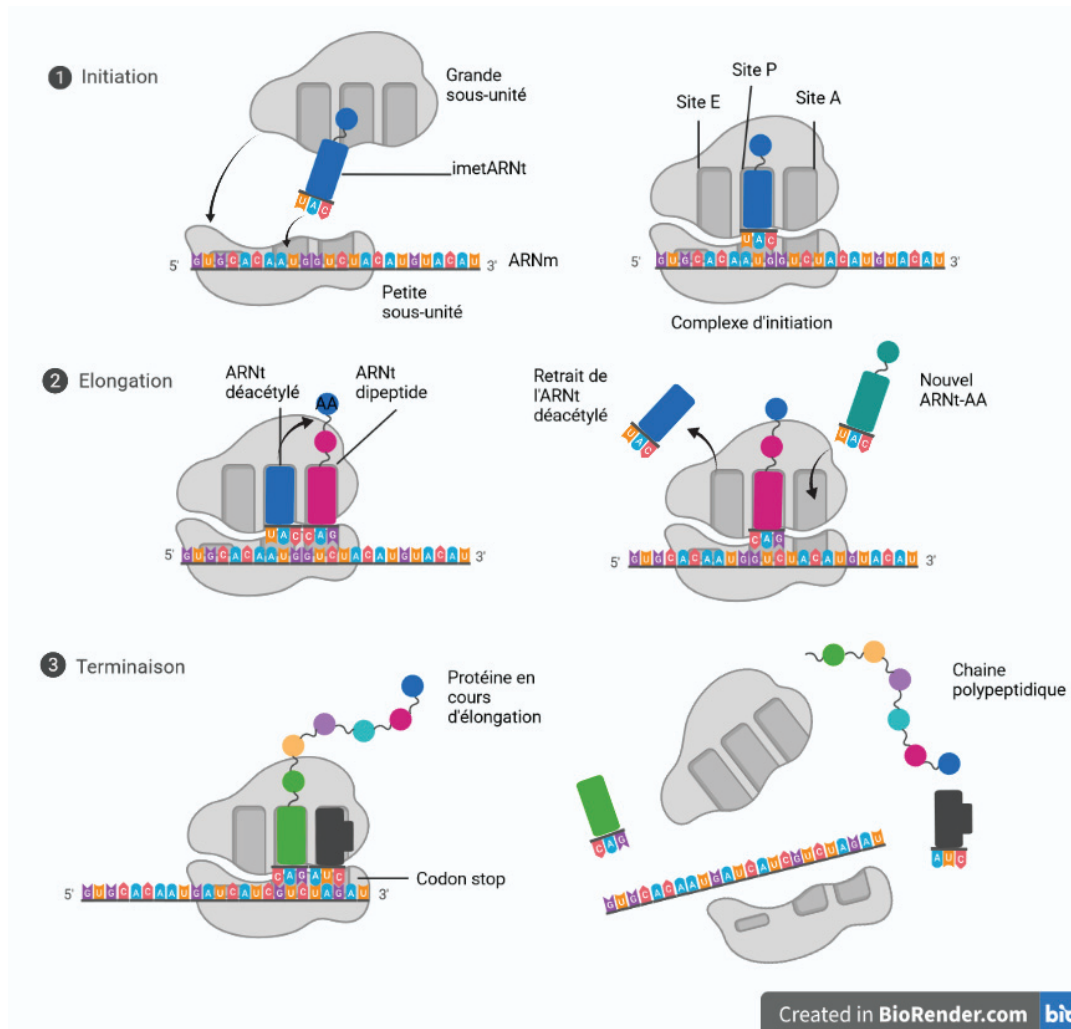


FIGURE 1.3 – La traduction comprend trois étapes qui sont l'initiation, l'élongation et la terminaison. Pendant l'initiation, le ribosome avec ARNt initiateur (imetARNt) dans le site P est recruté sur l'ARNm. L'élongation permet au ribosome de décoder l'ARNm. La terminaison a lieu lorsqu'un codon STOP est rencontré. La chaîne peptidique nouvellement synthétisée est libérée. Adapté de www.biorender.com

La traduction repose essentiellement sur le ribosome qui porte les activités catalytiques de traduction, mais aussi sur des facteurs canoniques de la traduction. Par exemple, lors du décodage de l'ARNm par l'ARNt et l'ARNr, l'aminocyl-ARNt est intégré dans le site A du ribosome via le facteur eEF1a-GTP, de manière que l'anticodon de l'ARNt soit complémentaire du codon de l'ARNm. Les ARNt sont cruciaux dans cet appariement. Les ARNt portent

l'un des 20 acides aminés attachés par une liaison ester à leur extrémité 3'-OH. Le bon appariement codon-anticodon produit un changement de conformation de eEF1a et l'hydrolyse du GTP en GDP. Le complexe eEF1a-GDP est libéré pour permettre l'intégration totale de l'aminocyl-ARNt dans le site A et positionner l'acide aminé près du centre PTC pour qu'il soit pris en charge par la LSU. Cette grande sous-unité catalyse ensuite la liaison peptidique entre l'acide aminé porté par aminocyl-ARNt du site A et la chaîne peptidique portée par le peptidyl-ARNt du site P. La formation de la liaison peptidique provoque une rotation, appelée translocation, de la SSU par rapport à la LSU. Cette translocation libère le site A qui peut ainsi de nouveau recruter le prochain aminocyl-ARNt correspondant. Le site P contient la chaîne peptidique liée à l'ARNt et le site E contient l'ARNt dé-acétylé. Au prochain cycle, l'ARNt dé-acétylé sera libéré du ribosome via le site E. Ce cycle se répète jusqu'à ce que le DC reconnaisse un codon STOP (UAA, UAG, UGA), où les facteurs de terminaison de la traduction comme eRF1 et eRF3 se lient au site A en induisant un changement de conformation du centre PTC et libèrent ainsi la chaîne peptidique synthétisée (JACKSON, HELLEN et PESTOVA, 2012).

Les protéines sont les éléments essentiels d'une cellule puisqu'elles définissent son identité et sa fonction. Ainsi, la synthèse des protéines est régulée au niveau basal. De plus, la traduction est un des processus le plus énergivore du métabolisme cellulaire et par conséquent elle doit être hautement régulée. Le taux de traduction est principalement régulé lors de l'étape de l'initiation. En régulant cette étape, la synthèse protéique peut être modulée rapidement. La régulation du taux de synthèse via l'élongation ou la terminaison est très coûteuse en énergie pour la cellule, même si certains mécanismes existent dans la littérature, mais sont moins nombreux et moins étudiés. La partie suivante sera dédiée à la description brève du processus d'initiation de la traduction avant d'aborder les mécanismes de régulation.

A.4.1 Recrutement de l'ARN messager par le ribosome

L'initiation de la traduction est l'étape pendant laquelle le ribosome est recruté sur l'ARNm et où commence le processus de décodage de l'ARNm (Fig : 1.3). L'objectif de l'initiation est d'assembler les deux petites sous-unités du ribosome en un ribosome mature 80S qui contient l'ARNt initiateur (généralement ARNt méthionine initiateur ou iMet-ARNt) sur le codon d'initiation (généralement AUG) de l'ARNm. Il existe deux grands mécanismes

d'initiation de la traduction chez les cellules eucaryotes : « coiffe dépendante » et « coiffe indépendante » (Fig : 1.4).

A.4.1.1 Initiation de la traduction par le mécanisme «coiffe dépendante»

La quasi-totalité des ARNm matures eucaryotes possède à leur extrémité 5' un nucléotide modifié par méthylation appelé coiffe (m⁷GpppN) et dans la plupart des cas une queue poly A à leur extrémité 3' (Fig : 1.4). La coiffe 5' est reconnue par le complexe protéique eIF4F composé des facteurs canoniques de la traduction eIF4E, eIF4G et de l'hélicase eIF4A. La queue poly A est elle reconnue par la protéine **PABP** (pour PolyA Binding Protein). PABP interagit avec la protéine eIF4G pour former une structure circulaire reliant les extrémités 5' et 3' de l'ARNm. Cette circularisation favorise l'initiation de la traduction, mais aussi le recyclage des ribosomes (Fig : 1.4).

En parallèle, il y a formation d'un complexe de pré-initiation 43S appelé **PIC** (Pre Initiation Complex). PIC est composé de plusieurs facteurs d'initiation de la traduction (eIF1, eIF1A, eIF3, eIF5), d'un complexe composé de eIF2-GTP, de l'ARNt initiateur de méthionine ainsi que de la petite sous-unité du ribosome 40S. PIC est recruté par le complexe eIF4F déjà fixé sur la coiffe 5' de l'ARNm par interaction directe entre les facteurs canoniques de la traduction. Une fois positionnée sur l'ARNm, la petite sous-unité du ribosome scanne l'ARNm à la recherche du codon d'initiation. Suite à la reconnaissance du codon d'initiation, l'hydrolyse du GTP associé à eIF2 entraîne la libération des facteurs eIF1, eIF3, eIF5, eIF2-GDP et d'un groupement phosphate. Cela permet le recrutement de la grande sous-unité 60S du ribosome, associée à eIF5B au niveau du codon initiateur, et donc la formation du ribosome 80S. Le ribosome est alors prêt à commencer l'élongation de la traduction jusqu'à la rencontre d'un codon STOP. Plusieurs ribosomes peuvent se fixer sur le même ARNm mature. On parle alors de polysomes.

A.4.1.2 Initiation de la traduction par le mécanisme «coiffe indépendante»

Outre l'initiation de la traduction et une régulation par le mécanisme dépendant de la coiffe, entre 5 et 10 % des ARNm d'une cellule eucaryote (HERSHEY, N. SONENBERG et MATHEWS, 2012) utilisent une méthode indépendante de la coiffe pour recruter directement le ribosome. Dans la plupart des cas, ce recrutement correspond à des ARNm qui possèdent, en plus d'une

coiffe, des structures secondaires particulières dans leurs séquences 5'UTR appelés **IRES** (pour Internal Ribosome Entry Site). Ces IRES ont été découvertes pour la première fois chez les virus de picornavirus et poliovirus à la fin des années 80 (JANG et al., 1988; PELLETIER et Nahum SONENBERG, 1988).

En réponse à différents stress tels qu'une infection virale, l'hypoxie, la privation en acide aminé ou même la différenciation cellulaire, la traduction des ARNm « coiffe dépendante » est réduite. Or, afin de répondre à ces changements et de s'adapter, la cellule doit synthétiser des protéines capables de permettre leur survie. Ainsi, la présence des structures IRES permet à des ARNm, dont la plupart jouent un rôle crucial dans le cycle cellulaire ou la survie, d'initier leur traduction et d'assurer la synthèse de protéines particulières malgré un arrêt global de la synthèse protéique. Parmi ces ARNm, on peut mentionner des facteurs de transcription comme c-Myc (STONELEY et al., 1998), p53 (RAY, GROVER et DAS, 2006), le facteur d'angiogenèse VEGF (HUEZ et al., 1998), le facteur de croissance IGF1R (GIRAUD et al., 2001), ou encore les facteurs de prolifération FGF-1 et FGF-2 (CRÉANCIER et al., 2000; MARTINEAU et al., 2004).

Selon la base de données IRESbase (ZHAO et al., 2020), il existerait 691 ARNm possédant une structure IRES chez l'Homme. Le mécanisme par lequel les IRES cellulaires initient la traduction reste relativement peu connu. Cela est dû au fait qu'il n'existe pas de séquence IRES consensus et que différentes structures secondaires ont été identifiées dans les IRES. Chez les virus, il a été clairement démontré que la structure IRES permettait le recrutement de la SSU du ribosome par trois mécanismes : interaction directe des IRES des ARNm et de la SSU, recrutement de la SSU via des facteurs canoniques de la traduction en fixant les IRES, ou via des protéines de liaison à l'ARN appelées ITAFs (pour IRES Trans-Acting Factors) (KWAN et THOMPSON, 2018). Pour les ARNm cellulaires, l'utilisation d'ITAFs reste le mécanisme le plus étudié.

A.4.1.3 Régulation de la traduction

La régulation de la traduction joue un rôle très important dans la plupart des processus biologiques de la cellule. Étant donné que la grande majorité des protéines eucaryotes ont des demi-vies assez longues (>2 h) (RAGHAVAN, 2002), il est important d'effectuer une régulation rapide du niveau protéique

en contrôlant leurs taux de traduction et de dégradation des protéines. Dans certains cas, comme les premiers stades de l'infection virale, ou dans les cellules dépourvues de transcription comme les ovocytes et les réticulocytes (HERSHEY, N. SONENBERG et MATHEWS, 2012; WALSH, MATHEWS et MOHR, 2012), la régulation de la traduction est le seul mécanisme permettant de réguler la synthèse des protéines. Cette régulation est aussi clé pour s'adapter aux stress et aux changements environnementaux. De plus, plusieurs études ont montré que le niveau d'expression de certains ARNm ne correspond pas à la quantité protéique retrouvée dans la cellule (SCHWANHÄUSSER et al., 2011). Sachant que la plupart des études utilisent le transcriptome pour identifier les fonctions biologiques d'une cellule résultantes des protéines, cette discordance entre transcriptome et protéome soutient l'importance de comprendre la régulation traductionnelle pour comprendre certains contextes physio-pathologiques.

Le contrôle de la traduction intervient essentiellement à l'étape d'initiation qui est vue comme l'étape limitante. Il existe deux grands types de régulation traductionnelle. Le premier mode de régulation implique une traduction dépendante de la coiffe. Dans ce cas, les principaux acteurs impliqués dans cette régulation sont les protéines qui interagissent avec les facteurs d'initiation de la traduction et impactent la synthèse protéique globale (Fig : 1.4). De nombreuses voies de signalisation régulent la synthèse protéique globale, la plus connue étant la voie de signalisation **mTOR** (pour mammalian Target of Rapamycin). En réponse à des facteurs de croissance ou d'hormones, mTOR phosphoryle différentes protéines dont 4E-BP (pour eIF4E-binding protein). Sous sa forme non-phosphorylée, 4E-BP inhibe la traduction en se liant et séquestrant le facteur d'initiation de la traduction eIF4E. La phosphorylation de 4E-BP catalysée par mTOR induit sa dissociation de la protéine eIF4E et favorise ainsi le recrutement du complexe d'initiation de la traduction contenant eIF4F sur la coiffe (QIN, B. JIANG et Y. ZHANG, 2016) (Fig : 1.4). D'autres voies de signalisation sont impliquées à différents niveaux dans la régulation de la traduction comme **MAPK** (pour Mitogen-Activated Protein Kinase).

En plus de la régulation de la synthèse protéique globale qui affecte tous les ARNm, il existe un second mode de régulation qui correspond à des régulations de la traduction spécifique à certains ARNm. Ces régulations reposent sur une interaction entre des éléments cis-régulateurs portés par l'ARNm et des éléments trans-régulateurs reconnaissant les éléments cis (Fig : 1.4). Les

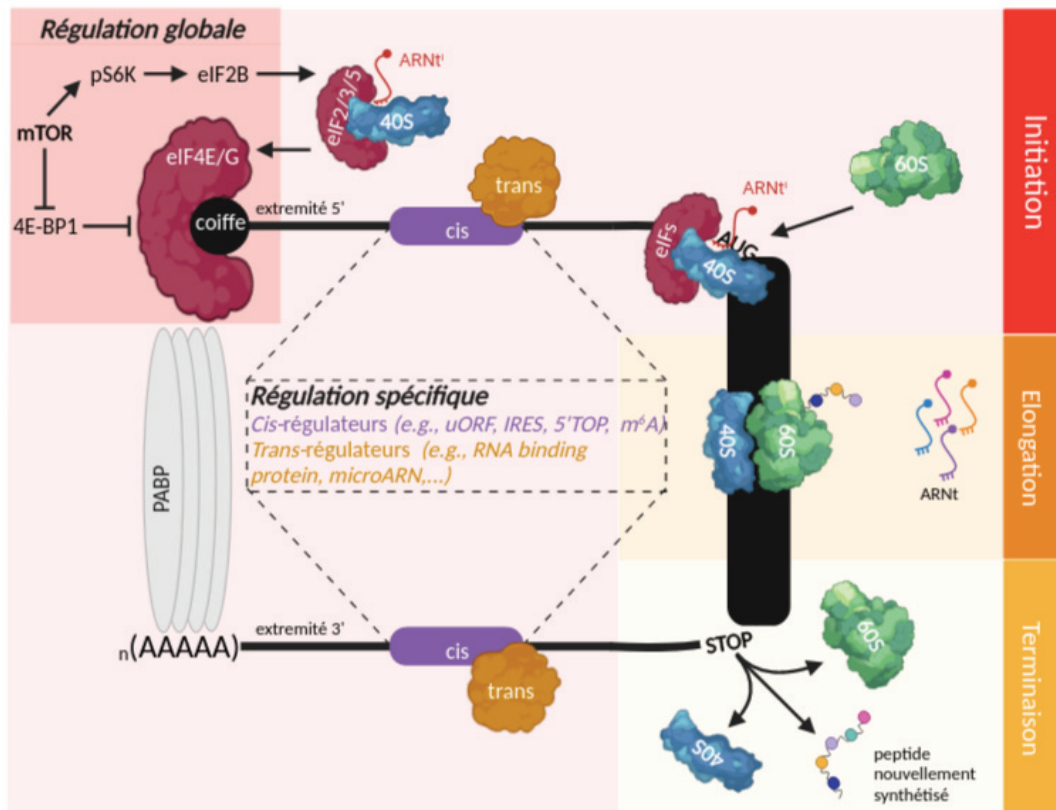


FIGURE 1.4 – Régulation de la traduction. La régulation de la traduction est un processus complexe qui permet de réguler la synthèse des protéines. Le premier mode de régulation est coiffe-dépendante et régule la traduction globale. Cette régulation dépend des facteurs d'initiation de la traduction (eIF) et des voies de signalisation comme mTOR. Le deuxième mode de régulation impacte la traduction d'ARNm spécifique comme les IRES et est coiffe-indépendante. Cette régulation dépend des cis-régulateurs (structuration d'ARNm, modification chimiques ...) et des trans-régulateurs qui peuvent être des petits ARN non codant ou des protéines de liaison à l'ARN. (Adapté de (MORIN et al., 2022))

trans-régulateurs interagissent généralement directement avec l'ARNm par reconnaissance d'une séquence nucléotidique d'ARNm spécifique (en 5'UTR ou 3'UTR) ou d'une structure particulière. Nous avons déjà vu les structures IRES qui permettent d'initier la traduction indépendamment de la coiffe (cf A.4.1.2). D'autres éléments cis-régulateurs ont été identifiés, par exemple les uORF (pour upstream Open Reading Frame), les séquences 5'TOP (pour 5' Terminal OligoPyrimidine Tract) qui sont retrouvés principalement sur les ARNm codant pour les protéines ribosomiques. Quant aux trans-régulateurs, ils peuvent être de nature différente comme des ARN non-codants (microARN,

long-non coding ARN. . .) ou des protéines de liaison aux ARN (RNA-binding protein). Plus récemment, il a été démontré que des modifications chimiques présentes sur l'ARNm, comme les méthylations de base, contribuent à la régulation traductionnelle. Par exemple, la N6-méthyladénosine (m6A), une des modifications les plus abondantes chez les ARNm eucaryotes, de l'ARNm SNAI1 augmente sa traduction en favorisant la liaison avec le facteur d'élongation eEF2 (LIN et al., 2019; MACHNICKA et al., 2012). Ce mécanisme de régulation de la traduction par les modifications chimiques des ARNm est nommé épitranscriptomique. Ces mécanismes restent largement méconnus (KWAN et THOMPSON, 2018).

L'ensemble de ces différents éléments contribue à la régulation de la traduction d'ARNm individuel ou d'un groupe d'ARNm permettant de spécifier un état physiologique ou une réponse cellulaire particulière. Comme nous venons de le voir, la régulation de la traduction implique de nombreux facteurs différents. En revanche, le ribosome n'a jamais été cité. En effet, le ribosome a toujours été considéré, depuis sa découverte, comme l'effecteur de la traduction, un acteur statique et invariable qui n'exerce aucun rôle sur la régulation de la traduction. Cette vision repose sur le fait que les deux composants du ribosome, les ARNr et les protéines, ainsi que leur structure, sont particulièrement conservés au cours de l'évolution. En effet, l'ARNr de la petite sous-unité, 16S pour les procaryotes et 18S pour les eucaryotes, est utilisé dans la construction d'arbres phylogénétiques ou les analyses métagénomiques (FIELD et al., 1988). De plus, le rôle très conservé du ribosome dans la synthèse protéique justifie ce point de vue statique dans le processus complexe de traduction.

Cependant, des données récentes indiquent que le ribosome pourrait être un acteur direct de la régulation traductionnelle. En effet, l'altération de la stœchiométrie ou de la modification chimique d'au moins un de ses 4 ARNr et 80 protéines pourrait induire à une hétérogénéité de composition du ribosome, chaque variant pouvant présenter des activités traductionnelles distinctes. Ces observations signifieraient que la population de ribosomes au sein d'une cellule ou d'un tissu ne serait pas identique, mais que différents types de ribosomes co-existeraient. Dans le prochain sous-chapitre je ferai un état de l'art sur l'hétérogénéité de composition du ribosome et comment ces données mettent en avant le ribosome comme un régulateur direct de la traduction.

B Hétérogénéité de composition du ribosome

Comme décrit précédemment, il existe des différences entre les ribosomes de deux espèces relativement éloignées dans l'arbre phylogénétique, concernant leur taille et leur densité (70S pour les procaryotes et 80S pour les eucaryotes), leurs RP et leurs ARNr. Au sein de la même espèce, la population de ribosomes a longtemps été vue comme homogène, sans changement de structure, de composition ou de fonction biologique. Cependant, peu après la découverte du ribosome, Francis Crick (CRICK, 1958) proposa l'hypothèse : « un gène, un ribosome, une protéine ». En effet, selon Crick, chaque ribosome possède l'information génétique pour produire une protéine unique. La découverte de l'ARNm par Gros, Jacob et Monod (BRENNER, JACOB et MESELSON, 1961 ; GROS et al., 1961 ; JACOB et MONOD, 1961), a montré qu'un ribosome peut traduire plusieurs ARNm en protéines. Brenner *et al* ont montré que l'infection de la bactérie *Escherichia coli* (*E.coli*) par un bactériophage ne résulte pas en la production de « nouveaux » ribosomes (BRENNER, JACOB et MESELSON, 1961). Cette observation implique que les ribosomes déjà présents dans la bactérie avant l'infection peuvent traduire les ARN du bactériophage. Cela a donc conduit à l'hypothèse suivante : les ribosomes seraient des structures non-spécialisées, homogènes et jouant un rôle passif dans la traduction en catalysant tout simplement les protéines codées par les ARNm.

Malgré la dominance quasi totale de cette hypothèse au sein de la communauté scientifique, plusieurs équipes, depuis les années 1980, ont rapporté l'existence d'une diversité de composition du ribosome pas seulement au sein d'un organisme, mais aussi entre deux types cellulaires et même au sein d'une seule cellule. Afin de décrire cette diversité, dans la littérature on peut trouver plusieurs termes. L'hypothèse du ribosome « filtrant » (MAURO et EDELMAN, 2002) suggère que la traduction peut être régulée par le ribosome via ses interactions plus ou moins spécifiques avec certains ARNm. Komili *et al* proposent « le code » du ribosome qui fonctionnerait de manière similaire au code histone (KOMILI et al., 2007). En effet, l'état de transcription d'une région de la chromatine dépend des combinaisons des protéines histones, de leurs modifications post-traductionnelles (PTM) et des modifications d'ADN (JENUWEIN et ALLIS, 2001). De façon similaire, les différentes combinaisons des protéines ribosomiques, leurs paralogues et leurs

modifications post-traductionnelles, les variations de séquences des ARNr et les modifications chimiques des ARNr permettraient de traduire des ARNm spécifiques en formant ainsi le « code » du ribosome. Plus récemment, Sulima *et al* utilisent une étonnante métaphore afin de décrire tous les ribosomes différents qui peuvent exister, le « riboverse », un amalgame lexical formé par la contraction des mots ribosome et univers (S. SULIMA, K. KAMPEN et KEERSMAECKER, 2019). Finalement, ces différents termes qui décrivent la diversité de composition du ribosome sont regroupés sous le concept du ribosome « spécialisé » selon lequel les ribosomes ne possèdent pas une structure constante, mais que leur composition/stœchiométrie varie en fonction des contextes physio-pathologiques et induit une conséquence fonctionnelle (DINMAN, 2016; GENUTH et BARNA, 2018; GILBERT, 2011; Z. SHI et BARNA, 2015; XUE et BARNA, 2012). Ainsi, certains « types » de ribosome vont traduire des ARNm très spécifiques. Tout au long de ma thèse, je vais utiliser le terme « hétérogénéité du ribosome » pour faire référence à l'hypothèse que les ribosomes n'ont pas une structure ou composition stable, mais variable au sein d'une cellule ou type cellulaire.

Les études autour de l'existence d'une hétérogénéité du ribosome ont été réalisées dans différents organismes comme la bactérie, la levure, la drosophile, le poisson zèbre, la souris, les lignées cellulaires humaines et les plantes. L'hétérogénéité du ribosome a été observée pour les deux éléments qui le constitue : les RP et les ARNr (Fig : 1.5).

B.1 Hétérogénéités des RP

Les premières observations concernant une hétérogénéité de composition du ribosome ont été faites sur les RP. Différentes études ont permis d'observer que les RP sont une source d'hétérogénéité, et ce, à différents niveaux (Fig : 1.5) :

- (i) Différence dans la stœchiométrie des RP entre les ribosomes. Une RP peut être présente ou absente dans la structure du ribosome.
- (ii) Existence de RP paralogues. Certaines RP sont codées par plusieurs gènes (paralogues) et peuvent coder pour des protéines ayant des différences dans leurs séquences en acides aminés ou même des séquences identiques.

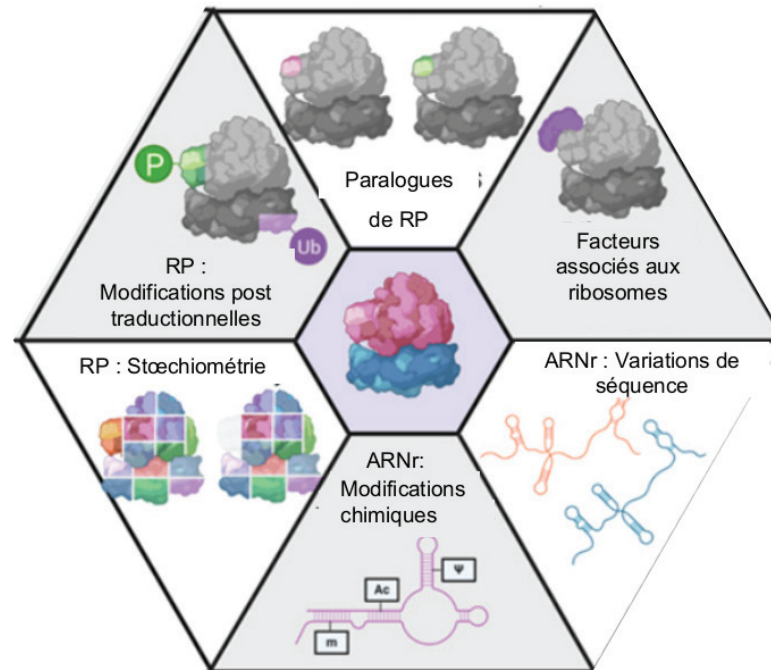


FIGURE 1.5 – Schéma représentant les différentes sources d’hétérogénéité de composition du ribosome. (Adapté de (NORRIS, HOPES et ASPDEN, 2021))

- (iii) Différences dues aux modifications post-traductionnelles (PTM) des RP. Les RP peuvent avoir des modifications chimiques ajoutées post-traductionnellement, comme des acétylations ou des phosphorylations des acides aminés.
- (iv) Mutations des RP. Les gènes codant les RP peuvent présenter des mutations qui induisent généralement la substitution d’un acide aminé par un autre dans la protéine synthétisée. Les mutations étant essentiellement associées au cancer, cette source d’hétérogénéité de composition du ribosome sera traité dans le chapitre [C.1.1](#)

B.1.1 Stœchiométrie des RP

L’idée que la stœchiométrie des RP est toujours de rapport 1 : 1 parmi la population de ribosome a été remise en question dès les années 1970. Les premières études faites sur les bactéries ont montré que la quantité de certaines protéines comme RPS6, RPS21 ou RPL12 varie en fonction des conditions de

culture (DEUSSER et WITTMANN, 1972; WEBER, 1972). Les développements technologiques, comme la spectrométrie de masse (MS) ou le séquençage haut-débit de l'ARN (RNA-seq) ainsi que l'amélioration de la précision dans les mesures (l'amélioration de la résolution pour les machines de MS et la profondeur de séquençage pour RNA-seq), ont aidé à mieux mettre en évidence l'hétérogénéité de composition du ribosome en RP.

Des études transcriptomiques ont révélé que les gènes codant les RP sont différenciellement exprimés entre les différents tissus et types de lignées cellulaires. Kondrashov *et al* ont étudié l'expression génique de 72 RP dans 14 régions et types cellulaires différents de l'embryon de souris (KONDRASHOV et al., 2011). L'expression des RP était très hétérogène entre les types cellulaires et régions, spécialement pour certaines RP comme les protéines de la grande sous-unité RPL10 ou RPL38. D'autres études ont montré que cette hétérogénéité est retrouvée entre différents tissus adultes humains. En 2001, Bortoluzzi *et al* ont analysé le niveau d'expression des ARNm des RP dans six tissus humains adultes. Ils ont identifié 13 gènes significativement différenciellement exprimés entre les tissus (BORTOLUZZI et al., 2001). Cinq RP étaient plus exprimées dans les ovaires (RPL14, RPL38, RPS9, RPS10 et RPS19) alors que huit RP étaient plus exprimés dans les muscles squelettiques (RPL37, RPL41, RPS17 et RPS25 étant les plus significatifs). Ces études, bien qu'analysant l'expression transcriptomique des gènes codant les RP et non pas la quantité protéique des RP et encore moins leur stœchiométrie au sein des ribosomes, suggèrent néanmoins l'existence de ribosomes de compositions différentes en fonction des conditions biologiques.

Des études plus récentes faites sur des cellules souches embryonnaires de souris ont montré pour la première fois des différences de stœchiométrie d'une partie des RP au niveau protéique dans les ribosomes (Z. SHI, FUJII et al., 2017; SLAVOV et al., 2015). En effet, dans ces deux études la quantité des RP a été mesurée dans différentes fractions de polysomes (c.-à-d., ARNm en cours de traduction et donc liés à des ribosomes) en utilisant la technologie de spectrométrie de masse. Les ribosomes en cours de traduction ont des compositions en RP différentes de celle des ribosomes libres dans le cytoplasme. Des résultats similaires ont été obtenus chez la levure, ou en plus de la variation de la stœchiométrie des RP dans les polysomes, une variation dans différentes conditions de culture a été observée pour RPL35b ou RPL26a (SLAVOV

et al., 2015). Shi *et al* ont réalisé une étude similaire en analysant la stœchiométrie de 15 RP dans des fractions de polysome et les ribosomes libre issus de cellules souches embryonnaires de souris (Z. SHI, FUJII et al., 2017). Parmi les 15 RP, quatre RP (RPL10, RPL38, RPL40 et RPS26) avaient une abondance significativement plus élevée et deux RP (RPL10A, RPS7) une abondance plus faible dans les fractions de polysomes comparé aux ribosomes libre dans le cytoplasme. Encore plus intéressant, les ribosomes enrichis en RPL10A traduisent de manière spécifique un sous-groupe de gènes ayant un IRES et codant des protéines de la matrice extracellulaire. Par ailleurs, la quantité de ribosomes enrichis en RPL25 est corrélée avec la traduction d'un sous-groupe d'ARNm impliqués dans la signalisation de la vitamine B12.

De plus, l'analyse de RP spécifiques soutient le concept d'hétérogénéité de composition du ribosome. Chez la levure, une exposition à un stress comme un taux élevé de sel ou de pH est associée à la génération de ribosomes déficients en RPS26. Cela entraîne une augmentation de la traduction de certains ARNm codant des protéines impliquées dans la réponse au stress (FERRETTI et al., 2017). Une autre RP qui participe à l'hétérogénéité de composition du ribosome est RPL38. RPL38 est non seulement exprimée de manière tissu-spécifique (KONDRASHOV et al., 2011), mais régule aussi la traduction d'un sous-groupe d'ARNm codant des protéines ayant une boîte homéotique (homeobox genes) qui sont essentiels au développement embryonnaire (XUE, TIAN et al., 2014). Ces résultats montrent dans un premier temps que des ribosomes ne possédant pas une ou plusieurs RP participent à la synthèse protéique. En effet, plusieurs études se sont intéressées à la composition en RP des ribosomes contenus dans les fractions de polysomes, donc actifs dans la traduction. Dans un second temps, on observe que la traduction tissu-spécifique d'un sous-groupe d'ARNm suite à une réponse au stress ou au cours du développement embryonnaire est associée à une hétérogénéité de composition du ribosome au niveau des RP. Certains ribosomes enrichis en certaines RP spécifiques traduisent davantage un sous-groupe d'ARNm, soutenant l'hypothèse que l'hétérogénéité du ribosome conduit à des ribosomes spécialisés.

Cependant, il est important de rappeler que certaines études se sont basées sur l'analyse du niveau d'expression des ARNm et des protéines RP au niveau de la cellule ou de tissus et non au niveau de fractions de ribosomes, ce qui ne reflète pas forcément la composition réelle du ribosome. Par ailleurs,

certaines RP peuvent être libres et de ce fait non incorporées dans le ribosome, et dans certains cas jouer un rôle biologique extra-ribosomique (Jonathan R. WARNER et MCINTOSH, 2009). Il est important de continuer à étudier le contenu des ribosomes en RP dans différentes conditions biologiques afin de mieux comprendre le rôle de la variation de la stœchiométrie des protéines ribosomiques dans la régulation de la traduction et des mécanismes associés.

B.1.2 Paralogue de RP

Le nombre de gènes paralogues codant des RP varie entre les différents organismes. *Arabidopsis thaliana*, la plante modèle par excellence en biologie, possède de 2 à 7 gènes paralogues pour chaque RP (BARAKAT et al., 2001). La levure *Saccharomyces cerevisiae* possède 2 copies de gènes pour 59 des 79 RP (KELLIS, BIRREN et LANDER, 2004). Certains gènes paralogues présentent des séquences identiques et d'autres présentent des variations dans leurs séquences. Ainsi, l'incorporation de ces RP paralogues dans le ribosome apporte une nouvelle source d'hétérogénéité de composition du ribosome. À l'aide de données transcriptomiques et protéomiques, on a pu observer, dans plusieurs études, une expression tissu-spécifique des RP paralogues. Cela suggère que des types cellulaires différents utilisent des RP paralogues différentes dans la composition de leur ribosome.

Chez la drosophile *D. melanogaster*, la RP RPL22 possède une RP paralogue appelée RPL22-like (RPL22L) avec une conservation d'identité de seulement 38 % (KEARSE, CHEN et V. C. WARE, 2010; MARYGOLD et al., 2007). RPL22L est exprimée spécifiquement dans les testicules, alors que RPL22 est exprimée de manière équivalente dans tous les tissus. Plus récemment, Mageeney *et al* ont démontré que les ribosomes ayant incorporé RPL22L sont nécessaires pour la traduction d'ARNm nécessaires au processus de spermatogenèse, supportant la notion d'une spécialisation de fonction du ribosome sur la base d'une composition différente (MAGEENEY et Vassie C. WARE, 2019). D'autres protéines paralogues de la drosophile sont exprimées de manière spécifique dans certains tissus comme RPS5B, RPS19A, RPL10AA et RPL37B (MARYGOLD et al., 2007).

La plante *A. thaliana* présente plusieurs exemples d'expression tissus-spécifiques ou au cours du développement de RP paralogues. RPS5A est exprimée au

cours de la différenciation cellulaire alors que sa protéine paralogue RPS5B est exprimée au cours de la division cellulaire (WEIJERS et al., 2001). Les deux RP paralogues de RPL23 (RPL23AA et RPL23AB) ont une homologie de séquence très élevée (94,8 % d'identité de la séquence protéique). En revanche, seule la déplétion de RPL23AA, et pas celle de RPL23AB, entraîne une altération de la croissance et du développement de la plante, démontrant qu'un petit changement de séquence protéique peut induire des conséquences fonctionnelles importantes (DEGENHARDT et BONHAM-SMITH, 2008).

Chez les mammifères, notamment chez l'Homme, la plupart des RP possèdent une seule copie de gène, à l'exception d'un petit nombre identifié dans la littérature. C'est le cas de RPS4 qui est codé par 3 gènes : RPS4X situé dans le chromosome X, RPS4Y1 et RPSY2 situés dans les chromosomes Y. Seul RPSY2 est exprimé de manière tissu-spécifique (testicules et prostate) (LOPES et al., 2010). Chez la souris, les gènes paralogues de RPL10, RPL22 et RPL39 sont aussi exprimés de manière tissu-spécifique (SUGIHARA et al., 2010).

Ces données suggèrent que les ribosomes présents dans différents tissus ou conditions biologiques contiennent des RP paralogues différentes afin de traduire de manière spécifique un sous-groupe d'ARNm.

B.1.3 Les modifications post traductionnelles des RP

Les modifications post-traductionnelles (PTM) jouent un rôle supplémentaire dans la génération de ribosomes de compositions différentes. Parmi les différentes PTM on peut citer les acétylations, méthylations, glycosylations, phosphorylations et ubiquitinations.

Une des premières PTM des RP découverte, et parmi les mieux caractérisées, est la phosphorylation de RPS6 qui a été identifiée en 1974 par Gressner et Wool en analysant des ribosomes de foie de souris après incorporation de phosphate radioactif (GRESSNER et WOOL, 1974). RPS6 est phosphorylée sur 5 sites et concerne l'acide aminée Sérine (MEYUHAS, 2015). De nombreuses études ont démontré que RPS6 est phosphorylée en réponse à l'induction ou l'activation de facteurs de croissance (voie de signalisation PI3K/mTORC1) (MEYUHAS, 2015), et également de l'activation physiologique ou pharmacologique des neurones chez la souris (KNIGHT et al., 2012). De ce fait, la phosphorylation de RPS6 est utilisée comme marqueur de l'activation de la

voie mTORC1 ou des neurones activés. D'ailleurs des anticorps reconnaissant spécifiquement la phosphorylation de RPS6 ont permis d'isoler les ribosomes et les ARNm associés qui sont traduits au cours de l'activation neuronale (KNIGHT et al., 2012). Malgré l'utilisation de cette PTM comme un marqueur, le rôle biologique et son mécanisme ne sont pas encore entièrement connus.

Chez la levure, des PTM sont observées sur plusieurs RP. Un exemple est l'hydroxylation (l'ajout de groupement OH sur un acide aminé, la proline dans ce cas) de la protéine RPS23. Loenarz *et al* ont démontré que la présence ou l'absence de l'hydroxylation sur RPS23 affecte la capacité du ribosome à lire les codons STOP. Chez un autre organisme modèle, *Dictyostelium discoideum*, des différences de méthylation et de phosphorylation ont été observées entre les ribosomes isolés pendant le stade unicellulaire et pluricellulaire (RAMAGOPAL et ENNIS, 1981).

Contrairement aux autres types de sources d'hétérogénéité de composition de ribosome qui surviennent essentiellement au cours de la biogenèse du ribosome, les PTM peuvent être introduites directement dans le ribosome mature. Cela pourrait permettre à la cellule de moduler rapidement la composition du ribosome et donc son activité de traduction afin de s'adapter et de traduire des ARNm spécifiques.

C Hétérogénéité de composition du ribosome liée aux ARNr

Les ARNr sont essentiels pour l'assemblage, la structure et les activités catalytiques du ribosome. Par conséquent, toute modification des ARNr, que ce soit (i) sur la séquence nucléotidique ou (ii) des modifications post-transcriptionnelles, pourrait avoir une conséquence directe sur la structure et la fonction du ribosome (Fig : 1.5).

C.0.1 Hétérogénéité de séquences des ARNr

Comme expliqué précédemment, dans les cellules eucaryotes ou procaryotes, les ARNr sont codés par plusieurs locus d'ADN ribosomiques qui sont répétés en tandem. Le nombre de répétitions varie entre les espèces avec un plus grand nombre de répétitions chez les eucaryotes comparés aux procaryotes.

Ainsi, ce nombre important de gènes peut induire des expressions d'ARNr différentes en fonction de certaines conditions.

Haloarcula marismortui, un organisme unicellulaire procaryote, possède trois opérons qui codent pour les ARNr 16S, 23S et 5S : opéron A, B et C (BALIGA et al., 2004). Les opérons A et C sont fortement identiques, tandis que l'opéron B présente un total de 135 variations nucléotidiques à travers les trois ARNr et est plus enrichi en GC. Ce dernier est exprimé jusqu'à 4 fois plus en réponse à de fortes températures comparé aux deux autres opérons, et jusqu'à trois fois moins à des températures faibles. Chez un autre organisme parasitaire qui cause la maladie de la malaria, *Plasmodium berghei*, il existe deux gènes qui codent pour l'ARNr de la petite sous-unité du ribosome. Un gène est exprimé principalement lorsque le parasite vit dans le moustique et l'autre gène lorsque le parasite infecte un hôte mammifère (SPAENDONK et al., 2001). Chez l'Homme et la souris, une étude datant de 2018 a identifié des variations nucléotidiques entre différents individus et même, ce qui paraît encore plus surprenant, certaines des variations nucléotidiques étaient tissu-spécifique chez l'Homme et la souris (PARKS et al., 2018). Au vu de ces observations, il est très tentant d'en conclure que l'expression de différents variants d'ARNr génère des ribosomes hétérogènes qui contribuent à s'adapter à des besoins traductionnels spécifiques.

Cette idée est supportée par des données récentes. Locati *et al* se sont intéressés à l'expression des ARNr au cours du développement de l'organisme modèle de poisson zèbre (*Danio rerio*) (LOCATI, PAGANO, ENSINK et al., 2016; LOCATI, PAGANO, GIRARD et al., 2017). Ils ont observé que les ARNr exprimés dans les ovocytes sont de types maternels et qu'ils sont remplacés au cours de l'embryogenèse par des ARNr de type somatique. Ces résultats laissent supposer, en se basant uniquement sur les variations nucléotidiques des ARNr, que deux types de ribosomes pourraient exister chez le poisson zèbre pendant son développement. Les analyses *in silico* ont montré que des ARNm maternels pourraient interagir avec certaines régions de l'ARNr 18S maternel, mais pas avec l'ARNr 18S somatique. De plus, chez l'Homme, toujours à l'aide d'analyses *in silico*, on a pu observer que des ARNm interagissent avec les segments d'expansions des ARNr 18S et 28S (PARKER et al., 2018). Par exemple, le segment d'expansion de la petite sous-unité du ribosome ES9S interagit avec la structure IRES de l'ARNm du gène *Hoxa9* et favorise sa traduction.

On notera cependant que le grand nombre de répétitions des ARNr dans le génome et le taux de GC qui peut atteindre jusqu'à 80 % sur certaines régions de l'ARNr 28S complexifie l'analyse des variations nucléotidiques. Compte tenu de ces premières observations, il paraît important d'effectuer de nouvelles études plus approfondies afin de mieux comprendre le rôle des variations nucléotidiques des ARNr et si elles participent à la génération de ribosomes spécialisés. Aujourd'hui il paraît donc nécessaire de réaliser des études fonctionnelles *in vitro* ou *in vivo* afin de mieux identifier les variations nucléotidiques des ARNr et comprendre leur rôle biologique.

C.1 Les modifications post-transcriptionnelles des ARNr

Les ARNr, comme les RP, sont sujets à des modifications chimiques. En général l'addition de ces modifications chimiques survient au cours de la transcription et de la maturation des ARNr (TUROWSKI et TOLLERVEY, 2014). On parle ainsi de modifications post-transcriptionnelles. On peut mentionner trois différents types de modifications chimiques des ARNr : la 2'-O-ribose méthylation (2'Ome), la pseudouridylation (Ψ) et la modification de base comme la méthylation et l'acétylation de bases azotée d'un nucléotide. En général, les modifications chimiques de l'ARNr servent à stabiliser la structure secondaire et tertiaire de l'ARNr ce qui va permettre de produire des ribosomes matures et fonctionnels. La plupart de ces modifications sont introduites sur des nucléotides bien spécifiques situés à proximité des centres catalytiques du ribosome, à sa surface ou à l'interface des deux sous-unités du ribosome. Chez les cellules eucaryotes, environ 2 % des nucléotides des ARNr sont modifiés chimiquement (chez l'Homme par exemple, il y a environ 200 nucléotides modifiés sur un total d'environ 7200 nucléotides). Les modifications les plus répandues sont les 2'Ome et les Ψ . Étant donné l'importance des modifications chimiques des ARNr dans la conservation de la structure du ribosome et en conséquence dans sa fonction, il a longtemps été pensé que les ARNr étaient toujours modifiées. Néanmoins, des études récentes indiquent que certains nucléotides ne sont pas toujours modifiés et cela apporte une nouvelle source d'hétérogénéité de composition du ribosome.

C.1.1 Les modifications de bases

Les modifications de bases des ARNr sont principalement de trois types : méthylation (m), acétylation (ac) et aminocarboxypropylation (acp). Elles sont moins abondantes que les 2'Ome ou les Ψ et sont catalysées par des enzymes spécifiques. Chez l'Homme il existe 12 modifications de bases identifiées à ce jour (TAOKA, NOBE, YAMAKI, SATO et al., 2018). Par exemple, N4-acétylcytidine (ac4C) est une acétylation de la base azotée catalysée par l'acetyltransférase NAT10 chez l'Homme (en position C1337), Kre33 chez la levure et chez certaines archaea. Chez *Thermococcus kodakarensis*, une espèce d'archaeas thermophile, il a été démontré que le taux de ac4C varie en fonction de la température. De plus, l'analyse de la structure par cryo-EM suggère que cette modification joue un rôle dans la stabilisation du ribosome à des températures élevées, en facilitant le repliement des ARNr (SAS-CHEN et al., 2020).

NSUN5 est une méthyltransférase qui catalyse la méthylation du carbone en position 5 de la cytosine en position 3782 de l'ARNr 28S chez l'Homme (28S_m5C3782). La perte de cette modification entraîne une diminution globale de la synthèse protéique mais augmente la traduction d'ARNm spécifiques impliqués dans l'adaptation au stress cellulaire (JANIN et al., 2019). Une autre étude montre des résultats similaires chez la souris où la délétion de NSUN5 réduit la traduction globale dans plusieurs tissus (HEISSENBERGER et al., 2019).

Un autre exemple très récent, correspond à une modification complexe et très conservée au cours de l'évolution : la 1-méthyl-3- α -amino- α -carboxylpropyl pseudouridine en position 1248 de l'ARNr 18S (18S_m1acp3Ψ1248) chez l'Homme qui est catalysée par la protéine TSR3. Cette modification chimique se trouve dans le centre de décodage du ribosome et proche du site P. Babaian *et al* ont observé qu'environ 45 % des patients atteints d'un cancer du colon présentent une perte de cette modification chimique. Cette absence de modification chimique est associée à une augmentation de la traduction de certains ARNm codant pour les RP (BABAIAN et al., 2020).

Ces exemples soutiennent l'hypothèse selon laquelle l'hétérogénéité de composition du ribosome ne serait pas induite que par des mécanismes génétiques, mais aussi par des mécanismes épitranscriptomiques : les modifications chimiques des ARNr.

C.1.2 Pseudourilation des ARNr

La pseudouridilation est l'une des modifications chimiques majeures des ARNr et consiste en une réaction chimique d'isomérisation qui convertit les nucléotides d'uridine en pseudouridine. Chez l'Homme, certaines études ont détecté jusqu'à 104 Ψ (TAOKA, NOBE, YAMAKI, SATO et al., 2018). Les Ψ sont catalysées par des complexes ribonucléoprotéiques (RNP) à boîte H/ACA. Chaque complexe est constitué d'un petit ARN non codant (snoARN) à boîte H/ACA et de quatre protéines : GAR1, NHP2, NOP10 et l'enzyme dyskérine (DKC1). Les snoARN H/ACA (SNORA) contiennent une séquence ANANNA (N peut être un des quatre nucléotides) appelé « hinge box » ou boîte H, et une séquence de trois nucléotides (ACA) appelé boîte ACA présente à l'extrémité 3'. Structurellement, les SNORA sont caractérisés par une double épingle à cheveux, qui contiennent une boucle simple brin (poche de pseudouridylation) complémentaire et spécifique à une séquence cible particulière de l'ARNr. Ainsi le SNORA va interagir avec l'ARNr par homologie de séquence et guider la dyskérine DKC1 sur l'uridine d'intérêt pour catalyser son isomérisation (GE et Y.-T. YU, 2013; PENZO et Lorenzo MONTANARO, 2018).

Les premières indications d'une variabilité de la pseudouridilation datent de la fin des années 1990 quand des mutations du gène DKC1 ont été identifiées dans les patients atteints de dyskératose congénitale liée à l'X (X-DC) (HEISS et al., 1998). X-DC est caractérisé par une défaillance des tissus en prolifération (peau, moelle osseuse...) et une susceptibilité accrue au cancer. Il s'accompagne d'une réduction globale de taux de pseudouridilation des ARNr et d'une altération de la traduction d'un groupe d'ARNm contenant des structures IRES (RUGGERO et al., 2003; YOON et al., 2006). Yoon *et al* ont utilisé un modèle de souris où ils ont introduit des mutations sur le gène DKC1 (DKC1^m) pour imiter les caractéristiques cliniques de la dyskératose congénitale chez l'Homme. Il s'est avéré que dans ce modèle murin, la mutation de DKC1 entraîne une diminution de la pseudouridilation des ARNr (RUGGERO et al., 2003; YOON et al., 2006). Alors que la synthèse protéique globale n'est pas impactée dans les lignées cellulaires primaires DKC1^m comparé aux lignées cellulaires normales, une altération de la traduction de certains ARNm a été observée. En effet, l'analyse des ARNm en cours de traduction dans les fractions de polysomes a montré une altération de la traduction

d'un groupe d'ARNm contenant des structures IRES, y compris des inhibiteurs de l'apoptose (BCL-xL et XIAP) et des suppresseurs de tumeur (p27 et p53) (BELLODI et al., 2010; Lorenzo MONTANARO, CALIENNI et al., 2010; YOON et al., 2006). Une autre étude a identifié d'autres ARNm contenant des structures IRES, comme le facteur de croissance VEGF ou la protéine du choc thermique Hsp70, dont le niveau de traduction est fortement augmenté suite à une extinction du gène de DKC1 dans des lignées épithéliales de cancer du sein (ROCCHI et al., 2013). Cependant, jusqu'en 2015 aucun lien direct entre altération de la pseudouridilation des ARNr et traduction n'a pas été établi. Pour répondre à la question, Penzo *et al* ont utilisé un système de traduction *in vitro* pour démontrer que les ribosomes purifiés à partir des cellules humaines dépourvues de DKC1 et présentant des altérations de pseudouridilations des ARNr affectent la traduction dépendante des IRES (PENZO, LUDOVINI et al., 2015). Les résultats ont confirmé que les ribosomes purifiés à partir des lignées cellulaires dépourvues de DKC1 augmentent (VEGF) ou réduisent (p53) la traduction de certains ARNm possédant une IRES. De plus, ces ARNm IRES dépendants sont plus susceptibles d'incorporer un mauvais acide aminé (c.-à-d. défaut de traduction nommé « mis-incorporation » par mauvais décodage de l'ARNm), indiquant que l'altération de la pseudouridilation des ARNr contribue aussi à une augmentation des erreurs de la traduction. Ces données établissent un lien direct entre l'altération globale de la pseudouridilation des ARNr et l'altération de la traduction, tant au niveau de la fidélité de traduction que de la traduction spécifique de certains ARNm. D'autres études chez la levure et des lignées cellulaires humaines ont permis de démontrer qu'une diminution globale de la pseudouridilation est liée à la qualité de la traduction effectuée par le ribosome et à la diminution de la traduction (JACK et al., 2011; PENZO, LUDOVINI et al., 2015). Le mécanisme par lequel l'altération de la pseudouridilation des ARNm affecte la traduction d'ARNm spécifique reste inconnu.

Des études de spectrométrie de masse sur des lignées cellulaires humaines ou chez la levure (TAOKA, NOBE, YAMAKI, SATO et al., 2018; TAOKA, NOBE, YAMAKI, YAMAUCHI et al., 2016; YAMAUCHI et al., 2015) ont montré que seulement certaines positions sont partiellement pseudouridilées. Cela signifie que le matériel biologique analysé correspond à un mélange d'ARNr qui, à un même nucléotide (ou site) porte ou non une Ψ . Ainsi, contrairement à

ce qui était attendu, l'ARNr peut exister sans Ψ . Plus récemment, le développement de technologies de séquençage haut-débit comme le Psi-seq ou l'HydraPsiSeq a permis d'analyser la totalité des sites Ψ en un seul séquençage (CARLILE, ROJAS-DURAN et GILBERT, 2015; MARCHAND, PICHOT et al., 2020). Marchand *et al.*, en utilisant l'HydraPsiSeq, ont identifié de 20 à 25 sites qui présentent une grande variabilité du taux de pseudouridilation, suggérant que certains sites sont plus susceptibles de présenter une absence de pseudouridilation que d'autres.

À ce jour, quelques études ont cherché à identifier la cause de la variation du niveau de pseudouridilation. En plus d'une mesure directe du taux de pseudouridilation dans un échantillon biologique, plusieurs études se sont intéressées au niveau d'expression des SNORA qui guident ces modifications chimiques (FAUCHER-GIGUÈRE et al., 2022; MCMAHON et al., 2019; RONCHETTI et al., 2013). Par exemple, la perte de l'expression de SNORA24 est associée à une diminution de la pseudouridilation de deux sites : 18S_ Ψ U609 et 18S_ Ψ U863. McMahon *et al.* ont observé que la biogenèse des ribosomes ou la synthèse protéique globale n'a pas été affectée. Cependant, l'efficacité de la sélection des bons aminoacyl-ARNt lors de l'étape de décodage de l'ARNm a été affectée dans les ribosomes dépourvus des deux pseudouridilations. Cela a entraîné l'augmentation d'incorporation de mauvais acides aminés ainsi que la mauvaise lecture des codons STOP. 18S_ Ψ U609 est localisé dans une région d'ARNr directement impliquée dans la reconnaissance du bon anticodon ARNt dans le site A du ribosome et 18S_ Ψ U863 est localisé dans un segment d'expansion de l'ARNr 18S assez éloigné par rapport au centre de décodage. Ainsi, une des hypothèses serait que l'absence de 18S_ Ψ U609 et 18S_ Ψ U863 modifient la capacité de la petite sous-unité et du centre de décodage à reconnaître le bon aminoacyl-ARNt dans le site A ou les codons STOP. Il serait intéressant de confirmer que ces résultats sont la conséquence directe de l'absence de la pseudouridilation en utilisant une autre approche comme la traduction *in vitro*, et non à cause d'autres mécanismes moléculaires qui pourrait impliquer SNORA24. En effet, des activités non-canoniques des snoARNs ont été décrits dans la régulation de l'expression génique (épissage, stabilité des ARNm...). Dans une étude de 2022, Faucher-Giguère *et al.* ont montré que l'expression de SNORA81 corrèle avec le taux de pseudouridilation de son site cible de l'ARNr 28S : 28S_ Ψ 4636 et son inhibition conduit à une réduction de la prolifération et de la migration des

cellules.

Ces résultats montrent que les SNORA peuvent avoir un impact direct dans la génération de ribosome de composition différente et des répercussions fonctionnelles. Néanmoins, une étude complète sur les différences d'expression des SNORA et la variation du taux de la pseudouridilation est nécessaire afin de mieux comprendre leurs rôles et leurs régulations dans différents contextes biologiques.

C.1.3 La 2'-O-ribose méthylation des ARNr

La 2'Ome est la modification des ARNr la plus répandue chez l'Homme. Elle consiste en l'ajout d'un groupement méthyl en position 2' du ribose d'un des quatre types de nucléotide. Parmi les 4 ARNr, des nucléotides 2'Ome ont été identifiés sur 3 ARNr : 18S, 5.8S et 28S chez l'Homme. Jusqu'à 112 sites 2'Ome ont été identifiés par certaines études (TAOKA, NOBE, YAMAKI, SATO et al., 2018). Les sites de 2'Ome des ARNr ont été identifiés par plusieurs méthodes de détection comme la spectrométrie de masse, les techniques de séquençage moyen ou haut-débit et par la présence d'un snoARN à boîte C/D. En effet, comme les pseudouridilations, les 2'Ome sont catalysées par des complexes ribonucléiques à boîte C/D. Ces complexes sont composés d'un snoARN à boîtes C/D associé à quatre protéines : SNU13 (15.5K ou NHP2L1), NOP56, NOP58 et la méthyltransférase fibrilarine FBL. Les snoARN C/D contiennent deux motifs de séquences très conservés : la boîte C (RUGAUGA ou R correspond à une purine) située à l'extrémité 5' et la boîte D (CUGA) à l'extrémité 3'. Pour une seule molécule de snoARN, les boîtes C/D existe en paire et sont appelées C'/D'. Du point de vue structurel, les snoARN C/D sont caractérisés par une épingle à cheveux contenant une boucle. La boucle est délimitée par les boîtes C/D et C'/D'. Les boîtes C et D peuvent s'apparier pour former un motif de structure appelé « k-turn ». La séquence à l'intérieur de la boucle est complémentaire à une région d'ARNr. Après l'hybridation avec l'ARNr, la FBL catalyse la réaction de méthylation au nucléotide situé cinq positions en amont de la boîte D.

Les premières évidences de l'altération de la 2'Ome chez l'Homme proviennent de l'étude de lignées cellulaires au sein de mon laboratoire d'accueil (BELIN et al., 2009). De plus, il s'est avéré que cette altération de la 2'Ome des ARNr

est associée à une diminution de la traduction dépendante des IRES de certains ARNm tels que p53, VEGF ou IGF1R, mais sans altération de la synthèse protéique globale. Dans une autre étude, mon équipe d'accueil s'est intéressée à l'impact de la surexpression de la méthyltransférase FBL sur la modulation du niveau de 2'Ome (Virginie MARCEL, GHAYAD et al., 2013). Contrairement à l'inactivation de DKC1 qui a un impact global sur le taux de Ψ , la modulation d'expression de FBL affecte davantage certains sites des ARNr que d'autres. Cependant, cette étude n'a pas montré de lien direct entre les ribosomes qui présentent une variabilité du niveau de 2'Ome et l'altération de la traduction. En effet, comme FBL participe aussi au clivage du pré-ARNr ou à la régulation de l'activité de l'ARN Pol I, il était nécessaire de démontrer le lien direct entre variation de 2'Ome et traduction chez l'Homme. En 2017, Erales et al ont purifié des ribosomes présentant une altération du niveau de la 2'Ome induite par une diminution d'expression de FBL (ERALES et al., 2017). Par la technique de traduction *in vitro* qui permet d'analyser l'activité intrinsèque d'une population de ribosomes purifiés, il a été montré que l'utilisation de ribosomes au profil de 2'Ome différents affecte la traduction des ARNm IRES dépendantes comme IGF1R ou VEGF. Ces données soutenaient l'idée que la 2'Ome des ARNr peut contribuer à l'activité de régulation traductionnelle du ribosome.

À mon début de thèse de 2019, les seules études chez l'Homme sur la 2'Ome correspondaient à l'analyse de la variation de la 2'Ome dans des modèles de lignées cellulaires. Mon projet de thèse avait donc pour objectif de démontrer qu'il existe dans la vraie vie, une hétérogénéité de composition des ARNr induite par une variation de 2'Ome par une approche haut-débit. En effet, historiquement, la détection et la quantification du 2Ome reposait sur des approches biochimiques et moléculaires, y compris, par exemple, la spectrométrie de masse, l'extension des amorces (c'est-à-dire la détection des pauses de transcription inverse à l'aide d'amorces radiomarquées) ou la RTL-PCR (Reverse Transcription en condition de faible dNTPs accompagné de PCR). Cependant, ces approches permettent d'analyser un nombre limité de sites, nécessitent une grande quantité de matériel biologique et sont très laborieuses. Dans le cas de matériel biologique précieux, provenant de patients atteint de cancer par exemple, il est compliqué d'analyser la totalité des sites avec ces méthodes. Des développements technologiques récents basés sur le séquençage haut-débit du transcriptome (RNA-seq) ont permis d'identifier la

totalité des sites 2'Ome et quantifier le taux de méthylation avec une quantité limitée de matériel biologique. En utilisant ces nouvelles méthodes, plusieurs équipes ont démontré qu'il existe une hétérogénéité de la 2'Ome à la fois dans une seule population cellulaire, un seul organe ou une seule tumeur (i.e., intra-variabilité) et entre différentes conditions (i.e., inter-variabilité). Au cours de ma thèse, j'ai participé à l'écriture d'une revue scientifique dont je suis second auteur, qui compile et discute des études fournissant des observations expérimentales qui démontrent que l'ARNr 2'Ome varie naturellement dans les organismes vivants (levure, plante, poisson zèbre, souris, humain), en fonction du contexte physio-pathologique. L'ensemble des données mentionnées dans la revue, y compris des données issues de mon projet de thèse, soutiennent l'hypothèse que la 2'Ome peut contribuer à l'établissement et/ou au maintien de phénotypes particuliers, probablement par un contrôle finement ajusté de la traduction, y compris dans le cancer (JAAFAR et al., 2021).

C.1.4 Revue : 2'O-Ribose methylation of rRNAs : Natural diversity in living organisms, biological processes and diseases

Review

2'-O-Ribose Methylation of Ribosomal RNAs: Natural Diversity in Living Organisms, Biological Processes, and Diseases

Mariam Jaafar , Hermes Paraqindes , Mathieu Gabut, Jean-Jacques Diaz , Virginie Marcel * 
and Sébastien Durand * 

Inserm U1052, CNRS UMR5286, Centre de Recherche en Cancérologie de Lyon, Université de Lyon, Université Claude Bernard Lyon 1, Centre Léon Bérard, CEDEX 08, F-69373 Lyon, France; Mariam.JAAFAR@lyon.unicancer.fr (M.J.); Hermes.PARAQINDES@lyon.unicancer.fr (H.P.); Mathieu.GABUT@lyon.unicancer.fr (M.G.); JeanJacques.DIAZ@lyon.unicancer.fr (J.-J.D.)
* Correspondence: virginie.marcel@lyon.unicancer.fr (V.M.); sebastien.durand@inserm.fr (S.D.);
Tel.: +33-04-2655-6745 (V.M.); +33-04-6985-6092 (S.D.)

Abstract: Recent findings suggest that ribosomes, the translational machineries, can display a distinct composition depending on physio-pathological contexts. Thanks to outstanding technological breakthroughs, many studies have reported that variations of rRNA modifications, and more particularly the most abundant rRNA chemical modification, the rRNA 2'-O-ribose methylation (2'Ome), intrinsically occur in many organisms. In the last 5 years, accumulating reports have illustrated that rRNA 2'Ome varies in human cell lines but also in living organisms (yeast, plant, zebrafish, mouse, human) during development and diseases. These rRNA 2'Ome variations occur either within a single cell line, organ, or patient's sample (i.e., intra-variability) or between at least two biological conditions (i.e., inter-variability). Thus, the ribosomes can tolerate the absence of 2'Ome at some specific positions. These observations question whether variations in rRNA 2'Ome could provide ribosomes with particular translational regulatory activities and functional specializations. Here, we compile recent studies supporting the heterogeneity of ribosome composition at rRNA 2'Ome level and provide an overview of the natural diversity in rRNA 2'Ome that has been reported up to now throughout the kingdom of life. Moreover, we discuss the little evidence that suggests that variations of rRNA 2'Ome can effectively impact the ribosome activity and contribute to the etiology of some human diseases.

Keywords: 2'-O-methylation; rRNA modification; ribosome; ribosome heterogeneity; RiboMeth-seq



Citation: Jaafar, M.; Paraqindes, H.; Gabut, M.; Diaz, J.-J.; Marcel, V.; Durand, S. 2'-O-Ribose Methylation of Ribosomal RNAs: Natural Diversity in Living Organisms, Biological Processes, and Diseases. *Cells* **2021**, *10*, 1948. <https://doi.org/10.3390/cells10081948>

Academic Editor: Alexander E. Kalyuzhny

Received: 2 July 2021
Accepted: 29 July 2021
Published: 31 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

First observed in the mid-1950s, the ribosome is the cellular machinery that synthesizes proteins using mRNAs during the process of translation. The understanding of ribosome activities and functions has evolved with the constant improvement of structural methods [1,2]. The recently-met atomic resolution level of the ribosome's structure demonstrated that its function has been conserved throughout evolution, from bacteria to higher vertebrates, due to the high conservation of its structure, which is based on two main components: the ribosomal proteins and the ribosomal RNAs (rRNAs) [2]. In humans, ribosomes are constructed of 80 ribosomal proteins organized around 4 rRNAs: 5S, 5.8S, 18S, and 28S [2–4]. Although the number and length of ribosomal proteins and/or rRNAs vary across species, the ribosome core displays a similar structuration on which its functions of mRNA-decoding and peptidyl-transferase activities depend [2].

The conserved structuration of rRNAs relies on ribosomal proteins but also to some extent on the presence of chemical modifications at some particular rRNA nucleotides that include 2'-O-ribose methylation (2'Ome, Figure 1A), pseudo-uridylation (ψ) and base methylations and acetylations [5,6]. These chemical modifications maintain the modified base in a particular orientation, stabilize RNA:RNA interactions or, in contrast, inhibit

RNA:protein association [7,8]. Therefore, based on the strong conservation of the structure/function relationship of the ribosome, the long-lasting dogma was that rRNAs are always and similarly chemically modified and that these modifications are critical for ribosome functions. Accordingly, this vision was supported by observations using non-quantitative methods that, in the two historical models *Saccharomyces cerevisiae* and *Xenopus laevis*, rRNAs are fully methylated at almost all specific positions in all ribosomes [6,9]. Similarly, knocking out the Fibrillar protein (i.e., the rRNA 2'-O-methyltransferase, FBL, Figure 1B) induces lethality in yeast, zebrafish, and mouse models, supporting an essential role of the rRNA 2'Ome across species [10,11]. Therefore, it was broadly admitted that rRNA, and more generally the ribosome, always displays the same composition to sustain its conserved structure and thus functions, and that a disruption of this structure is detrimental for the ribosome activity in living organisms.

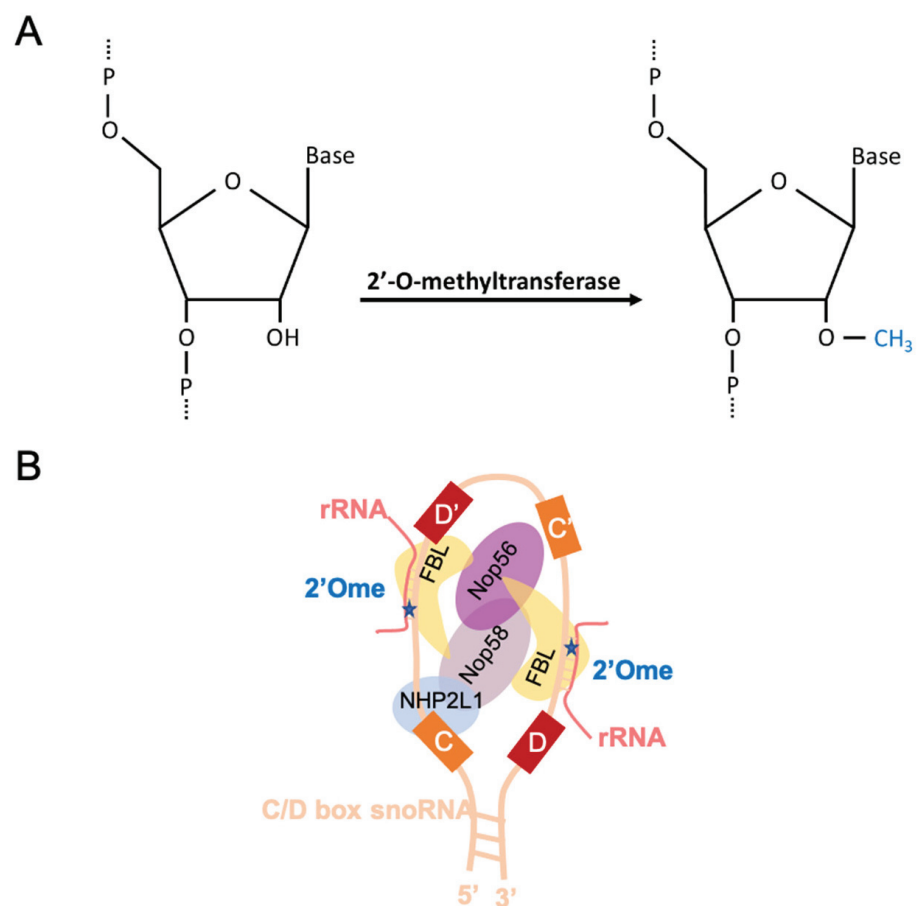


Figure 1. Schematic representation of a 2'-O-methylated ribose (A) and a C/D box snoRNP (also named the methylation complex) (B). The C/D box snoRNP includes the methyltransferase protein Fibrillar protein (light yellow), NOP56 (purple), NOP58 (light brown), NHP2L1 (light blue), and the C/D snoRNA (light orange). Boxes C/C' (orange) and D/D' (red) are shown. The rRNA forms an RNA:RNA duplex with the C/D snoRNA. The FBL catalytic domain 2'-O-methylates the fifth nucleotide downstream of the boxes D or D' and is indicated by a blue star.

However, recently, several data have suggested that ribosomes can actually display distinct and variable compositions, both at ribosomal protein and RNA levels [1,7,12–16]. The current hypothesis proposes that these changes in ribosome composition may directly affect its intrinsic translational properties and thus provide the ribosome with translational regulatory capacities. Thus, rather than simply being a machinery that synthesizes proteins, the ribosome emerges as a direct regulator of translation. However, only some data has supported this hypothesis up to now and efforts must continue to definitively universalize

this model. In contrast, we have reached a moment where technological breakthroughs provide strong and numerous evidence that variations in chemical rRNA modifications occur naturally in living organisms, biological processes, and diseases. Indeed, the scientific community has yet accumulated numerous pieces of data in less than five years, demonstrating that an rRNA molecule can subsist even in the absence of chemical modifications at some particular positions, most of these studies focusing on 2'Ome. Here, we compiled and discussed studies providing experimental observations that definitively demonstrate that rRNA 2'Ome varies naturally in living organisms, specifically depending on the physio-pathological context, however.

2. A Brief Overview of rRNA 2'Ome

2'Ome consists in the addition of a methyl group (CH₃) onto the hydroxyl (OH) of the second carbon of the ribose using the S-adenosyl-methionine (SAM) as a methyl donor (Figure 1A). Out of the 4 rRNAs, modified nucleotides have been reported on the 3 rRNAs of that derived from the common 47S precursor (i.e., 5.8S, 18S and 28S) [6]. The addition of 2'Ome has been shown to occur co-transcriptionally and to participate in the precise shaping of the rRNA structure on which the ribosome activity relies [17,18]. This observation pushed the scientific community to consider that the rRNA 2'Ome is required for building and maintaining ribosome activities.

2'Ome is catalyzed by a methylation complex, the composition of which is conserved in archaea and eukaryotes (Figure 1B). This complex contains four different proteins and one non-coding small nucleolar RNA (snoRNA) and thus forms particles called small nucleolar ribonucleoproteins (snoRNPs) [6,7,19]. The addition of 2'Ome occurs at some particular nucleotides along the rRNAs and is guided by the family of snoRNAs that exhibits C/D boxes, called SNORDs [7,20]. Indeed, SNORDs directly interact with rRNAs by sequence homology to guide the rRNA 2'O-methyltransferase FBL onto the nucleotide to be modified. Interactions between SNORDs and rRNAs depend on the particular structure of the methylation complex provided by NOP58, NOP56, and NHP2L1 (also known as 15.5kDa protein) in humans. Among archaea and eukaryotes, some 2'Ome sites do not depend on the snoRNP methylation complex but instead on individual site-specific rRNA methyl-transferases [21]. For example, the yeast 25S-Gm2922 and 25S-Um2921 are catalyzed by the stand-alone methyl-transferase Spb1, independently of C/D box snoRNAs [22,23]. Such site-specific rRNA methyl-transferases and their associated 2'Ome sites are conserved from bacteria, where only four 2'Ome sites have been described as 2'O-methylated, the 16S-Cm1402, 23S-Gm2251, 23S-Cm2498, and the highly conserved 23S-Um2552, which are catalyzed by the stand-alone rRNA methyl-transferases RsmI (YraL), RlmB (YfjH), RlmM (YgdE), and RlmE (RrmJ or Ftsj), respectively [24]. Thus, one could easily hypothesize that snoRNA-dependent methylation most likely overcame the complexification of the number of 2'Ome through evolution. Indeed, historical models reported increasing numbers of 2'Ome rRNA positions, from 3 in bacteria, 55 in yeast, and more than 106 sites in humans, which also correlates with the increased length of rRNAs (www-snoRNA-biotoul.fr, accessed on 29 July 2021). Moreover, most of the 2'Ome sites specifically found in eukaryotes concentrate in ribosome functional domains, which are not 2'O-methylated in bacteria, despite being strongly conserved [6]. This observation supports the notion that the presence or absence of 2'Ome at some particular sites may occur without affecting the conserved function of the ribosome.

Similar to the structure of the ribosome, the mapping of the 2'Ome positions across species evolves concomitantly with the improvement of technologies [25,26]. Historically, the 2'Ome detection relied on biochemical and molecular approaches, including, for example, mass spectrometry, primer extension (i.e., detection of reverse transcription pauses using radiolabeled primers) or RTL-PCR (Reverse Transcription Low dNTPs-PCR), a brief description of which is provided in Table 1. The main limitation of these methods remains that several independent experiments are required to fully map the total > 7kb of rRNAs, thus requiring a huge amount of biological materials and labor. The recent development of

RNA-seq-based approaches to map rRNA 2'Ome has opened up several avenues in the field [17,27] (briefly described in Table 2; for review and comparison of these methods, see [25]). From 2016, these methods generated an accumulation of data for all rRNA nucleotides in a single experiment using a limited amount of biological material. These novel approaches not only allow the mapping of 2'Ome in novel living organisms, and thus the study of the conservation of rRNA 2'Ome localization through evolution, but also permit researchers for the first time to quantitatively compare rRNA 2'Ome levels at all sites, both in a single cellular population, organ, or tumor (i.e., intra-variability) and between different conditions (i.e., inter-variability).

Table 1. Description of the different methods used to quantify and detect the 2'Ome.

Method	Principle	2'Ome Site	2'Ome Detection	Reference
Primer Extension	The presence of a 2'Ome residue causes the Reverse Transcriptase to pause at low dNTP concentration, the pause being directly observable on a gel using a radioactive/fluorescent primer	site specific	qualitative	[28]
RTL-P	The presence of a 2'Ome residue decreases the reverse transcription efficiency at low dNTP concentration, the pause being quantified by regular or quantitative PCR	site specific	semi-quantitative	[29]
Mass Spectrometry	Detect rRNA fragments or nucleosides, the mass of which is shifted by specific modification including 2'Ome	global overview of all sites + site specific	quantitative (absolute)	[30]
RiboMeth-seq	The presence of a 2'Ome inhibits alkaline-mediated hydrolysis of the 3'-adjacent phosphodiester bond and is then detected by RNA-seq (Illumina or Ion Torrent sequencing)	global overview of known sites + site specific	quantitative (absolute)	[17]
2'Ome-seq	The presence of a 2'Ome residue causes the Reverse Transcriptase to pause at low dNTP concentration and is then detected by RNA-seq (Illumina)	global overview of all sites + site specific	quantitative (relative)	[31]
RibOxi-seq	The 3'-terminal ribose of a 2'Ome residue is resistant to periodate cleavage. After RNA fragmentation using the endonuclease Benzoylase and periodate oxidation, RNA fragments with a 2'Ome group at their 3' end are enriched and detected by RNA-seq	global overview of all sites	qualitative	[32]

3. Natural Variability of rRNA 2'Ome in Yeast

To date, biochemical mapping and sequencing-based profiling approaches have identified 55 2'Ome sites in *Saccharomyces cerevisiae* rRNAs: 18 sites in the 18S rRNA, 37 sites in the 25S rRNA and no sites in neither the 5S nor the 5.8S rRNAs [6,17,27,33,34]. Several pieces of evidence have supported the notion that rRNA 2'Ome is critical for ribosome biogenesis in yeast. First, the recent comparison of the 2'Ome patterns in chromatin-associated (i.e., nascent transcripts) and mature rRNAs showed that the majority of the 2'Ome occurring on the 18S rRNA was co-transcriptional, while the 2'Ome on the 25S rRNA seemed to occur both co- and post-transcriptionally [17], in agreement with previous results obtained by kinetic labeling experiments [18]. Therefore, rRNA 2'Ome appears absolutely crucial for the early first steps of pre-rRNA folding and the assembly of ribosomal particles. Second, similar to other eukaryotes, the catalysis of 2'Ome in *S. cerevisiae* is

driven by the methyl-transferase Nop1, the yeast homologue of vertebrate FBL, which is highly conserved among all species [35]. Conditional depletions of Nop1 alleles are lethal and result in both impairments of pre-rRNA 2'Ome and processing over long periods of times [36]. Similarly, a Nop1 mutant unable to bind the SAM, and thus abolishing its methyl-transferase activity, severely impaired growth, supporting the importance of 2'Ome in yeast [35]. Altogether, these data support the essential role of rRNA 2'Ome in bona fide ribosome biogenesis and yeast growth.

Interestingly, rRNA modifications, including 2'Ome, are not randomly scattered across the ribosome. The 2D and 3D structures of the 18S and 25S rRNAs revealed that the majority of chemically modified nucleotides conserved between different yeast species, such as *S. pombe* and *S. cerevisiae*, cluster in functionally important domains of the ribosome [30,33]. Accordingly, these conserved modified nucleotides are mainly located within the peptidyl-transferase center (i.e., the formation of the peptidyl-transferase bound between two amino-acids), the decoding center (i.e., the recognition between mRNA codon:tRNA anti-codon) and the inter-subunit bridges (i.e., the interaction between the small and the large subunits), suggesting that their functional importance is fundamental and universal. It was therefore predicted that individual depletions of any rRNA chemical modifications could highly impact both ribosome biogenesis and ribosome translational activity. Surprisingly, however, the loss of individual modifications in yeast only impacted slightly, or not at all, with cell viability or ribosome functions [37,38]. Chemical modifications instead act in a cumulative manner, as significant phenotypes were only observed upon the loss of multiple modifications. For instance, the helix 69 (H69) from the inter-subunit bridge, which interacts with both A and P sites, contains five modifications in yeast that are all conserved in humans: one 2'Ome (25S-Am2256) and four pseudo-uridines (25S-Ψ2258, 25S-Ψ2260, 25S-Ψ2266 and 25S-Ψ2264). The loss of one or two chemical modifications through the depletion of corresponding snoRNAs results in little or no effect, whereas the combinatorial loss of three to five modifications has many effects, including an impairment of cell growth, an increased translation rate and termination at stop codons, a reduced accumulation of rRNA, as well as alterations of ribosome structure and an increase in sensitivity to antibiotics [37]. In addition, the loss of three modifications (18S-Gm1428, 18S-Cm1639, and 18S-m1acp3Ψ1191) in the decoding center reduced growth rates and protein synthesis efficiency and caused a strong deficiency in small subunit production [38]. Although these studies do not address phenotypes caused by only 2'Ome depletions, it nonetheless demonstrates that different combinations of rRNA chemical modification depletions affects the ribosome to various extents, however without being systematically lethal. Hence, unlike initially thought, yeast ribosomes can handle partial losses of chemical modifications.

Importantly, although it was originally thought that individual 2'Ome were of little impact on cell growth and translation, this idea was later challenged in yeast by high-resolution phenotyping experiments. This approach indeed revealed that individual snoRNA depletions could result in important underlying deregulations under stress conditions [39]. Systematic individual deletions of 20 SNORDs result, in most cases, in the alteration of growth adaptation, while the strain's growth rate is slightly or not at all affected. In addition, the majority of these SNORD deletions result in resistance to several antibiotics including neomycin, paromomycin, and anisomycin, while others result in antibiotic-sensitivity [39]. Interestingly, differences in phenotypic impacts on growth adaptation were not correlated to the spatial distribution of 2'Ome, therefore suggesting that each single 2'Ome fulfills a specific function during rRNA structuration. These data also suggest that the loss of individual rRNA 2'Ome is not strictly critical for ribosomes, although it can modulate ribosome functions in translational regulations in response to environmental changes.

Interestingly, such data not only challenged the long-standing assumption that 2'Ome sites must be constitutively present on specific rRNA residues, but also fed the arising concept that ribosomal heterogeneity naturally occurs. Intriguingly, the level of 2'Ome at each individual position in yeast rRNAs is not the same, and a subset of 2'Ome positions

was found to be partially methylated [17,33]. This implies that, in a single yeast population, a mix of non-methylated and fully methylated rRNA molecules at a given position co-exists, corresponding to the notion of rRNA 2'Ome intra-variability. The first study to demonstrate a ribosome heterogeneity at the rRNA 2'Ome level in *S. cerevisiae* used RNA-cleaving DNazymes coupled to LC-UV MS/MS analysis to show that the methyl group at the ribose of nucleotide A at position 100 of the 18S rRNA (18S-Am100) was missing in one-third of 18S rRNA molecules, therefore revealing an intra-variability of this residue in a single yeast strain [40]. Notably, ribosomes devoid of this 2'Ome still participated in translation, but the exact function of these particular ribosomes was not yet elucidated [40]. Later on, many other studies using different high throughput techniques, including high resolution RP-HPLC analysis, SILNAS-based MS-technology and RiboMeth-seq, revealed that 18S-Am100 was not the only site to be partially methylated in yeast. In fact, while for a given ribosome population the majority of the sites are fully methylated with a methylation score > 0.95, other positions are only partially methylated and display a methylation score < 0.8–0.9 [17,27,33,34]. Interestingly, partially methylated nucleotides are located at similar positions in both *S. cerevisiae* and *S. pombe* rRNAs, therefore indicating that 2'Ome intra-variability displays some degree of conservation [33].

Biological explanations on how and why intra-variability takes place in yeast still remains to be elucidated. We can, however, easily hypothesize that rRNA 2'Ome variability may be adjusted by the regulation of the methylation complex assembly and/or association to targeted pre-rRNA under different contexts. In particular, the modulation of the pre-rRNA/SNORD association is a crucial event during ribosome biogenesis, that is usually regulated by SNORD expression and processing or the catalytic activity of RNA helicases [41,42]. Along these lines, it was shown that Ppr43 and Dbp3, two RNA helicases implicated in the release of several snoRNAs from pre-rRNAs, may modulate rRNA 2'Ome through the regulation of snoRNAs dynamics [43]. Aquino and colleagues showed that Dbp3 depletion in yeast reduced the level of 2'Ome at 18 sites of the 25S rRNA and at 1 site of the 18S rRNA, most likely by trapping snoRNAs on pre-rRNA and therefore interfering with the access of adjacent snoRNPs to mutually exclusive overlapping regions required for the proper completion of 2'Ome. Similarly, Prp43 depletion led to the reduction of 2'Ome level at 8 sites in the 18S rRNA and 25 in the 25S rRNA by a mechanism similar to Dbp3, albeit regulating a specific subset of SNORDs. Although this experimentally driven study constitutes nonetheless strong evidence of 2'Ome modulations by RNA helicases, additional mechanisms that regulate the natural intra-variability in rRNA 2'Ome observed in yeast remain to be explored.

4. Variation of rRNA in Metazoan Cell Lines

In addition to yeast models, the variability in rRNA 2'Ome level has also been reported in metazoan cell lines, mostly in human cells and to a lower extent in mouse cells. Hence, in addition to supporting the notion of intra-variability of rRNA 2'Ome within a single cell population as observed in yeast, ex vivo metazoan models also raised the notion that rRNA 2'Ome varies depending on the physio-pathological context, including cell lines harboring clinically relevant specific disease-related gene mutations (Figure 2).

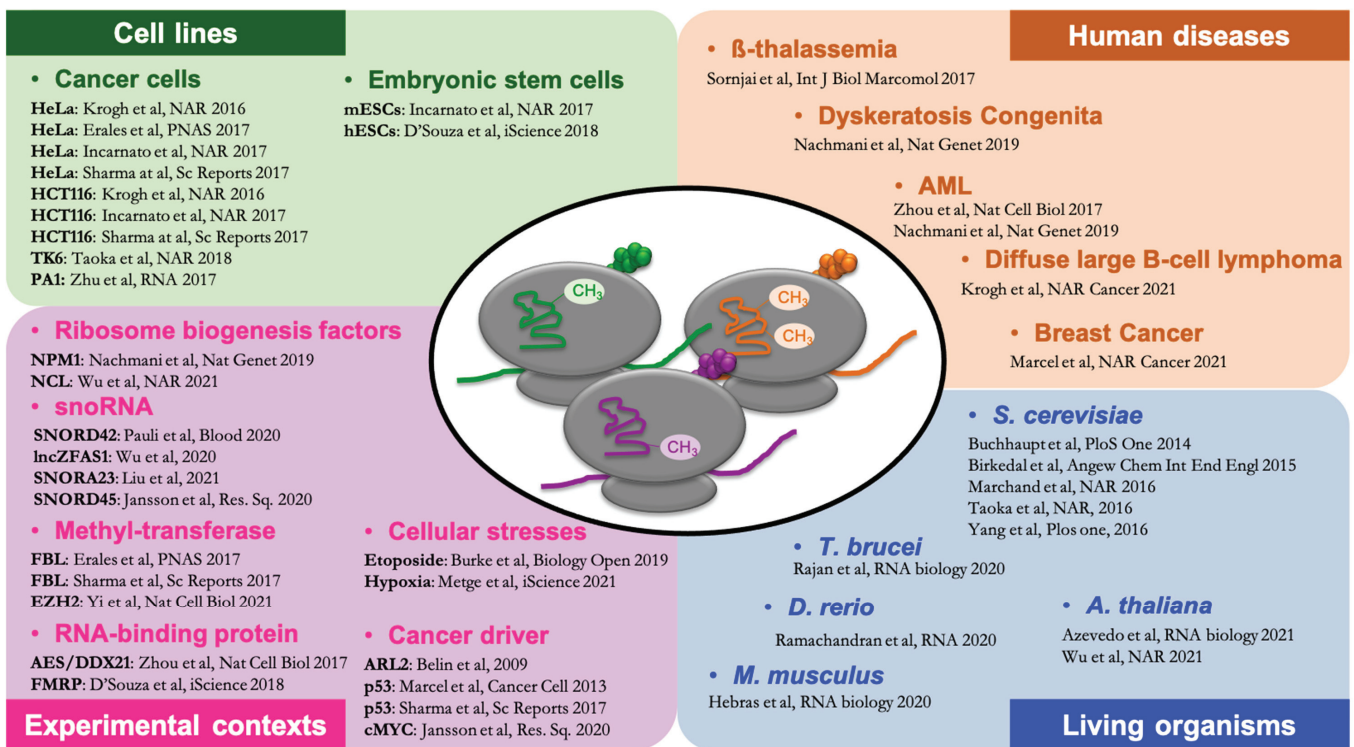


Figure 2. Variability of rRNA 2'Ome in living organisms, biological processes, and human diseases.

4.1. Intra-Variability of rRNA 2'Ome in Human Cell Lines

The existence of variations in rRNA 2'Ome in humans first emerges from analyses of individual human cell lines. Albeit supported by a relatively low number of studies, evidence for the existence of partially methylated rRNA within a single human cell line has started to emerge in 2016 with the description of RNA-seq based approaches [17,26]. Using the RiboMeth-seq approach, which relies on the property of the 2'Ome to protect the methylated nucleotide from alkaline hydrolysis, it was first demonstrated that about one-third of the 106 sites are partially methylated within a single human cell line [26,44,45]. Indeed, the comparison of rRNA 2'Ome levels within a cell population issued from either the cervix adenocarcinoma HeLa cell line or the colorectal carcinoma HCT116 cell line indicated that most of the rRNA 2'Ome positions are fully methylated and that only a subset of rRNA 2'Ome positions are partially methylated. Therefore, the existence of partially methylated sites at particular positions implies a mix of fully methylated and non-methylated molecules in the bulk of analyzed human rRNAs. Furthermore, another study using the high-throughput RibOxi-seq approach confirmed the presence of partially methylated sites in the human ovary teratoma-derived PA1 cell lines [32].

While the next-generation sequencing methods are starting to be widely used, especially RiboMeth-seq, Taoka and colleagues used a quantitative mass spectrometry approach to quantify the level of rRNA modifications in the TK6 cell line (human lymphoblast from spleen) [46]. In addition to quantifying methylation at the 106 known sites, they also identified and quantified six additional rRNA 2'Ome sites (18S-Um354, 18S-Cm621, 18S-Um1668, 28S-Am1310, 28S-Um1760, 28S-Gm3606). Among all of these, about one-third ($n = 40$) was partially methylated and displayed a methylation ratio $< 90\%$, therefore cohering with the proportion of partially methylated sites observed in HeLa cells using RiboMeth-seq [26,44,46]. Interestingly, high-resolution structures of HeLa ribosomes obtained by cryo-EM also identified a subset of rRNA sites partially methylated, although all the sites have not been analyzed by cryo-EM due to heterogeneity in high-resolution within the ribosome structure [1,5].

Altogether, these studies demonstrated that, as in yeast, the presence of 2'Ome at some particular rRNA positions is not compulsory in human cell lines. Cells can thus tolerate the production of ribosome with an absence of 2'Ome, however not at all of the rRNA 2'Ome positions, but only at about one-third of them. However, the distinct sensitivity and scoring of the diverse technical approaches prevent a direct comparison of the intra-variability in these different human cell lines. Moreover, analyses of cell lines in bulk do not allow us to establish whether the 2'Ome intra-variability results from intra-cellular 2'Ome variations of rRNA molecules or, in contrast, reflects a cell-to-cell heterogeneity of 2'Ome. Nevertheless, these data indicate that ribosomes of distinct 2'Ome compositions co-exist in human cell lines, supporting the notion of ribosome heterogeneity.

4.2. Physiopathological Variability of rRNA 2'Ome in Cell Lines

In parallel to the increasing evidence for natural intra-variability of 2'Ome in cell lines, several studies have pointed out that artificial and experimental manipulations of genes, involved in ribosome biogenesis or physio-pathological processes, may induce variations of 2'Ome between two different biological conditions, supporting the notion of inter-variability in rRNA 2'Ome (Figure 2).

4.2.1. rRNA 2'Ome Variability in Response to Modulations of Ribosome Biogenesis Factor (RBF) Expression and Activity

Some studies first examined the impact of RBFs on rRNA 2'Ome regulations, such as the recent work that Wu and colleagues dedicated to the long non-coding RNA *ZFAS1*, a host gene that encodes three SNORD12 family members (SNORD12, SNORD12B, and SNORD12C) [47]. In a panel of human colorectal cancer cell lines, it appeared that the depletion of *ZFAS1* decreases the level of rRNA 2'Ome at 28S-Gm3878 and 28S-Gm4593, guided by SNORD12C and SNORD78, respectively. Indeed, in addition to encoding SNORDs,

ZFAS promotes the assembly of SNORD12C- but also SNORD78-containing snoRNP complexes, through a direct interaction with a member of the methylation complex, NOP58. Although the ZFAS1 knockdown decreases cell proliferation and invasion of colorectal cancer in vitro and in vivo, the direct implication of SNORDs and rRNA 2'Ome has not been demonstrated. In addition, the impacts of other snoRNAs, the expression of which is deregulated in cancer, on variations of rRNA 2'Ome has also been reported. For instance, SNORA23, which is an H/ACA box snoRNA downregulated in hepatocellular carcinoma tissues and cell lines, surprisingly functions by impairing the 2'Ome at 28S-Cm4506 in Huh7 cell lines [48]. In this study, SNORA23 expression was shown to be regulated by the PI3K/AKT/mTOR signaling pathway and the increased SNORA23 expression inhibits proliferation, migration, and invasion, although the contribution of 28S-Cm4506 has not been addressed.

Along the lines of the ZFAS study, many other studies have demonstrated that modulations of methylation complex expression and formation induced variations of rRNA 2'Ome in cell lines. For instance, *FBL* depletions in HeLa cells caused a global decrease in rRNA 2'Ome levels at most, albeit not all, positions [44,49]. As a matter of fact, RiboMeth-seq analyses showed that the decrease in rRNA 2'Ome was surprisingly site-specific, rather than systematic. In addition, partially methylated sites exhibited a peculiar sensitivity to *FBL* depletion, as sites with a methylation score < 0.8 in control HeLa cells decreased by at least 10% upon *FBL* depletions. Importantly, functionally critical regions of the ribosome, including the peptidyl-transferase and decoding centers, which contain the most conserved modification sites, were devoid of 2'Ome alterations upon *FBL* depletions, in contrast to the inter-subunit bridges, peptide exit tunnel, and A and P-sites [44]. Notably, the decrease in rRNA 2'Ome at particular sites upon *FBL* depletions was neither the result of a reduction in rRNA production nor a consequence of changes in corresponding snoRNA expression [44,49]. These data suggest that the full methylation at specific sites seems to be required for ribosome structure and functions, while partial methylation of others can be tolerated. Thus, one can hypothesize that modulations of *FBL* activity and expression could be a means to finely tune ribosome functions by modulating the 2'Ome of specific sites without affecting the full methylation at key positions. Another study has recently highlighted the importance of the methylation complex formation in the regulation of rRNA 2'Ome [50]. EZH2, a lysine methyltransferase catalyzing histone H3 lysine 27 trimethylation (H3K27me3) and member of the polycomb repressive complex 2 (PRC2) was found to interact with components of the rRNA methylation complex and thus play a role in 2'Ome independently from its chromatin-related functions through histone methylation [50]. More precisely, it was shown that EZH2 binds strongly and directly to both the methylation complex proteins *FBL* and NOP56. EZH2 is required for the assembly of the methylation complex as its depletion impairs the assembly of core proteins NOP56, NOP58, and SNU13. Accordingly, EZH2 depletions reduce site-specific 2'Ome in prostate C4-2 cancer cells compared to control cells, with 2 sites affected in the 5.8S rRNA, 31 in the 18S rRNA, and 54 in the 28S rRNA. Importantly, the *FBL* overexpression was found to rescue the decrease in 2'Ome caused by EZH2 depletions, therefore suggesting that EZH2 modulates 2'Ome through *FBL* methyl-transferase activity. Finally, EZH2 promotes the Internal Ribosome Entry Site (IRES)-dependent translation of several genes, including XIAP, although the implication of *FBL*/NOP56 or rRNA 2'Ome in this process remains to be proven [50].

4.2.2. rRNA 2'Ome Variability in Physio-Pathological Contexts

The question of an inter-variability of 2'Ome occurring in physio-pathological contexts as well as its contribution to disease's etiology has recently gained much attention. Indeed, by using human cell lines harboring clinically relevant genetic mutations, some recent studies have provided evidence that a 2'Ome variability occurs in specific pathologies, although only few of them provide convincing data supporting the key contributions of 2'Ome to disease's phenotypes (Figure 2).

The fragile X syndrome results from the loss-of-function of the RNA-binding Fragile X Mental Retardation Protein (FMRP), which has been recently connected to the regulation of rRNA 2'Ome in human Embryonic Stem cells (hESC) [51]. D'Souza and colleagues observed that FMRP interacts with several SNORDs to form FBL-independent snoRNPs. Importantly, RiboMeth-seq analyses showed that *FMRP* knock-out in hESC significantly affected 2'Ome at 17 positions, including 18S-Cm174, 18S-Um428, 18S-Gm867, 18S-Um1248, 28S-Um2402, 28S-Cm2409, 28S-Um4276, and 28S-Cm4917 (nomenclature from original article corrected). As reported with FBL, *FMRP* depletions mostly resulted in alterations (although both up and down) of partially methylated sites, while fully methylated ones remained unaffected [44,49,51]. Since levels of SNORDs guiding methylations at these positions were not affected, further studies will be required to get more of an insight into the mechanisms by which FMRP either promotes or represses specific rRNA 2'Ome. Moreover, the authors did not investigate whether patients with FMRP-dependent fragile X syndrome display 2'Ome alterations and whether these potential alterations contribute to disease phenotypes.

Another outstanding illustration of the implication of rRNA 2'Ome variations in human genetic disorders has been provided by a recent work focused on the ribosome biogenesis factor nucleophosmin 1 (NPM1) [52]. Nachmani and colleagues observed that artificial depletions of *NPM1* in mouse cell lines altered 2'Ome at positions 28S-Cm1327, 28S-Am3764, 28S-Cm3866, 28S-Um3904, and 28S-Gm4198 (human nomenclature). In addition, the human leukemia OCI-AML3 cell line, which expresses a mutated form of NPM1 sequestered in the cytoplasm (NPMc⁺) and frequently found in Acute Myeloid Leukemia (AML) [53], harbors reduced 2'Ome at positions 28S-Cm1327, 28S-Cm3866, 28S-Um3904, and 28S-Gm4198, while respective SNORD levels remains unchanged. In the same study, Nachmani and colleagues also identified two NPM1 mutations, NPM1^{D178H} and NPM1^{D180del}, in patients with Dyskeratosis Congenita (DC), where causative mutations could not have been identified. These mutations reduced NPM1 abilities to interact with snoRNAs (SNORD15/47/52/58/104) and decreased the 2'Ome at corresponding positions (28S-Am3764, 28S-Cm3866, 28S-Um3904, 28S-Gm4198, and 28S-Cm1327). Additionally, a mouse model carrying the NPM1^{D180del} mutation recapitulates DC pathological features. Interestingly, NPM1 depletions, cytoplasmic retention of NPM1 (NPMc⁺), and NPM1^{D178H}/NPM1^{D180del} mutations impacted the translation of specific IRES-containing transcripts. Altogether, the study by Nachmani et al. provided solid evidence that a variability of 2'Ome at specific positions exists in certain forms of NPM1-dependent DC and leukemia, and most likely participates in the physiopathology of these diseases. However, further investigations should carefully address the contribution of rRNA 2'Ome-dependent translational modulations to the etiology of the diseases.

Several other studies have highlighted the existence of a 2'Ome variability in other forms of AML and have demonstrated its contribution to leukemogenesis in a NPM1-independent manner [54,55]. Hence, the oncogenic AML1-ETO fusion protein, which results from the most common balanced translocation in AML (t(8:21)), plays essential roles in hematopoietic progenitor cell (HPC) self-renewal, in part by promoting the expression of groucho-related Amino-terminal Enhancer of Splicing (AES) [56–58]. Strikingly, *AES* knock-down in the human Kasumi-1 cell line impacted the abundance of many SNORDs and, accordingly, *AES* depletion tremendously reduced 2'Ome at three particular positions (18S-Cm1703, 18S-Gm1328, and 28S-Cm4506) while pseudo-uridylation was only weakly affected [55]. Notably, the authors did not perform high-throughput analyses of rRNA 2'Ome but most likely focused on positions guided by candidate SNORDs, the level of which was affected by *AES* depletion. Mechanistically speaking, the authors showed that AES interacts with the RNA helicase DDX21 and is required for DDX21 interactions with proteins involved in the methylation complex, including FBL, NOP56, and NOP58, as well as in pre-rRNA and rRNA processing complexes. Interestingly, AML1-ETO positive patients' primary AML blasts overexpress specific SNORDs, including SNORD14D/14E/20/32A/34/35A/43/53/104 and displays increased 2'Ome at positions

18S-Um1805, 28S-Um4197, 28S-Cm3848, 18S-Cm1703, and 18S-Gm1328. Furthermore, a set of snoRNAs, including SNORD20/34/35A/43, were found to be associated with AML displaying a high leukemic stem cell content and a strong treatment resistance. Altogether, this remarkable work by Zhou et al. demonstrated that increased snoRNA expression, and therefore 2'Ome levels at specific rRNA positions, occurs during AML1-ETO-dependent leukemogenesis and contributes to both the physiopathology and the severity of the disease. In a follow-up snoRNA screening study, the team of Müller-Tidow also identified SNORD42A and 18S-Um117 as major actors of leukemogenesis in AML cell lines and patient's blast [54]. Altogether, these studies shed light on the role of specific 2'Ome in leukemogenesis and possibly normal hematologic functions. On one hand, rRNA 2'Ome are required for normal hematopoiesis and bone marrow functions, and the variability of 2'Ome at these positions is observed in NPM1-related leukemia or DC and contributes to the disease's phenotypes [52]. On the other hand, high expressions of specific SNORDs and elevated 2'Ome levels at related specific positions are observed in AML1-ETO-, and possibly C-MYC- and MLL-AF9-associated AML, and participate in leukemogenesis [54,55]. Of note, while the FMRP- and NPM1-mediated 2'Ome variabilities imply changes in methylation complex formations without affecting the SNORD expression, AML1-ETO alternatively controls specific 2'Ome changes by regulating the expression of specific SNORDs.

Additional key oncogenes and tumor suppressors have been associated with alterations of rRNA 2'Ome. A preprint study available on Research Square focused on the MYC proto-oncogene [59]. BJ^{hTERT} fibroblasts expressing MYC in an inducible manner showed that MYC overexpression is associated with an increase in *FBL* expression and changes in 2'Ome levels at three sites: a decrease at 28S-Um2031 and 28S-Um2402, and an increase at 18S-Cm174 [59]. Interestingly, SNORD45C, which targets 18S-Cm174, was also found up-regulated under MYC overexpression. SNORD45C knock-out and the related abrogation of 2'Ome at 18S-Cm174 in HeLa cells resulted in translational deregulations of a large number of specific mRNAs involved in cell cycle, mitosis, metabolism, and intracellular transport, for example [59]. Interestingly, mRNAs up- and down-regulated upon SNORD45C depletion displayed a strong variation in elongation-related codon compositions, suggesting that translational defects caused by SNORD45C depletion might depend on ORF sequences and codon usage. However, it is unclear whether or not the effects observed by the authors were directly dependent on perturbations of rRNA 2'Ome.

In addition, several studies focused on *TP53*, a tumor suppressor gene that is frequently inactivated in solid tumors mainly by mutations (for reviews [60,61]) and investigated its functions in rRNA 2'Ome. It is now well established that p53 suppresses RNA Pol I transcription and its inactivation in cancer is therefore directly associated with the up-regulation of RNA Polymerase I activity and ribosome production [62]. Thus, it was hypothesized that p53 could also affect rRNA 2'Ome. Using an isogenic colorectal cancer cell line HCT116, Sharma and colleagues reported using RiboMeth-seq, and that, among the 22 partially methylated sites, 13 sites showed a decrease in rRNA 2'Ome level in p53 -/- cell line compared to p53 +/- one, demonstrating that some 2'Ome sites are particularly sensitive to the absence of functional p53 [49]. Of note, it was reported that *FBL* is a direct target of p53 [63]. In human mammary epithelial cells (HME) and in the HCT116 colorectal cell lines, p53 inactivation increases *FBL* expression and promotes a significant alteration of 2'Ome levels at several specific sites in the 5.8S, 18S and 28S rRNAs, although RTL-qPCR technic failed to draw a full overview of alterations of rRNA 2'Ome. Importantly, the alteration of rRNA 2'Ome was not correlated to expression levels of corresponding SNORDs, suggesting once more that 2'Ome defects are strictly dependent upon SNORDs expression [63]. Importantly, in agreement with 2'Ome responses to NPM1 alterations, p53- and *FBL*-induced alterations of rRNA 2'Ome impact IRES-dependent translation of some mRNAs, and in vitro translation assays have demonstrated that variations in rRNA 2'Ome are responsible for changes in IRES-dependent translation [44,63]. Altogether, these data demonstrate a direct link between p53, *FBL* and site-specific rRNA 2'Ome regulation, although the reasons behind the differential sensitivity of specific 2'Ome to p53 remains

an open question that needs to be addressed. Interestingly, variations in rRNA 2'Ome levels have also been shown to increase at some specific sites in cancer cells displaying enhanced aggressivity [64]. Hence, human breast cancer MCF-derived cells, which exhibit a reduced ARL2 (ADP ribosylation factor like 2) expression associated with enhanced in vivo and in vitro tumor aggressivity, display increased 2'Ome levels at six out of eleven sites analyzed in the 18S and 28S rRNAs compared to wild-type cells. Although this observation needs to be broadened by systematically analyzing all rRNA 2'Ome sites, it is nonetheless a clear indicator of qualitative alterations of ribosome composition in aggressive tumors.

Finally, a few pieces of evidence suggest that cellular stresses could also impact 2'Ome. Hence, Metge and colleagues investigated the effect of hypoxia on ribosome biogenesis in T47D and MCF10 breast cancer cell lines and found that not only does hypoxia up-regulate RNA pol I activity and rRNA synthesis compared to normoxia but also modifies 2'Ome of seven out of fourteen sites measured by RTL-qPCR [65]. Interestingly, ribosomes bound to the IRES located in the 5'UTR of *VEGF-C* mRNA displayed distinct 2'Ome patterns in hypoxia compared to normoxia, therefore suggesting that *VEGF-C* mRNA could be translated by a distinct pool of ribosomes upon changes in oxygen supply, with potential consequences during tumorigenesis. Although the molecular mechanism behind the modulation of 2'Ome during hypoxia remains to be fully elucidated, it seems that a decrease in the expression of NMI ("N-Myc and STAT interactor") may impede the formation of methylation complexes containing specific snoRNAs. In addition to hypoxia, genotoxic stresses induced by etoposide, a Topoisomerase II inhibitor commonly used as a treatment for multiple cancers, has been shown to promote changes in 2'Ome [66]. Accordingly, etoposide increases levels of SNORD68 and SNORD111B and therefore augmented 2'Ome at 28S-Am2388 and 28S-Gm3923 in HeLa cells. Considering the wild impact of etoposide on cellular processes, the mode of action could be quite complex and multimodal. However, the authors proposed that etoposide could interfere with SMN and Drosha functions in the processing of nucleolus-enriched, small Cajal body-specific (sca)RNAs, which were hypothesized to regulate rRNA methylation complex activity [66–68].

Altogether, these data support that intra- and inter-variability of rRNA 2'Ome occurs in physiologically and clinically relevant cellular models (Figure 2). Profiling of rRNA 2'Ome should now be performed in human disease-related samples to definitively demonstrate that rRNA 2'Ome can occur in these different pathological contexts.

5. Diversity of rRNA 2'Ome Profiles in Human Cancer Tissues

Until the recent advances in omics approach to analyzing rRNA 2'Ome using RNA-seq-based methodologies, demonstrating the natural variation in rRNA 2'Ome, as well as evaluating its diversity in human tissues, was a real challenge. Indeed, due to the usual small quantity of available human tissue material, only methods related to RTL-P could be performed and therefore only a limited number of sites could be addressed [29,64]. To overcome the material quantity limitation, the earliest studies mainly investigated human samples from hematological diseases. Cell isolations from blood samples indeed allowed purifications of large amounts of total RNA that can be analyzed by RTL-P. Thus, by using human blood samples, two studies reported alterations in rRNA 2'Ome in acute myeloid leukemia (AML), albeit in different contexts of oncogenic alterations [52,55]. A significant increase in rRNA 2'Ome level at two out of the eleven rRNA sites analyzed was observed in five AML patients carrying the fusion AML1-ETO alteration compared to five healthy donors (i.e., 18S-Cm1703 and 18S-Gm1328) [55]. However, only three of the AML samples causes this increase, suggesting that the specific increase in rRNA 2'Ome at 18S-Cm1703 and 18S-Gm1328 is not a universal trait of AML1-ETO AML patients. In contrast to AML1-ETO1-driven AML, a comparison of 16 NPM1c⁺-driven AML patient samples with 14 samples from patients with clinical remission using RTL-P revealed a decrease at the four sites analyzed (i.e., 28S-Cm1327, 28S-Cm3866, 28S-Um3904, and 28S-Gm4198). Interestingly, the two studies found different changes in rRNA 2'Ome at different sites in AML patients compared to healthy donors, suggesting a specificity of 2'Ome

variations depending on the nature of the oncogenic activation. At present only one study investigated variation in rRNA 2'Ome in non-cancer-related disease [69]. In a cohort of 15 heterozygous β -thalassemic patients and 15 healthy donors, a significant increase in the level of rRNA 2'Ome in two out of the three rRNA sites analyzed by RTL-qPCR (i.e., 28S-Am1858 and 28S-Cm4506) was observed, further supporting the fact that specific alterations of rRNA 2'Ome at some particular sites occur in human disease. Although RTL-P-based approaches allowed researchers to gain an insight into the variability of rRNA 2'Ome at some particular sites in human pathological tissues, such approaches failed to draw the complete landscape of rRNA 2'Ome profile diversity in these conditions.

In late 2020, the first evaluation of the complete diversity of rRNA 2'Ome in human cancer samples was established using the RiboMeth-seq technology by two labs in back-to-back publications [70,71]. As in previous RTL-P analyses performed on blood samples, one study established the rRNA 2'Ome profile of 17 diffuse large B-cell lymphoma (DLBCL) patient samples and three healthy donors using an IonTorrent-based RiboMeth-seq approach [70]. In this cohort, the difference between rRNA 2'Ome levels of each individual DLBCL samples and the mean of rRNA 2'Ome levels of normal samples was estimated for 110 rRNA 2'Ome sites. Although some DLBCL samples showed a significant increase in their rRNA 2'Ome at some sites compared to mean normal samples (i.e., 12 sites), most of the significant difference corresponded to a decrease in rRNA 2'Ome at some sites (i.e., 65 sites). Two main hypotheses might explain these variations in rRNA 2'Ome. First, the decrease in SNORD expression may be responsible for the decreased rRNA 2'Ome levels, especially for the three sites with the most important variation, as observed in AML1-ETO-dependent AML [55]. Second, the global decrease in rRNA 2'Ome could be caused by a passive effect driven by the hyperproliferation of cells, as the significant decrease in 2'Ome correlates with an increase in the proliferative Ki67 labelling index [70]. Indeed, rRNA synthesis is hyperactivated in cancer cells to maintain a high level of protein synthesis, essential for sustaining an hyperproliferative rate [72]. In this condition, a decoupling between the hyperproduction of rRNA and the limited expression of factors involved in rRNA 2'Ome, such as SNORDs, might explain a broad decrease in rRNA 2'Ome level. Based on this latter hypothesis, variations in rRNA 2'Ome can be interpreted as a biological noise. However, the observed increase in the rRNA 2'Ome level in some DLBCL and AML1-ETO-driven AML patients, and the correlation between SNORD expression and rRNA 2'Ome levels suggests that the variation of rRNA 2'Ome, at least at some sites, is cancer-specific or even tumor-specific [32,52,70]. In addition to being specific in regard to the nature of the tumor, rRNA 2'Ome levels are also different at some particular sites when comparing three distinct normal human organs (i.e., brain, liver, and skeletal muscle), supporting the tissue-specificity of rRNA 2'Ome [70].

The second study performed the world's first analysis of rRNA 2'Ome (i) as a global profile rather than in site-by-site, thanks to the Illumina-based RiboMeth-seq approach, (ii) in solid tumors, and (iii) using a large series compiling more than 200 human samples [71]. Thus, the analysis of 195 primary, non-metastatic breast tumors at 106 rRNA 2'Ome sites supports the previous observations obtained in cell lines: within a single human sample, a mix of differentially methylated rRNA molecules was present, with sites being always methylated while others existed either methylated or not. As in previous studies, specific rRNA 2'Ome alterations were observed in tumors compared to healthy tissues. Interestingly, a significant decrease in rRNA 2'Ome levels was identified for one particular site in a second series of 10 malignant mammary tumors and 8 benign tumors (i.e., 18S-Am576) [71]. Thus, this observation supports the notion that some alterations of the rRNA 2'Ome level can correlate with the pathological degree of the sample. Moreover, as observed in DLBCL, the analysis of the 195 breast tumors revealed that rRNA 2'Ome, at some sites, are naturally different from one patient to another, suggesting a tumor-specific variability of rRNA 2'Ome [70]. More importantly, it appears that, while 60% of the rRNA sites displayed a very small variation in their 2'Ome level between the 195 breast cancer patients reflecting the full heterogeneity of breast tumors, about 40% of the rRNA sites exhibited a

high variation [71]. The “stable” rRNA 2’Ome sites are conserved throughout evolution and are concentrated directly within main functional domains of the ribosome. In contrast, the “variable” rRNA 2’Ome sites appear during evolution and localize both at the second layer of the main ribosomal functional domains and at positions of direct interaction with ribosomal proteins or core translational factors. This demonstration of the co-existence of compulsory and plastic 2’Ome at specific rRNA positions reconciles two divergent visions of rRNA 2’Ome. First, it unambiguously challenges the long-lasting notion that the presence of rRNA chemical modifications is strictly required for maintaining a proper ribosome structure, which is necessary to support the bona fide ribosome’s intrinsic activity and therefore cell fitness [6,10,11,73]. In contrast, this finding supports the emerging notion that rRNA chemical modifications may provide plasticity to the ribosome structure and therefore provide the ribosome with regulatory abilities [44,52,63].

Finally, the study by Marcel and colleagues reported that differences in rRNA 2’Ome at the “variable” sites was also associated with biological and clinical features [71]. For example, a significant decrease in rRNA 2’Ome levels was observed in triple-negative breast cancers compared to luminal breast cancers at 18S-Gm1447, while a significant increase was observed at 18S-Am576, therefore supporting specific alterations of rRNA 2’Ome associated with the biological and clinical context. In addition, based on rRNA 2’Ome profiles at 106 positions, four groups of breast tumors, which displayed different overall survival and grade repartition and intrinsic breast cancer subtypes, have been identified. Notably, the breast tumor group that displayed a global decrease in rRNA 2’Ome compared to others exhibited the poorest overall survival and the highest tumor grade, usually associated with a hyperproliferative state. Therefore, in the same manner as the transcriptomic approaches, the rRNA 2’Ome signature may recapitulate several oncogenic and biological traits of tumors. To this regard, we may expect to not only identify a high diversity of rRNA 2’Ome profiles in cancer and, more largely, in human diseases, but also to be able to use rRNA 2’Ome profiling as a direct readout of the inter-tumoral heterogeneity. Further studies will have to address the challenging question of whether the intra-tumoral variability of rRNA 2’Ome reflects either an intra-tumoral cell heterogeneity or a cell-to-cell variability in rRNA modifications.

Overall, this first profiling of rRNA 2’Ome in tumor samples demonstrated the high diversity of rRNA 2’Ome in humans and the substantiate occurrence of natural variation of rRNA 2’Ome in real-life. More importantly, these studies emphasize the notion that variation in rRNA 2’Ome, and even more so at particular rRNA 2’Ome sites, is not random but rather correlates with something biological and clinical, therefore unraveling the potential role of the rRNA 2’Ome in biology.

Table 2. The twelve most referenced human rRNA 2’Ome sites showing variability.

Site ¹	Biological Context of Variability	Variation	Related snoRNA (Host Gene)	Method	Reference
18S-Um116	FBL depletion	Down	SNORD42A/B (RPL23A)	RMS ² -Illumina	[44]
	SNORD42A depletion	Down		RMS-Illumina + RTL-P	[54]
	EZH2 depletion	Down		RMS-Illumina	[50]
	Cell lines, between breast cancer patients	Up and down		RMS-Illumina	[71]
18S-Cm174	FMRP depletion	Up	SNORD45C (RABGGTB)	RMS-Illumina	[51]
	FBL depletion	Down		RMS-Illumina	[44]
	c-MYC overexpression	Up		RMS-Ion Torrent	[59]
	EZH2 depletion	Down		RMS-Illumina	[50]
	Cell lines	Up and down		RMS-Ion Torrent	[26]
18S-Gm867	FMRP depletion	Down	SNORD98 (CCAR1)	RMS-Illumina	[51]
	EZH2 depletion	Down		RMS-Illumina	[50]
	FBL depletion	Down		RMS-Illumina	[44]
	Cell lines, between breast cancer patients	Up and down		RMS-Illumina	[71]

Table 2. Cont.

Site ¹	Biological Context of Variability	Variation	Related snoRNA (Host Gene)	Method	Reference
28S-Cm1327	EZH2 depletion	Down	SNORD104	RMS-Illumina	[50]
	NPM1 depletion and loss-of-function, between AML patients	Down	(lncSNHG25)	RTL-P	[52]
	FBL depletion	Down		RMS-Illumina	[44]
	Cell lines, between breast cancer patients	Up and down		RMS-Illumina	[71]
28S-Am1858	Reduced Arl2 expression	Up	SNORD38A/B (RPS8)	RTL-P	[64]
	FBL depletion	Down		RMS-Illumina	[44]
	EZH2 depletion	Down		RMS-Illumina	[50]
	DLBCL patients vs. healthy donor	Down		RMS-Ion Torrent	[70]
	β-thalassemic patients vs. healthy donor	Up		RTL-P	[69]
28S-Um2402	FMRP depletion	Down	SNORD143, SNORD144 (SEC31A)	RMS-Illumina	[51]
	FBL depletion	Down		RMS-Illumina	[44]
	c-MYC overexpression cell lines	Down		RMS-Ion Torrent	[59]
	Cell lines, between breast cancer patients	Up and down		RMS-Ion Torrent	[26]
	DLBCL patients vs. healthy donor	Up and down		RMS-Illumina	[71]
28S-Cm2409	EZH2 depletion	Down	SNORD5 (TAF1D)	RMS-Illumina	[50]
	FMRP depletion	Up		RMS-Illumina	[51]
	FBL depletion	Down		RMS-Illumina	[44]
	cell lines	Up and down		RMS-Ion Torrent	[26]
	cell lines, between breast cancer patients	Up and down		RMS-Illumina	[71]
28S-Am2774	FMRP depletion	Down	SNORD99 (SNHG12)	RMS-Illumina	[51]
	FBL depletion	Down		RMS-Illumina	[44]
	Cell lines, between breast cancer patients	Up and down		RMS-Illumina	[71]
	Brain, liver and skeletal muscle	Down in Liver		RMS-Ion Torrent	[70]
28S-Cm2848	EZH2 depletion	Down	SNORD50A/B (SNHG5)	RMS-Illumina	[50]
	FBL depletion	Down		RMS-Illumina	[44]
	Cell lines	Up and down		RMS-Ion Torrent	[26]
	Cell lines, between breast cancer patients	Up and down		RMS-Illumina	[71]
28S-Gm3923	FMRP depletion	Down	SNORD111/B (SF3B3)	RMS-Illumina	[51]
	FBL depletion	Down		RMS-Illumina	[44]
	Cell lines	Up and down		RMS-Ion Torrent	[26]
	cell lines, between breast cancer patients	Up and down		RMS-Illumina	[71]
28S-Cm4506	FBL depletion	Down	SNORD35A (RPL13A)	RMS-Illumina	[44]
	SNORA23 depletion/overexpression	Up/down	SNORD35B (RPS11)	RTL-P	[48]
	AES depletion	Down		RTL-P	[55]
	β-thalassemic patients vs. healthy donor	Up		RTL-P	[69]
28S-Gm4593	FBL depletion	Down	SNORD78 (GAS5)	RMS-Illumina	[44]
	ZFAS1 depletion/overexpression	Up/down		RTL-P/DBPST	[47]
	Cell lines, between breast cancer patients	Up and down		RMS-Illumina	[71]
	DLBCL patients vs. healthy donor, brain, liver and skeletal muscle	Up in DLBCL, Up in brain and liver		RMS-Ion Torrent	[70]

¹ Based on Biotoul database; ² RMS = RiboMeth-Seq.

6. In Vivo Variability of rRNA 2'Ome in Model Organisms

In addition to revealing rRNA 2'Ome diversity in yeast, human cell lines, and samples, the development of RNA-seq-based approaches brought a significant insight into the mapping of rRNA 2'Ome sites at the entire organism/organ level in different plant and animal models, therefore enlarging the number of models available to elucidate the role of rRNA 2'Ome variability in living organisms (Figure 2).

6.1. *Trypanosoma Brucei*

In contrast to multicellular organisms, which have differentiated and specialized cells to perform a variety of functions and adapt to environmental conditions, *Trypanosoma brucei* is a unicellular eukaryotic parasite that cycles between two hosts (fly vs. affected animals including humans) and shows extremely coordinated adaptations to their surrounding environment [74]. Recently, a map of 99 highly confident 2'Ome sites was established

on *T. brucei* rRNA using the combination of three different high-throughput technologies (RibOxi-seq, RiboMeth-seq and 2'-OMe-seq) [75]. These sites identified at the two life stages of *T. brucei*, including 31 2'OMe sites in the small subunit rRNA, 34 in the large subunit α , and 38 in the large subunit β .

In addition to the mapping of rRNA 2'OMe, the quantitative data generated by RiboMeth-seq allowed researchers to evaluate the intra- and inter-variability in *T. brucei*. First, this study showed that not all sites are equally methylated, but a subset of rRNA 2'OMe are partially methylated, as observed in cell lines and human samples. Thus, within a single sample, a mix of rRNA molecules methylated and non-methylated co-exist. Second, it appears that some sites display different levels of 2'OMe when comparing the two life cycles of *T. brucei*. Interestingly, most of these modulated 2'OMe sites are located around the functional domains of the ribosome (A, P, and peptidyl-transferase center sites). As in many cellular models and human samples, no correlation between the level of 2'OMe and SNORDs expression was established. Therefore, these results support the natural existence of different populations of ribosomes differentially methylated, that might support specific functions during *T. brucei* life cycles and contribute to the slight but recurrent adaptation of this living organism to differences in nutrients and temperature between the two hosts.

6.2. *Arabidopsis Thaliana*

Arabidopsis thaliana, the first plant to have its genome sequenced [76], is a widely used model organism in biochemistry, genetic development, and physiology. However, a complete cartography of rRNA 2'OMe sites was missing until recently. Up to now, the identification of rRNA 2'OMe sites in *A. thaliana* has mainly consisted in predictions through sequence homology with SNORDs and all predicted sites could not be validated. Recently, two studies employed RiboMeth-seq for mapping rRNA 2'OMe sites in *A. thaliana*, thus providing a more confident cartography [77,78]. Hence, both studies identified 117 and 111 2'OMe sites, the difference potentially being due to either distinct statistical methods to identify new candidates or distinct biological materials, i.e., 3-week-old leaves plant seedlings [78] vs. ~9 days light-grown seedlings [77]. Nevertheless, 110 2'OMe sites were found in common (2 sites in 5.8S, 35 in 18S, and 73 in 28S rRNAs), indicating that we are close to a complete mapping of methylated sites in *A. thaliana*.

Finally, these two studies also showed that the modulation of ribosome biogenesis factors affects rRNA 2'OMe in plants, as observed in cell lines [77,78]. Thus, the expression of Nucleolin (transcription and processing of pre-rRNA) and Fibrillarin mutants induced a general decrease in rRNA 2'OMe levels compared to wild-type plants. Moreover, it has to be noted that, in wild-type plants, only a subset of sites is partially methylated, indicating the existence of intra-variability in rRNA 2'OMe as well as inter-variability, although experimentally-driven, in *A. thaliana*.

6.3. *Danio Rerio*

Danio rerio (zebrafish) is another widely used model organism in biomedical research and developmental biology. Recent studies have shown that zebrafish expressed two distinct versions of 5S, 5.8S, 18S, and 28S rRNAs during early embryogenesis [79–81], where oocyte-specific rRNAs (i.e., maternal rRNAs) switch to somatic rRNA variants, the nucleotide sequences of which are different from maternal rRNAs. It has been hypothesized that these differences in rRNA sequences may have major functional consequences by affecting the ribosome:mRNA interactions and leading to specific translations of either maternal or somatic mRNAs. Interestingly, this shift from maternal to somatic rRNA molecules is also observable at the rRNA 2'OMe level [82].

First, a complete map of 98 2'OMe sites was established by compiling data issued from (i) RiboMeth-seq, (ii) primer extension assays and/or (iii) putative SNORD expression analyses, in pooled samples issued from four developmental stages (unfertilized eggs, the 32-cell, 12-somite, protruding mouth and tail) in 1-year old (adult) fish. Three 2'OMe sites were identified in 5.8S, 35 in 18S, and 60 in 28S rRNAs. Second, it appears that,

although 97 sites were common between the two maternal and somatic rRNA types, one site was somatic-specific (28S-Cm3916). Moreover, a comparison of rRNAs issued from the maternal and somatic stages and adults revealed an increased methylation level of a subset of 2'Ome sites. In adult zebrafish tissues, the majority of 2'Ome sites seemed almost fully methylated and only few sites were partially methylated. As in other model organisms, natural variations of 2'Ome occurs in zebrafish, within particular developmental stages and adult tissues, therefore questioning the role of rRNA 2'Ome in development and tissue homeostasis.

6.4. *Mus Musculus*

Mouse remains an important model organism belonging the class of Mammalia and the super-order of Euarchontoglires, which contains both Rodentia and Primate. Therefore, mice share many developmental, biological, and genomic features with primates which makes it a perfect model for studying human biology and diseases, and for developing therapeutic strategies. Therefore, the perspective of studying the regulation and the function of rRNA 2'Ome and its potential use as a therapeutic target in humans required a deep understanding of the 2'Ome in mice. Despite the lack of 2'Ome studies in mice using biochemical approaches, such as mass spectrometry or primer extension assays, recent works have started the mapping of mouse rRNA 2'Ome through high-throughput methods. In contrast to methods such as RiboMeth-seq, which relies on the assumption that a site is modified to determine a methylation score, and therefore requires validation with gold-standard methods, the 2OME-seq developed by Incarnato et al. has recently allowed researchers to uncover potential bona fide 2'Ome sites in mice without needing prior extensive biochemical studies [31]. Interestingly, comparisons of 2'Ome sites between human HeLa cells and mouse Embryonic Stem Cells (mESCs) indicated that 2'Ome sites are extremely conserved in mice and humans (92 out of 93 in common). Interestingly, 2OME-seq identified 12 de novo 2'Ome sites in humans, 11 of which are conserved in mouse. Surprisingly, 28S-Gm4196, which is one of the most conserved 2'Ome site from *E. coli* to humans [24] was not found in mice in this study, in contrast to others (Durand S and Marcel V, unpublished data; [83]). Finally, although 2OME-seq is not the most robust method to quantify 2'Ome levels, these data suggest nonetheless that the level of 2'Ome on specific sites is different between human and mouse. These species-specific 2'Ome cartographies suggest (i) that this diversity in location of 2'Ome sites in mouse ESCs is distinct from that of human HeLa cells; and (ii) that mouse ESCs may possess ribosomes with compositional heterogeneity at the rRNA 2'Ome level.

Another milestone has been recently reached by the team of Cavallé, in respect to the understanding of the 2'Ome biological diversity in mice [83]. By performing RiboMeth-seq analysis on adult and E16.6/E9.5 embryo tissues, they uncovered a fascinating dynamic of 2'Ome during mouse development. First, albeit mouse adult tissues display close-to-full methylation rates on the majority of rRNA sites, few positions exhibit partially methylated rates among the same tissues (intra-variability), while some other sites present differential 2'Ome levels when comparing distinct tissues (inter-variability). Thus, adult tissues present a low, albeit existing, degree of rRNA 2'Ome intra-variability, which is relatively conserved in all tested adult tissues (i.e., brain, liver, lung, heart, and kidneys). In contrast, many 2'Ome sites are considerably less methylated in developing tissues compared to their adult counterparts, therefore emphasizing a stunning developmental regulation of 2'Ome during mouse development. Notably, two 2'Ome sites, the 18S-Am576 and 28S-Gm4593, exhibited an opposite regulation, the latter one being associated with a reduction in 2'Ome level in adult tissues, most likely due to a developmental regulation of its related-SNORD78. Altogether, this work demonstrates the existence of a regulated 2'Ome pattern in adult mice and embryo tissues and sheds the light on the rRNA-related, developmentally-regulated ribosomal heterogeneity in mice.

7. Conclusions and Perspectives

The development of omic approaches (Table 1) dedicated to profiling rRNA 2'Ome in cell lines, fresh tissues, and human biopsies during the last 5 years has definitively demonstrated that rRNA 2'Ome varies in real-life depending on the physio-pathological context (Figure 2).

Intriguingly, not all methylated rRNA positions can tolerate a lack of 2'Ome. Indeed, it appears that only about one-third of methylated rRNA positions co-exists as a mix of methylated and non-methylated nucleotides. Due to the high diversity of technical, methodological, and statistical approaches used for profiling rRNA 2'Ome, it nevertheless remains difficult to firmly establish a genuine list of stable and variable sites, and further meta-analyses will therefore be required (Table 2). Similarly, we cannot rule out that unknown rRNA positions could be methylated in particular biological contexts. In addition, such techniques could not determine whether variations of rRNA 2'Ome reflect an intracellular 2'Ome heterogeneity (i.e., a mix of differentially methylated rRNA molecules within a single cell), or, in contrast, reveal a cell-homogenous 2'Ome that fluctuates between cell sub-populations. Finally, further studies should focus on establishing whether 2'Ome at variable positions co-exist or not on the same rRNA molecules or if they are, in contrast, mutually exclusive, therefore complexifying even further the degree of rRNA 2'Ome heterogeneity. As a matter of fact, the recent demonstration that single-molecule sequencing using nanopore technology can be used to map rRNA chemical modifications opens up novel perspectives in the field [84].

Altogether, these data clearly strengthen the hypothesis that rRNA 2'Ome may contribute to the establishment and/or maintenance of particular phenotypes, probably through a finely tuned control of translation, including in cancer. However, we cannot totally rule out that some variations of 2'Ome in vivo rather reflect a technical issue or biological noise and do not possess any functional roles. Few studies have ever addressed this question and many further in-depth studies are required to fully understand the role of rRNA 2'Ome in regulating particular phenotypes. One could hypothesize that, similar to observations made in *S. cerevisiae*, variable 2'O methylated positions may not be strictly essential for sustaining the bulk of cellular functions at a steady state, but could alternatively play a subtle role in particular conditions, such as, for instance, by adapting to rapid changes in environmental cues or responses to stress. Nevertheless, thanks to the recent advance in technologies dedicated to the rRNA epitranscriptomics, we now can firmly claim that rRNA 2'Ome varies naturally in real-life.

Author Contributions: Conceptualization, V.M. and S.D.; writing—original draft preparation, M.J., H.P., V.M. and S.D.; intellectual contribution, M.J., H.P., J.-J.D., M.G., V.M. and S.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by ITMO Cancer, grant number 20CN063-00 (REMOTE); by INCa, grant number 2019-138 (MARACAS); and by Ligue Nationale Contre le Cancer, grant number PPE-2021 (sMETHY). H.P. was PhD fellow from Ligue Nationale Contre le Cancer.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Natchiar, S.K.; Myasnikov, A.G.; Hazemann, I.; Klaholz, B.P. Visualizing the Role of 2'-OH rRNA Methylations in the Human Ribosome Structure. *Biomolecules* **2018**, *8*, 125. [[CrossRef](#)]
2. Melnikov, S.; Ben-Shem, A.; Garreau de Loubresse, N.; Jenner, L.; Yusupova, G.; Yusupov, M. One core, two shells: Bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.* **2012**, *19*, 560–567. [[CrossRef](#)]
3. Khatler, H.; Myasnikov, A.G.; Natchiar, S.K.; Klaholz, B.P. Structure of the human 80S ribosome. *Nature* **2015**, *520*, 640–645. [[CrossRef](#)]
4. Anger, A.M.; Armache, J.P.; Berninghausen, O.; Habeck, M.; Subklewe, M.; Wilson, D.N.; Beckmann, R. Structures of the human and *Drosophila* 80S ribosome. *Nature* **2013**, *497*, 80–85. [[CrossRef](#)]

5. Natchiar, S.K.; Myasnikov, A.G.; Kratzat, H.; Hazemann, I.; Klaholz, B.P. Visualization of chemical modifications in the human 80S ribosome structure. *Nature* **2017**, *551*, 472–477. [[CrossRef](#)]
6. Decatur, W.A.; Fournier, M.J. rRNA modifications and ribosome function. *Trends Biochem. Sci.* **2002**, *27*, 344–351. [[CrossRef](#)]
7. Monaco, P.L.; Marcel, V.; Diaz, J.J.; Catez, F. 2'-O-Methylation of Ribosomal RNA: Towards an Epitranscriptomic Control of Translation? *Biomolecules* **2018**, *8*, 106. [[CrossRef](#)] [[PubMed](#)]
8. Dimitrova, D.G.; Teyssset, L.; Carre, C. RNA 2'-O-Methylation (Nm) Modification in Human Diseases. *Genes* **2019**, *10*, 117. [[CrossRef](#)] [[PubMed](#)]
9. Maden, B.E. Methylation map of *Xenopus laevis* ribosomal RNA. *Nature* **1980**, *288*, 293–296. [[CrossRef](#)] [[PubMed](#)]
10. Schimmang, T.; Tollervey, D.; Kern, H.; Frank, R.; Hurt, E.C. A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability. *EMBO J.* **1989**, *8*, 4015–4024. [[CrossRef](#)] [[PubMed](#)]
11. Newton, K.; Petfalski, E.; Tollervey, D.; Caceres, J.F. Fibrillarin is essential for early development and required for accumulation of an intron-encoded small nucleolar RNA in the mouse. *Mol. Cell. Biol.* **2003**, *23*, 8519–8527. [[CrossRef](#)]
12. Dalla Venezia, N.; Vincent, A.; Marcel, V.; Catez, F.; Diaz, J.J. Emerging Role of Eukaryote Ribosomes in Translational Control. *Int. J. Mol. Sci.* **2019**, *20*, 1226. [[CrossRef](#)]
13. Gabut, M.; Bourdelais, F.; Durand, S. Ribosome and Translational Control in Stem Cells. *Cells* **2020**, *9*, 497. [[CrossRef](#)]
14. Genuth, N.R.; Barna, M. The Discovery of Ribosome Heterogeneity and Its Implications for Gene Regulation and Organismal Life. *Mol. Cell* **2018**, *71*, 364–374. [[CrossRef](#)] [[PubMed](#)]
15. Sharma, S.; Lafontaine, D.L.J. 'View From A Bridge': A New Perspective on Eukaryotic rRNA Base Modification. *Trends Biochem. Sci.* **2015**, *40*, 560–575. [[CrossRef](#)]
16. Xue, S.; Barna, M. Specialized ribosomes: A new frontier in gene regulation and organismal biology. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 355–369. [[CrossRef](#)]
17. Birkedal, U.; Christensen-Dalsgaard, M.; Krogh, N.; Sabarinathan, R.; Gorodkin, J.; Nielsen, H. Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew. Chem.* **2015**, *127*, 461–465. [[CrossRef](#)]
18. Kos, M.; Tollervey, D. Yeast pre-rRNA processing and modification occur cotranscriptionally. *Mol. Cell* **2010**, *37*, 809–820. [[CrossRef](#)]
19. Gautier, T.; Berges, T.; Tollervey, D.; Hurt, E. Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Mol. Cell. Biol.* **1997**, *17*, 7088–7098. [[CrossRef](#)]
20. Lafontaine, D.L. Noncoding RNAs in eukaryotic ribosome biogenesis and function. *Nat. Struct. Mol. Biol.* **2015**, *22*, 11–19. [[CrossRef](#)] [[PubMed](#)]
21. Motorin, Y.; Helm, M. RNA nucleotide methylation. *Wiley Interdiscip. Rev. RNA* **2011**, *2*, 611–631. [[CrossRef](#)]
22. Bonnerot, C.; Pintard, L.; Lutfalla, G. Functional redundancy of Spb1p and a snR52-dependent mechanism for the 2'-O-ribose methylation of a conserved rRNA position in yeast. *Mol. Cell* **2003**, *12*, 1309–1315. [[CrossRef](#)]
23. Lapeyre, B.; Purushothaman, S.K. Spb1p-directed formation of Gm2922 in the ribosome catalytic center occurs at a late processing stage. *Mol. Cell* **2004**, *16*, 663–669. [[CrossRef](#)]
24. Ayadi, L.; Galvanin, A.; Pichot, F.; Marchand, V.; Motorin, Y. RNA ribose methylation (2'-O-methylation): Occurrence, biosynthesis and biological functions. *Biochim. Biophys. Acta Gene Regul. Mech.* **2019**, *1862*, 253–269. [[CrossRef](#)]
25. Helm, M.; Motorin, Y. Detecting RNA modifications in the epitranscriptome: Predict and validate. *Nat. Rev. Genet.* **2017**, *18*, 275–291. [[CrossRef](#)]
26. Krogh, N.; Jansson, M.D.; Hafner, S.J.; Tehler, D.; Birkedal, U.; Christensen-Dalsgaard, M.; Lund, A.H.; Nielsen, H. Profiling of 2'-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity. *Nucleic Acids Res.* **2016**, *44*, 7884–7895. [[CrossRef](#)]
27. Marchand, V.; Blanloeil-Oillo, F.; Helm, M.; Motorin, Y. Illumina-based RiboMethSeq approach for mapping of 2'-O-Me residues in RNA. *Nucleic Acids Res.* **2016**, *44*, e135. [[CrossRef](#)] [[PubMed](#)]
28. Maden, B.E. Mapping 2'-O-methyl groups in ribosomal RNA. *Methods* **2001**, *25*, 374–382. [[CrossRef](#)] [[PubMed](#)]
29. Dong, Z.W.; Shao, P.; Diao, L.T.; Zhou, H.; Yu, C.H.; Qu, L.H. RTL-P: A sensitive approach for detecting sites of 2'-O-methylation in RNA molecules. *Nucleic Acids Res.* **2012**, *40*, e157. [[CrossRef](#)]
30. Taoka, M.; Nobe, Y.; Hori, M.; Takeuchi, A.; Masaki, S.; Yamauchi, Y.; Nakayama, H.; Takahashi, N.; Isobe, T. A mass spectrometry-based method for comprehensive quantitative determination of post-transcriptional RNA modifications: The complete chemical structure of *Schizosaccharomyces pombe* ribosomal RNAs. *Nucleic Acids Res.* **2015**, *43*, e115. [[CrossRef](#)] [[PubMed](#)]
31. Incarnato, D.; Anselmi, F.; Morandi, E.; Neri, F.; Maldotti, M.; Rapelli, S.; Parlato, C.; Basile, G.; Oliviero, S. High-throughput single-base resolution mapping of RNA 2-O-methylated residues. *Nucleic Acids Res.* **2017**, *45*, 1433–1441. [[CrossRef](#)]
32. Zhu, Y.; Pirnie, S.P.; Carmichael, G.G. High-throughput and site-specific identification of 2'-O-methylation sites using ribose oxidation sequencing (RibOxi-seq). *RNA* **2017**, *23*, 1303–1314. [[CrossRef](#)]
33. Taoka, M.; Nobe, Y.; Yamaki, Y.; Yamauchi, Y.; Ishikawa, H.; Takahashi, N.; Nakayama, H.; Isobe, T. The complete chemical structure of *Saccharomyces cerevisiae* rRNA: Partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9. *Nucleic Acids Res.* **2016**, *44*, 8951–8961. [[CrossRef](#)]
34. Yang, J.; Sharma, S.; Watzinger, P.; Hartmann, J.D.; Kotter, P.; Entian, K.D. Mapping of Complete Set of Ribose and Base Modifications of Yeast rRNA by RP-HPLC and Mung Bean Nuclease Assay. *PLoS ONE* **2016**, *11*, e0168873. [[CrossRef](#)] [[PubMed](#)]

35. Tollervey, D.; Lehtonen, H.; Jansen, R.; Kern, H.; Hurt, E.C. Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell* **1993**, *72*, 443–457. [[CrossRef](#)]
36. Tollervey, D.; Lehtonen, H.; Carmo-Fonseca, M.; Hurt, E.C. The small nucleolar RNP protein NOP1 (fibrillarin) is required for pre-rRNA processing in yeast. *EMBO J.* **1991**, *10*, 573–583. [[CrossRef](#)]
37. Liang, X.H.; Liu, Q.; Fournier, M.J. rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. *Mol. Cell* **2007**, *28*, 965–977. [[CrossRef](#)] [[PubMed](#)]
38. Liang, X.H.; Liu, Q.; Fournier, M.J. Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA* **2009**, *15*, 1716–1728. [[CrossRef](#)]
39. Esguerra, J.; Warringer, J.; Blomberg, A. Functional importance of individual rRNA 2'-O-ribose methylations revealed by high-resolution phenotyping. *RNA* **2008**, *14*, 649–656. [[CrossRef](#)]
40. Buchhaupt, M.; Sharma, S.; Kellner, S.; Oswald, S.; Paetzold, M.; Peifer, C.; Watzinger, P.; Schrader, J.; Helm, M.; Entian, K.D. Partial methylation at Am100 in 18S rRNA of baker's yeast reveals ribosome heterogeneity on the level of eukaryotic rRNA modification. *PLoS ONE* **2014**, *9*, e89640. [[CrossRef](#)] [[PubMed](#)]
41. Martin, R.; Straub, A.U.; Doebele, C.; Bohnsack, M.T. DEXD/H-box RNA helicases in ribosome biogenesis. *RNA Biol.* **2013**, *10*, 4–18. [[CrossRef](#)] [[PubMed](#)]
42. Rodriguez-Galan, O.; Garcia-Gomez, J.J.; de la Cruz, J. Yeast and human RNA helicases involved in ribosome biogenesis: Current status and perspectives. *Biochim. Biophys. Acta* **2013**, *1829*, 775–790. [[CrossRef](#)]
43. Aquino, G.R.R.; Krogh, N.; Hackert, P.; Martin, R.; Galesio, J.D.; van Nues, R.W.; Schneider, C.; Watkins, N.J.; Nielsen, H.; Bohnsack, K.E.; et al. RNA helicase-mediated regulation of snoRNP dynamics on pre-ribosomes and rRNA 2'-O-methylation. *Nucleic Acids Res.* **2021**, *49*, 4066–4084. [[CrossRef](#)] [[PubMed](#)]
44. Erales, J.; Marchand, V.; Panthu, B.; Gillot, S.; Belin, S.; Ghayad, S.E.; Garcia, M.; Laforets, F.; Marcel, V.; Baudin-Baillieu, A.; et al. Evidence for rRNA 2'-O-methylation plasticity: Control of intrinsic translational capabilities of human ribosomes. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 12934–12939. [[CrossRef](#)] [[PubMed](#)]
45. Gumienny, R.; Jedlinski, D.J.; Schmidt, A.; Gypas, F.; Martin, G.; Vina-Vilaseca, A.; Zavolan, M. High-throughput identification of C/D box snoRNA targets with CLIP and RiboMeth-seq. *Nucleic Acids Res.* **2017**, *45*, 2341–2353. [[CrossRef](#)] [[PubMed](#)]
46. Taoka, M.; Nobe, Y.; Yamaki, Y.; Sato, K.; Ishikawa, H.; Izumikawa, K.; Yamauchi, Y.; Hirota, K.; Nakayama, H.; Takahashi, N.; et al. Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.* **2018**, *46*, 9289–9298. [[CrossRef](#)]
47. Wu, H.; Qin, W.; Lu, S.; Wang, X.; Zhang, J.; Sun, T.; Hu, X.; Li, Y.; Chen, Q.; Wang, Y.; et al. Long noncoding RNA ZFAS1 promoting small nucleolar RNA-mediated 2'-O-methylation via NOP58 recruitment in colorectal cancer. *Mol. Cancer* **2020**, *19*, 95. [[CrossRef](#)]
48. Liu, Z.; Pang, Y.; Jia, Y.; Qin, Q.; Wang, R.; Li, W.; Jing, J.; Liu, H.; Liu, S. SNORA23 inhibits HCC tumorigenesis by impairing the 2'-O-ribose methylation level of 28S rRNA. *Cancer Biol. Med.* **2021**. online ahead of print. [[CrossRef](#)]
49. Sharma, S.; Marchand, V.; Motorin, Y.; Lafontaine, D.L.J. Identification of sites of 2'-O-methylation vulnerability in human ribosomal RNAs by systematic mapping. *Sci. Rep.* **2017**, *7*, 11490. [[CrossRef](#)]
50. Yi, Y.; Li, Y.; Meng, Q.; Li, Q.; Li, F.; Lu, B.; Shen, J.; Fazli, L.; Zhao, D.; Li, C.; et al. A PRC2-independent function for EZH2 in regulating rRNA 2'-O methylation and IRES-dependent translation. *Nat. Cell Biol.* **2021**, *23*, 341–354. [[CrossRef](#)]
51. D'Souza, M.N.; Gowda, N.K.C.; Tiwari, V.; Babu, R.O.; Anand, P.; Dastidar, S.G.; Singh, R.; James, O.G.; Selvaraj, B.; Pal, R.; et al. FMRP Interacts with C/D Box snoRNA in the Nucleus and Regulates Ribosomal RNA Methylation. *iScience* **2018**, *9*, 399–411. [[CrossRef](#)] [[PubMed](#)]
52. Nachmani, D.; Bothmer, A.H.; Grisendi, S.; Mele, A.; Bothmer, D.; Lee, J.D.; Monteleone, E.; Cheng, K.; Zhang, Y.; Bester, A.C.; et al. Germline NPM1 mutations lead to altered rRNA 2'-O-methylation and cause dyskeratosis congenita. *Nat. Genet.* **2019**, *51*, 1518–1529. [[CrossRef](#)] [[PubMed](#)]
53. Kunchala, P.; Kuravi, S.; Jensen, R.; McGuirk, J.; Balusu, R. When the good go bad: Mutant NPM1 in acute myeloid leukemia. *Blood Rev.* **2018**, *32*, 167–183. [[CrossRef](#)]
54. Pauli, C.; Liu, Y.; Rohde, C.; Cui, C.; Fijalkowska, D.; Gerloff, D.; Walter, C.; Krijgsveld, J.; Dugas, M.; Edemir, B.; et al. Site-specific methylation of 18S ribosomal RNA by SNORD42A is required for acute myeloid leukemia cell proliferation. *Blood* **2020**, *135*, 2059–2070. [[CrossRef](#)] [[PubMed](#)]
55. Zhou, F.; Liu, Y.; Rohde, C.; Pauli, C.; Gerloff, D.; Kohn, M.; Misiak, D.; Baumer, N.; Cui, C.; Gollner, S.; et al. AML1-ETO requires enhanced C/D box snoRNA/RNP formation to induce self-renewal and leukaemia. *Nat. Cell Biol.* **2017**, *19*, 844–855. [[CrossRef](#)] [[PubMed](#)]
56. Mulloy, J.C.; Cammenga, J.; MacKenzie, K.L.; Berguido, F.J.; Moore, M.A.; Nimer, S.D. The AML1-ETO fusion protein promotes the expansion of human hematopoietic stem cells. *Blood* **2002**, *99*, 15–23. [[CrossRef](#)]
57. Steffen, B.; Knop, M.; Bergholz, U.; Vakhrusheva, O.; Rode, M.; Kohler, G.; Henrichs, M.P.; Bulk, E.; Hehn, S.; Stehling, M.; et al. AML1/ETO induces self-renewal in hematopoietic progenitor cells via the Groucho-related amino-terminal AES protein. *Blood* **2011**, *117*, 4328–4337. [[CrossRef](#)]
58. Tonks, A.; Tonks, A.J.; Pearn, L.; Pearce, L.; Hoy, T.; Couzens, S.; Fisher, J.; Burnett, A.K.; Darley, R.L. Expression of AML1-ETO in human myelomonocytic cells selectively inhibits granulocytic differentiation and promotes their self-renewal. *Leukemia* **2004**, *18*, 1238–1245. [[CrossRef](#)]

59. Jansson, M.D.; Häfner, S.J.; Altinel, K.; Tehler, D.; Krogh, N.; Jakobsen, E.; Andersen, J.V.; Andersen, K.L.; Schoof, E.M.; Ménard, P.; et al. Ribosomal RNA methylation induced by MYC regulates translation. *Res. Sq.* **2020**. preprint. [[CrossRef](#)]
60. Joerger, A.C.; Fersht, A.R. The p53 Pathway: Origins, Inactivation in Cancer, and Emerging Therapeutic Approaches. *Annu. Rev. Biochem.* **2016**, *85*, 375–404. [[CrossRef](#)] [[PubMed](#)]
61. Marcel, V.; Dichtel-Danjoy, M.L.; Sagne, C.; Hafsi, H.; Ma, D.; Ortiz-Cuaran, S.; Olivier, M.; Hall, J.; Mollereau, B.; Hainaut, P.; et al. Biological functions of p53 isoforms through evolution: Lessons from animal and cellular models. *Cell Death Differ.* **2011**, *18*, 1815–1824. [[CrossRef](#)] [[PubMed](#)]
62. Zhai, W.; Comai, L. Repression of RNA polymerase I transcription by the tumor suppressor p53. *Mol. Cell Biol.* **2000**, *20*, 5930–5938. [[CrossRef](#)]
63. Marcel, V.; Ghayad, S.E.; Belin, S.; Therizols, G.; Morel, A.P.; Solano-Gonzalez, E.; Vendrell, J.A.; Hacot, S.; Mertani, H.C.; Albaret, M.A.; et al. p53 acts as a safeguard of translational control by regulating fibrillarin and rRNA methylation in cancer. *Cancer Cell* **2013**, *24*, 318–330. [[CrossRef](#)]
64. Belin, S.; Beghin, A.; Solano-Gonzalez, E.; Bezin, L.; Brunet-Manquat, S.; Textoris, J.; Prats, A.C.; Mertani, H.C.; Dumontet, C.; Diaz, J.J. Dysregulation of ribosome biogenesis and translational capacity is associated with tumor progression of human breast cancer cells. *PLoS ONE* **2009**, *4*, e7147. [[CrossRef](#)]
65. Metge, B.J.; Kammerud, S.C.; Pruitt, H.C.; Shevde, L.A.; Samant, R.S. Hypoxia re-programs 2'-O-Me modifications on ribosomal RNA. *iScience* **2021**, *24*, 102010. [[CrossRef](#)] [[PubMed](#)]
66. Burke, M.F.; McLaurin, D.M.; Logan, M.K.; Hebert, M.D. Alteration of 28S rRNA 2'-O-methylation by etoposide correlates with decreased SMN phosphorylation and reduced Drosha levels. *Biol. Open* **2019**, *8*, bio041848. [[CrossRef](#)]
67. Burke, M.F.; Logan, M.K.; Hebert, M.D. Identification of additional regulatory RNPs that impact rRNA and U6 snRNA methylation. *Biol. Open* **2018**, *7*, bio036095. [[CrossRef](#)] [[PubMed](#)]
68. Poole, A.R.; Vicino, I.; Adachi, H.; Yu, Y.T.; Hebert, M.D. Regulatory RNPs: A novel class of ribonucleoproteins that potentially contribute to ribosome heterogeneity. *Biol. Open* **2017**, *6*, 1342–1354. [[CrossRef](#)] [[PubMed](#)]
69. Sornjai, W.; Lithanatudom, P.; Eroles, J.; Joly, P.; Francina, A.; Hacot, S.; Fucharoen, S.; Svasti, S.; Diaz, J.J.; Mertani, H.C.; et al. Hypermethylation of 28S ribosomal RNA in beta-thalassemia trait carriers. *Int. J. Biol. Macromol.* **2017**, *94*, 728–734. [[CrossRef](#)] [[PubMed](#)]
70. Krogh, N.; Asmar, F.; Côme, C.; Munch-Petersen, H.F.; Grønbaek, K.; Nielsen, H. Profiling of ribose methylations in ribosomal RNA from diffuse large B-cell lymphoma patients for evaluation of ribosomes as drug targets. *NAR Cancer* **2020**, *2*, zcaa035. [[CrossRef](#)] [[PubMed](#)]
71. Marcel, V.; Kielbassa, J.; Marchand, V.; Kundhavi, S.N.; Paraqindes, H.; Van Long, F.N.; Ayadi, L.; Bourguignon-Igel, V.; Lo Monaco, P.; Monchiet, D.; et al. Ribosomal RNA 2'-O-methylation as a novel layer of inter-tumour heterogeneity in breast cancer. *NAR Cancer* **2020**, *2*, zcaa036. [[CrossRef](#)]
72. Truitt, M.L.; Ruggero, D. New frontiers in translational control of the cancer genome. *Nat. Rev. Cancer* **2016**, *16*, 288–304. [[CrossRef](#)] [[PubMed](#)]
73. Baudin-Baillieu, A.; Fabret, C.; Liang, X.H.; Piekna-Przybylska, D.; Fournier, M.J.; Rousset, J.P. Nucleotide modifications in three functionally important regions of the *Saccharomyces cerevisiae* ribosome affect translation accuracy. *Nucleic Acids Res.* **2009**, *37*, 7665–7677. [[CrossRef](#)]
74. Matthews, K.R. The developmental cell biology of *Trypanosoma brucei*. *J. Cell Sci.* **2005**, *118*, 283–290. [[CrossRef](#)] [[PubMed](#)]
75. Rajan, K.S.; Zhu, Y.; Adler, K.; Doniger, T.; Cohen-Chalamish, S.; Srivastava, A.; Shalev-Benami, M.; Matzov, D.; Unger, R.; Tschudi, C.; et al. The large repertoire of 2'-O-methylation guided by C/D snoRNAs on *Trypanosoma brucei* rRNA. *RNA Biol.* **2020**, *17*, 1018–1039. [[CrossRef](#)] [[PubMed](#)]
76. Arabidopsis Genome, I. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **2000**, *408*, 796–815. [[CrossRef](#)]
77. Wu, S.; Wang, Y.; Wang, J.; Li, X.; Li, J.; Ye, K. Profiling of RNA ribose methylation in *Arabidopsis thaliana*. *Nucleic Acids Res.* **2021**, *49*, 4104–4119. [[CrossRef](#)]
78. Azevedo-Favory, J.; Gaspin, C.; Ayadi, L.; Montacie, C.; Marchand, V.; Jobet, E.; Rompais, M.; Carapito, C.; Motorin, Y.; Saez-Vasquez, J. Mapping rRNA 2'-O-methylations and identification of C/D snoRNAs in *Arabidopsis thaliana* plants. *RNA Biol.* **2021**, 1–18. [[CrossRef](#)]
79. Locati, M.D.; Pagano, J.F.; Ensink, W.A.; van Olst, M.; van Leeuwen, S.; Nehrdich, U.; Zhu, K.; Spaink, H.P.; Girard, G.; Rauwerda, H.; et al. Linking maternal and somatic 5S rRNA types with different sequence-specific non-LTR retrotransposons. *RNA* **2017**, *23*, 446–456. [[CrossRef](#)]
80. Locati, M.D.; Pagano, J.F.B.; Girard, G.; Ensink, W.A.; van Olst, M.; van Leeuwen, S.; Nehrdich, U.; Spaink, H.P.; Rauwerda, H.; Jonker, M.J.; et al. Expression of distinct maternal and somatic 5.8S, 18S, and 28S rRNA types during zebrafish development. *RNA* **2017**, *23*, 1188–1199. [[CrossRef](#)]
81. Tao, B.; Lo, L.J.; Peng, J.; He, J. rDNA subtypes and their transcriptional expression in zebrafish at different developmental stages. *Biochem. Biophys. Res. Commun.* **2020**, *529*, 819–825. [[CrossRef](#)] [[PubMed](#)]
82. Ramachandran, S.; Krogh, N.; Jorgensen, T.E.; Johansen, S.D.; Nielsen, H.; Babiak, I. The shift from early to late types of ribosomes in zebrafish development involves changes at a subset of rRNA 2'-O-Me sites. *RNA* **2020**, *26*, 1919–1934. [[CrossRef](#)] [[PubMed](#)]

-
83. Hebras, J.; Krogh, N.; Marty, V.; Nielsen, H.; Cavaille, J. Developmental changes of rRNA ribose methylations in the mouse. *RNA Biol.* **2020**, *17*, 150–164. [[CrossRef](#)] [[PubMed](#)]
 84. Begik, O.; Lucas, M.C.; Prysycz, L.P.; Ramirez, J.M.; Medina, R.; Milenkovic, I.; Cruciani, S.; Liu, H.; Vieira, H.G.S.; Sas-Chen, A.; et al. Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat. Biotechnol.* **2021**. [[CrossRef](#)] [[PubMed](#)]

2 Ribosomes et Cancers

A Le(s) cancer(s)

Le mot cancer vient du mot grec *Καρκίνος* (carcinus) qui signifie littéralement en grec ancien « crabe ». Le terme a été utilisé pour la première fois par le philosophe grec Hippocrate à cause de la ressemblance des lésions d'une tumeur solide avec la forme d'un crabe. Aujourd'hui le terme cancer regroupe plus de 200 formes différentes de pathologies. En 2020, selon l'OMS, on estime à 19,3 millions le nombre de nouveaux cas de cancer par an dans le monde et environ 10 millions de décès liés au cancer (FERLAY et al., 2021 ; SUNG et al., 2021) (Fig : 2.1). Dans plusieurs pays (112 sur 183), le cancer est la première ou seconde cause de décès avant l'âge de 70 ans. De plus, le cancer présente un taux de mortalité et d'incidence très important dans le monde entier, quel que soit le niveau de développement du pays.

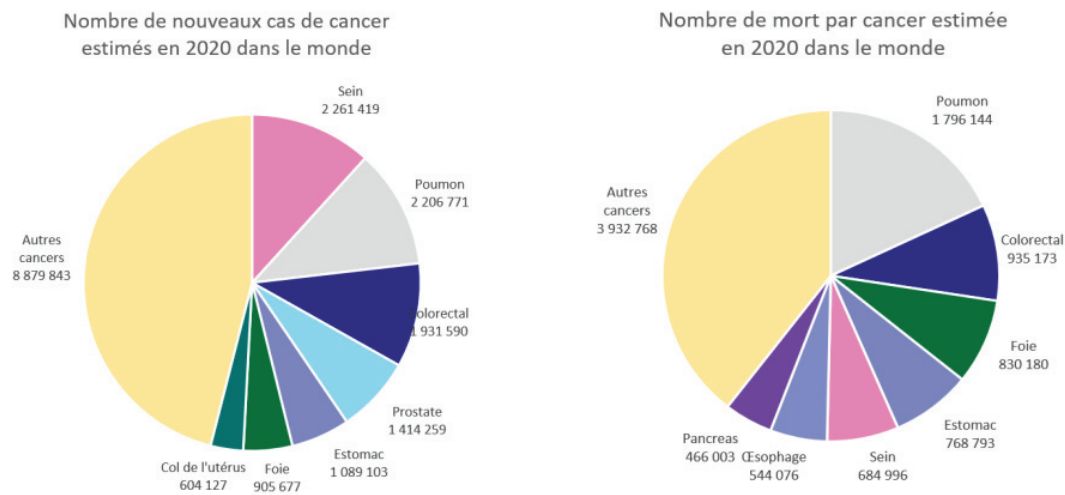


FIGURE 2.1 – À gauche les chiffres de nouveaux cas par cancer dans le monde en 2020. À droite les chiffres de mortalité par cancer dans le monde en 2020 (GLOBOCAN 2020).

Les 30-40 trillions de cellules d'un corps humain en bonne santé vivent dans

un environnement complexe et interdépendant, régulant notamment la prolifération des unes et des autres. Une cellule normale se multiplie quand elle reçoit des signaux d'autres cellules et quand l'organisme en a besoin. Cela permet à chaque tissu de conserver une architecture normale et d'assurer un rôle biologique adapté aux besoins de l'organisme. Quand une cellule normale vieillit ou accumule des dommages, elle meurt et laisse la place à de nouvelles cellules. Cependant, certaines cellules ne suivent pas ce schéma très simplifié. Elles accumulent des dommages, ou mutations, dans leurs ADN et se multiplient de manière incontrôlée. Ces cellules, dites cancéreuses, commencent alors à envahir le tissu et peuvent même migrer et envahir de nouveaux tissus voisins pour former de nouveaux cancers. Dans ce cas, on parle de métastases. Pour résumé, à l'origine, le cancer était essentiellement vu comme une maladie génétique qui est due à une accumulation de mutations ce qui entraîne une prolifération incontrôlée des cellules mutées.

En effet, des mutations génétiques ont souvent été observées dans les tumeurs et trois types de gènes impliqués dans le développement tumoral ont été identifiés : les oncogènes, les gènes suppresseurs de tumeurs et les gènes responsables de la réparation de l'ADN. Les oncogènes (ex : Her2, KRAS ou c-myc) sont des gènes qui portent des mutations dites activatrices. La présence de mutation induit en effet une hyperactivité de la protéine synthétisée qui participe directement à la transformation d'une cellule normale en cellule tumorale, souvent via une augmentation de la prolifération cellulaire. À l'inverse, les gènes suppresseurs de tumeurs (ex : p53, PTEN, Rb) sont inactivés dans les cellules cancéreuses et ne peuvent plus exercer leurs fonctions qui inhibent la transformation cellulaire, en induisant l'apoptose ou un arrêt du cycle cellulaire en réponse à un stress. De la même façon, les gènes réparateurs de l'ADN (ex : BRCA1) sont très souvent inactivés dans les cancers ce qui conduit à une augmentation du nombre de mutations.

Ainsi, le progrès des techniques de diagnostics et de séquençage haut débit ont pu mettre en évidence la complexité des cancers tout d'abord au niveau génomique. Eifert et Pwers résument très bien l'état actuel de notre compréhension :

Diversity and complexity are hallmarks of cancer genomes. Even cancers that arise from the same cell type can harbour a range of

different genetic alterations that facilitate their unrestrained expansion and eventual metastasis. As a result, the behaviour of individual tumours how they progress and eventually respond to therapy can be varied and difficult to predict. (EIFERT et POWERS, 2012)

La diversité et la complexité sont les caractéristiques des génomes du cancer. Même les cancers qui proviennent d'un même type cellulaire peuvent héberger une gamme d'altérations génétiques différentes qui facilitent leur expansion effrénée et leur éventuelle métastase. En conséquence, le comportement des tumeurs individuelles, leur évolution et leur réponse finale au traitement, peuvent varier et être difficiles à prédire.

Ces mêmes progrès techniques ont permis de montrer que les mutations ne sont pas les seuls facteurs déterminant le développement d'un cancer. La complexité de cette pathologie est due à une forte hétérogénéité qui se retrouve à plusieurs niveaux : génétiques, mais aussi phénotypiques, transcriptomiques, épigénétiques, etc. De plus, l'hétérogénéité peut être séparée en deux catégories : entre les cellules d'un même patient (intra-tumorale) et entre les patients atteints du même type de tumeur (inter-tumorale) (DAGOGO-JACK et SHAW, 2017). L'hétérogénéité intra-tumorale se manifeste par des cellules cancéreuses qui sont par exemple génétiquement différentes au sein d'une même tumeur, tandis que l'hétérogénéité inter-tumorale résulte de facteurs intrinsèques aux patients, comme les variations génétiques ou les expositions à des facteurs environnementaux particuliers. Les données de séquençage haut-débit (OMICS) ont largement contribué à identifier, comprendre et caractériser la complexité de l'hétérogénéité des cancers. Le projet Cancer Genome Atlas (TCGA), par exemple, contient des données OMICS (génomiques, transcriptomiques, protéiques, épigénétiques) de 33 cancers différents provenant de plus de 11 000 patients.

Malgré cette hétérogénéité importante, les cellules cancéreuses présentent des caractéristiques biologiques communes, quel que soit le type de cancer (Fig : 2.2). Hanahan et Weinberg ont proposé une liste de ces marqueurs des cellules cancéreuses (les fameuses « Hallmarks of cancer ») qui comprend une prolifération soutenue, une résistance aux suppresseurs de croissance, une résistance à la mort cellulaire, une réplication continue, une induction

de l'angiogenèse, un contournement du système immunitaire, et une activation de l'invasion et de métastases (HANAHAN, 2022; HANAHAN et Robert A WEINBERG, 2000, 2011) (Fig : 2.2).

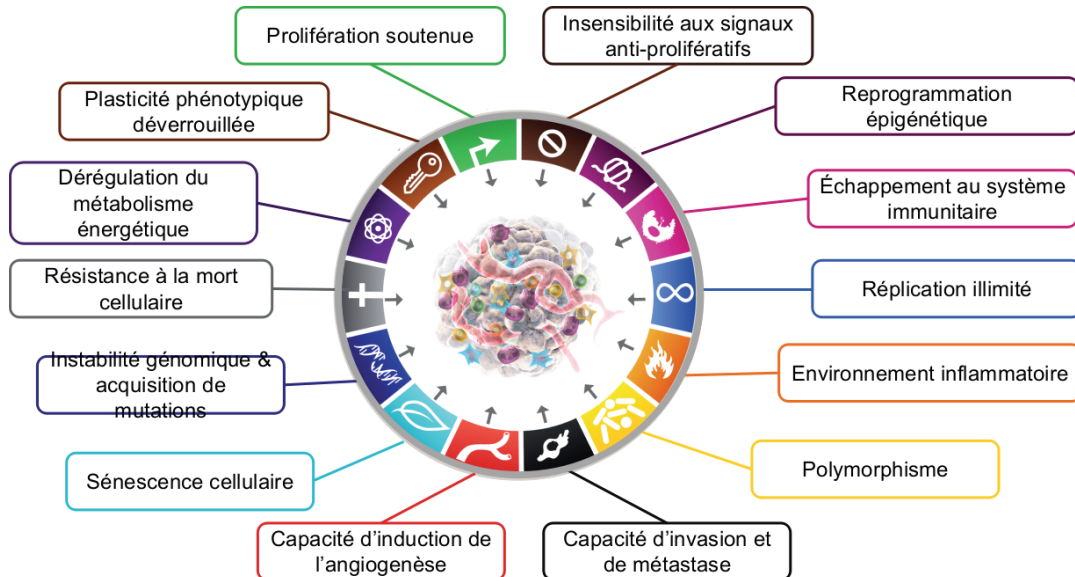


FIGURE 2.2 – Les caractéristiques des cancers. Schéma général des caractéristiques des cellules cancéreuses proposé par Hanahan et Weinberg (HANAHAN, 2022; HANAHAN et Robert A WEINBERG, 2000, 2011)

La prolifération soutenue et incontrôlée reste une des caractéristiques clés dans le développement d'une tumeur. Celle-ci dépend directement du ribosome. Dans les deux prochains sous-chapitres, je vais décrire comment le ribosome est impliqué dans le processus de la tumorigénèse (ie, formation de la tumeur) et de la progression tumorale (ie, évolution de la tumeur). Tout d'abord, les cellules cancéreuses présentent une altération quantitative du ribosome. En effet, il y a une augmentation du nombre de ribosomes dans les cellules cancéreuses pour soutenir une synthèse des protéines et ainsi leur prolifération accrue. Dans un deuxième temps, il apparaît que les cellules cancéreuses présentent des ribosomes portant des altérations qualitatives (c.-à-d., ribosomes de compositions différentes), qui pourraient permettre la traduction de certains ARNm en particulier.

B Biogenèse des ribosomes et cancer

B.1 Altération de la biogenèse des ribosomes dans les cancers

Comme nous l'avons vu dans la partie [A.2](#), la biogenèse du ribosome a lieu principalement dans le nucléole. Dès la fin du 19^{ème} siècle, Pianese [Pianese, 1896. Beitrag zur histologie und aetiologie des carcinoms (Vol. 1). G. Fischer] a rapporté que les nucléoles hypertrophiques et de forme irrégulière étaient caractéristiques des cellules cancéreuses malignes. Cependant, l'hypertrophie du nucléole est aussi observée dans des cellules normales prolifératives. Ainsi, ces caractéristiques n'ont pas été considérées comme une caractéristique spécifique des cellules cancéreuses et n'ont jamais été utilisées seules comme biomarqueur des cellules tumorales. En revanche, associée à d'autres caractéristiques morphologiques comme la forme du noyau, l'hypertrophie des nucléoles est utilisée par les anatomopathologistes pour identifier le grade des tumeurs. (PENZO, Lorenzo MONTANARO et al., 2019)

Au cours des dernières années, plusieurs études ont montré que l'hypertrophie du nucléole des cellules cancéreuses est essentiellement liée à une hyperactivation de la biogenèse des ribosomes. En effet, il s'avère que non seulement la biogenèse des ribosomes contribue à réguler le cycle cellulaire des cellules prolifératives, y compris des cellules cancéreuses, mais aussi que certains oncogènes qui contrôlent le cycle cellulaire, régulent aussi la production des ribosomes (DERENZINI, 2000; Massimo DERENZINI, Lorenzo MONTANARO et TRERÉ, 2009; Lorenzo MONTANARO, TRERÉ et Massimo DERENZINI, 2008). Ainsi, le dogme actuel est que les cellules cancéreuses ont une augmentation de la biogenèse des ribosomes permettant ainsi une hyperactivité de synthèse protéique pour soutenir leur taux de prolifération élevé. L'ensemble de ces études ont contribué à reconnaître que la morphologie, la taille et le nombre de nucléoles pourraient être des marqueurs de l'agressivité et pronostique en cancérologie (Massimo DERENZINI, Lorenzo MONTANARO et TRERÉ, 2009)

B.2 Hyperactivation de la biogenèse des ribosomes induite par des oncogènes et des suppresseurs de tumeurs clés

À ce jour, plusieurs oncogènes et suppresseurs de tumeurs ont été identifiés comme régulateurs de la biogenèse des ribosomes. Pour rappel, la synthèse d'un ribosome mature fonctionnel nécessite les trois ARN polymérase (ARN Pol), 80 protéines ribosomiques, 4 ARNr et environ 200 facteurs d'assemblage. La transcription des ARNr par les ARN PolII et PolIII est la première étape de la biogenèse des ribosomes qui a lieu principalement dans le nucléole et constitue l'étape limitante de ce processus. Dans les cellules cancéreuses, l'activité de ces deux polymérase peut être régulée par de nombreuses voies de signalisation clés en cancérologie, comme l'oncogène c-Myc ou le suppresseur de tumeur p53.

c-Myc est un facteur de transcription qui régule l'expression des gènes et participe ainsi à la croissance et à la prolifération cellulaire. De plus, c-Myc est un des oncogènes principaux en cancérologie, puisque son expression est augmentée dans de nombreux cancers suite à des remaniements chromosomiques par exemple, et permet l'initiation et la progression tumorale (DANG, 2012). En plus de ces activités de régulation de l'expression génique par l'ARN PolII, il a été montré que c-Myc augmente l'activité des PolII et PolIII et par conséquent la synthèse des quatre ARNr. Par ailleurs, c-Myc régule aussi la transcription des gènes codant pour des facteurs d'assemblage des ribosomes ou pour les protéines ribosomiques en activant l'ARN pol II (SCHLOSSER, 2003; SLIMANE et al., 2020; VENEZIA et al., 2019). Enfin, une étude récente sur des lignées cellulaires humaines de cancer de colon, a montré que la déplétion de c-Myc diminue la production de ribosome, soutenant le lien entre c-Myc et biogenèse des ribosomes (MORCELLE et al., 2019). Ainsi, l'augmentation de l'expression de cMyc dans les cancers est généralement associée à une hyperactivation de la biogenèse des ribosomes. Une étude clé a démontré l'importance de cette hyperactivation de la biogenèse des ribosomes dans l'activité oncogénique de c-Myc. Barna *et al* ont montré que dans un modèle murin qui développe des lymphomes spontanément via l'activation de l'oncogène c-Myc, la surexpression de c-Myc dans les cellules tumorales est associée à une augmentation de la synthèse protéique (BARNA et al., 2008). Chez ces souris, une haploinsuffisance du gène RPL24 entraîne

un retour à la normale du niveau de synthèse protéique, mais aussi une diminution du nombre de lymphomes spontanément développés. Ces données ont démontré pour la première fois que le pouvoir oncogénique de c-Myc repose au moins en partie sur une augmentation de la synthèse protéique. D'autres oncogènes régulent aussi la biogenèse des ribosomes. Par exemple, il a été montré qu'une isoforme de netrin-1 se lie aux composants de la PolII exclusivement dans les cellules cancéreuses. Cela augmente la transcription de l'ADN ribosomique, la biogenèse des ribosomes, mais aussi la prolifération des cellules et la tumorigenèse (DELLOYE-BOURGEOIS et al., 2012).

De l'autre côté, les gènes suppresseurs de tumeurs comme p53 inhibent la prolifération cellulaire et ainsi l'initiation et la progression tumorale. Il a été montré dans des cellules cancéreuses que p53 inhibe l'activité de transcription de PolII et PolIII et donc participe à l'inhibition de la biogenèse des ribosomes (CAIRNS, 1998; ZHAI et COMAI, 2000). Par ailleurs, de nombreux mécanismes moléculaires ont montré l'implication de p53 dans la régulation de la biogenèse des ribosomes, mais aussi de la synthèse protéique (MARCEL, CATEZ et DIAZ, 2015). En revanche, l'importance de l'inhibition de la biogenèse des ribosomes dans l'activité suppressive de p53 n'a pas été démontrée clairement.

B.3 Ciblage thérapeutique de la biogenèse des ribosomes

Ces études fonctionnelles montrent que la biogenèse des ribosomes n'est pas seulement nécessaire pour soutenir les besoins des cellules cancéreuses en synthèse protéique, mais contribue directement à la tumorigenèse. De plus, des travaux plus récents ont montré que l'effet de nombreuses chimiothérapies repose sur l'inhibition de la biogenèse des ribosomes, comme par exemple l'oxaliplatine (BRUNO et al., 2017; Frédéric CATEZ et al., 2019). Ainsi, la biogenèse des ribosomes s'avère être une cible thérapeutique intéressante en cancérologie (Frédéric CATEZ et al., 2019). Deux études princeps ont montré que l'inhibition de la biogenèse des ribosomes en inhibant l'activité de la PolII tue spécifiquement les cellules cancéreuses sans affecter les cellules normales (BYWATER et al., 2012; Frédéric CATEZ et al., 2019; PELTONEN et al., 2014). L'hypothèse est que les cellules qui ont un taux de prolifération plus bas que les cellules tumorales, comme les cellules normales, ont besoin de moins de ribosomes néo-synthétisés en raison de la stabilité de ces derniers et donc sont moins sensibles à l'inhibition de la biogenèse.

C Hétérogénéité de composition des ribosomes et cancer

En plus d'une augmentation de la biogenèse des ribosomes, il s'avère que les cellules tumorales présentent une hétérogénéité de la composition des ribosomes. L'hypothèse est que ces ribosomes spécifiques des cellules cancéreuses favoriseraient la traduction de certains oncogènes et inhiberaient celle de suppresseur de tumeurs pour contribuer à l'initiation et à la progression tumorale. Ici, je reprendrai le plan utilisé dans le chapitre B pour illustrer le lien entre les différents composants du ribosome participant à sa diversité et le cancer.

C.1 Altération des RPs dans les cancers

Depuis la fin des années 1990, de nombreuses études ont identifié des altérations au niveau des RPs dans les cancers, que ce soit la présence de mutation ou une dérégulation de leur expression. En revanche, la contribution de ces altérations de RPs dans l'acquisition de phénotypes tumorales commence seulement à émerger. Aujourd'hui, de nombreuses interrogations restent quant à leur contribution, tant au niveau moléculaire que cellulaire, dans le développement tumoral.

C.1.1 Mutation dans les gènes codant les RPs, ribosomopathies et cancer

Les ribosomopathies sont un ensemble de maladies génétiques ayant pour caractéristique commune un ribosome ou une synthèse du ribosome défectueux (S. SULIMA, K. KAMPEN et KEERSMAECKER, 2019). La plupart des ribosomopathies sont des maladies génétiques congénitales (c.-à-d., présentes à la naissance) causées par l'haploinsuffisance des gènes codant pour un facteur de la biogenèse du ribosome ou pour des protéines ribosomiques (RPs). On peut citer l'anémie de Diamond-Blackfan (DBA), le syndrome de Shwachman-Diamond (SDS), le syndrome de Treacher Collins (STC), la délétion du chromosome 5q (ou syndrome du cri de chat), la dyskératose congénitale ou encore l'asplénie congénitale familiale isolée. Beaucoup de ces syndromes présentent des symptômes cliniques communs et induits par des

troubles de l'hypoprolifération (c.-à-d., diminution de la prolifération) : malformation, déficience mentale et motrice, ou insuffisance médullaire. Les patients atteints de ribosomopathies ont un risque accru de développer un cancer au cours de leur vie. Pour certaines ribosomopathies, ce risque est augmenté de 2,5 à 8,5 fois par rapport au risque de développer un cancer par la population générale (S. SULIMA, K. KAMPEN et KEERSMAECKER, 2019). Les ribosomopathies soutiennent un lien fort entre altération de la biogenèse des ribosomes et développement tumoral. Par ailleurs, la description de mutation dans les gènes codant les RP dans les ribosomopathies suggèrent l'existence d'une hétérogénéité de composition du ribosome dûe aux RPs. Ici, je donnerai quelques exemples de mutations dans les gènes codant pour les RPs identifiées dans les ribosomopathies et ayant montré un rôle dans le développement tumoral.

La DBA est la première ribosomopathie identifiée et une des plus étudiées. En 1999, des mutations de type translocations et délétions ont été identifiées sur le gène codant pour la protéine RPS19 dans environ 25 % des patients atteints de DBA, qui résultent majoritairement toutes en mutations de type faux-sens (c.-à-d., changement d'un acide aminé par un autre) (CAMPAGNOLI et al., 2008; DRAPTCHINSKAIA et al., 1999). Depuis, d'autres mutations ont été identifiées dans au moins 20 RPs de la petite et de la grande sous-unité du ribosome : RPL5 (7 %), RPS10 (6 %), RPL11 (5 %), RPS26 (3 %) RPL35A (3 %), RPS24 (2 %), RPS17 (1 %), RPS7 (1 %), RPL26, RPL27, RPS27, RPL15, RPS28, RPL31, RPS29 (FARRAR et al., 2008; GAZDA et al., 2008; QUARELLO et al., 2009; M.-J. SONG et al., 2009). Toutes ces mutations sont hétérozygotes, car les mutations homozygotes sont souvent létales, comme démontré à l'aide de modèle animaux (COSTA, LEBLANC et MOHANDAS, 2020; S. O. SULIMA et al., 2017). La DBA est associée à une augmentation du risque de cancer et favorise le développement de leucémie myéloïde aigue, syndromes myélo-dysplasiques et d'autres types de cancer, soutenant le lien entre la mutation de ces différentes RPs et le développement tumoral.

Ce lien entre mutation des RPs et cancer est renforcé par l'étude de Fancello *et al*, qui a analysé les données de mutations et du nombre de copies provenant de la base de données publique The Cancer Genome Atlas (TCGA) dans 16 types de cancer. Ils ont identifié six gènes codant pour des RPs présentant une mutation ou une délétion hétérozygote dans les cancers qui sont aussi clés pour la DBA : RPL5, RPL11, RPL23A, RPS5, RPS20 et RPSA. Par

exemple, RPL5 est un des gènes majoritairement mutés dans la DBA, et est altéré dans de nombreuses tumeurs : 11 % des glioblastomes, 28 % des mélanomes, 34 % des cancers du sein, et 20 % des tumeurs de myélome multiple (FANCELLO et al., 2017). Par ailleurs, il a été montré que la réduction de 50 % du niveau d'expression de RPL5 dans des lignées cellulaires de cancer du sein augmente la prolifération cellulaire *in vitro*, mais aussi la progression tumorale *in vivo*, à l'aide de xénogreffe chez la souris (FANCELLO et al., 2017). Ces résultats soutiennent l'hypothèse d'un rôle de tumeur suppresseur de RPL5.

D'autres mutations dans les gènes codant les RPs ont été identifiées par différentes études chez des patients atteints de leucémie aiguë lymphoblastique (LAL). Par exemple, RPL10 est muté chez 8 % des patients atteints de LAL (KEERSMAECKER et al., 2012). Les mutations observées sont de type faux-sens (c.-à-d., remplacement d'un nucléotide par un autre) et entraînent sur la protéine synthétisée, le remplacement de l'acide aminé de l'arginine en position 98 par une sérine (RPL10-R98S). Plus récemment, Kampen *et al* ont montré que les lignées cellulaires leucémiques exprimant la protéine mutante RPL10-R98S survivent à des niveaux élevés de stress oxydatif et augmente la viabilité cellulaire et la prolifération cellulaire (K. R. KAMPEN et al., 2018). Il s'avère que cette protéine RPL10-R98S favorise la traduction de la protéine anti-apoptotique Bcl2 par un mécanisme dépendant des IRES. Par ailleurs, l'inhibition de Bcl2 par une molécule utilisée en clinique (Venotoclax) chez des souris exprimant RPL10-R98S suffit à supprimer la présence de cellules leucémiques dans le sang des animaux. Cet exemple démontre qu'une altération de protéine ribosomique par la présence d'une mutation affecte la traduction de certains ARNm et favorise la tumorigenèse.

D'autres observations *in vivo* soutiennent l'hypothèse de l'implication des mutations des RP dans la tumorigenèse. Une étude de 2004 chez le poisson zèbre a permis d'identifier 11 altérations génétiques (mutation ou délétion hétérozygote) dans des gènes codants des RP différentes qui augmente la probabilité de développer un cancer (AMSTERDAM et al., 2004). De plus, l'haplosuffisance de RPL11 et de RPL22 favorise le développement tumoral respectivement dans des modèles de lymphome et de sarcome chez la souris (MORGADO-PALACIN et al., 2015; RAO et al., 2012).

Ces quelques exemples montrent que les mutations dans les gènes codant des

RP contribuent à la formation de tumeur. Cependant, encore peu d'études se sont intéressées au rôle de ces mutations dans le développement tumoral et l'impact des mutations des RPs dans les cancers reste probablement sous-estimé. De plus, seulement peu d'études ont identifié le rôle moléculaire de ces RPs mutantes, une seule étude ayant montré un rôle dans la régulation traductionnelle. Enfin, aucune étude n'a clairement démontré que ces mutations affectaient la composition du ribosome mature et fonctionnel, et donc l'activité intrinsèque du ribosome.

C.1.2 Altération de l'expression des RP dans le cancer

En plus de la présence de mutations, il apparaît que l'expression des gènes codant pour les RP est altérée dans les cancers. En effet, plusieurs RP sont surexprimées dans différents cancers, comme les cancers de la prostate, du côlon, du foie, du pancréas, de l'estomac, du poumon et du sein, entre autres (GOUDARZI et LINDSTRÖM, 2016; KHOURY et NASR, 2021; PENZO, Lorenzo MONTANARO et al., 2019). Par exemple, des analyses immunohistochimiques de tissus tumoraux provenant de 60 patients atteints de carcinome hépatocellulaire (c.-à-d., cancer primitif du foie, ou CHC) ont montré que RPL36 est surexprimée dans 75 % des cas par rapport aux tissus non tumoraux environnants. En particulier, RPL36 est plus exprimée dans les premiers stades du cancer (stade I et II) et est associé à une meilleure survie globale (M. J. SONG et al., 2011). Plusieurs études suggèrent un lien entre facteurs environnementaux et altération de l'expression des RPs, notamment dans les CHC. Par exemple, RPS8 est surexprimée dans les CHC des patients présentant une forte consommation d'alcool comparé aux patients atteints de CHC sans consommation d'alcool (BI et al., 2020). Par ailleurs, une infection chronique par le virus de l'hépatite B peut favoriser le développement d'un CHC par le biais de la protéine virale HBx qui active plusieurs voies de signalisation cellulaire et induit la tumorigenèse. Or, plusieurs RPs (ex : RPS20, RPS24, RPL8, RPL31...) sont surexprimées dans les CHC associés à une infection par le virus de l'hépatite B (KIM et al., 2001). Enfin, dans un modèle cellulaire de CHC induit par l'hépatite B, deux études ont déterminé que les protéines RPS3a et RPS27a sont surexprimées dans ces conditions (FATIMA, MATHAN et KUMAR, 2012; LIM et al., 2011). De la même façon que pour le CHC, plusieurs RPs peuvent être surexprimées dans un même type de cancer, par exemple RPL23, RPS27, RPL6 et RPL13 dans les cancers de l'estomac

(Y. SHI et al., 2004; WU et al., 2011). Néanmoins, ces études ne se sont focalisées que sur l'analyse de l'expression de RP individuellement. La question reste de savoir si la surexpression des RP est (i) coordonnée ou non et (ii) induit la production de ribosomes de compositions différentes.

En plus de ces données observationnelles, quelques données suggèrent que la surexpression de RPs favorise l'initiation et la progression tumorale, bien qu'une démonstration du lien direct soit toujours manquante. Par exemple, RPS13 est plus exprimée dans les tissus et les lignées cellulaires du cancer de l'estomac que dans les tissus de la muqueuse gastrique et des lignées cellulaires épithéliales normales, indiquant une surexpression de RPS13 dans le cancer de l'estomac. De plus, la réduction de l'expression de cette protéine inhibe la croissance des cellules cancéreuses gastriques, suggérant que la surexpression de RPL13 contribuerait au développement tumoral. Des études plus récentes s'intéressant à la résistance aux thérapies des cancers de l'estomac, ont montré une association entre expression des RPs et résistance. Le 5-Fluorouracil (5-FU) est une molécule largement utilisée pour le traitement de différents cancers, y compris de l'estomac. Des analyses de survie de type Kaplan-Meier ont indiqué que l'expression élevée de RPL11 chez les patients atteints de cancer de l'estomac traités par 5-FU était significativement associée à un bon pronostic (KAWAHATA et al., 2020). En plus d'être un potentiel biomarqueur pronostique, ces données suggèrent un rôle potentiel de la surexpression de RPL11 en combinaison avec le 5-FU dans l'efficacité du traitement du cancer de l'estomac.

En plus d'être surexprimées dans les cancers indiquant un rôle oncogénique, il apparaît que certaines RP sont sous-exprimées dans certains types de cancers indiquant que les RPs pourraient aussi avoir un rôle suppresseur de tumeur. Par exemple, RPL27L est sous-exprimée dans les cancers du sein et du colon (HUANG et al., 2013; XIONG et al., 2018). De plus, des analyses immunohistochimiques, de western-blot et de RT-qPCR ont montré que RPL22 est sous-exprimée au niveau ARNm et protéine dans le cancer du poumon non à petites cellules (CPNPC). Ces résultats d'analyse d'expression différentielle des RP dans différents types de cancers, ainsi que les quelques données fonctionnelles, montre un lien entre protéines ribosomiques et cancer. La modulation de l'expression des RP a été associée à différentes fonctions, comme la régulation du cycle cellulaire, du métabolisme cellulaire, de la survie cellulaire jusqu'à la tumorigénèse. Ainsi, les RPs sont des candidats intéressants

en clinique. Elles pourraient servir de biomarqueurs pronostique en fonction du type de cancer, mais aussi de potentielles cibles thérapeutiques. Il est cependant important de noter que l'expression différentielle des RPs n'implique pas forcément une hétérogénéité de composition du ribosome. En effet, certaines RPs, en plus d'être des constituants essentiels de la structure du ribosome, peuvent aussi exercer des fonctions indépendantes du ribosome. Parmi les fonctions indépendantes du ribosome, on peut mentionner un rôle de co-facteur de la transcription ou de régulation de la maturation et de la stabilité des ARNm, qui contribue aussi à la régulation de l'apoptose, de l'autophagie ou de la migration cellulaire (PECORARO et al., 2021). Par ailleurs, il faut noter que l'altération d'expression des RPs peut aussi résulter d'une dérégulation de leur traduction et ainsi participer à la tumorigenèse et à la résistance aux traitements. Des métastases pulmonaires prélevées chez la souris présentent une surexpression de plusieurs RPs, dont RPL15, qui lorsque surexprimée dans des cellules tumorales primaires induit une augmentation de la traduction d'autres RPs mais aussi de facteurs de transcription pour induire la formation de métastases *in vivo* (EBRIGHT et al., 2020).

C.1.3 Conclusion

En conséquence, l'hypothèse que l'altération de l'expression et/ou des séquences protéiques des RPs participe à la génération de ribosome de composition différente et affecte ainsi son activité, ne peut être exclue, mais requiert des travaux plus poussés. Afin de mieux comprendre le rôle des RPs dans le développement et la progression tumorale, il est donc nécessaire d'analyser les altérations de RPs survenant à plusieurs niveaux : (i) au niveau génique par l'analyse de mutation et du nombre de copies, (ii) au niveau de son expression, tant au niveau de l'expression de l'ARNm que protéique et (iii) au niveau de son incorporation ou non dans le ribosome. Une telle approche intégrée permettrait de démontrer le rôle de l'altération des RPs dans les cancers et de définir précisément leurs rôles moléculaires, comme protéines libres ou comme constituants d'un ribosome « spécialisé ».

C.2 Altération des modifications chimiques des ARNr dans les cancers

Alors que les altérations des RPs dans les cancers ont été décrites dès la fin des années 1990, c'est seulement au cours des 10 dernières années que des études ont permis de clairement démontrer le rôle des modifications chimiques des ARNr dans les cancers. Au début de ma thèse en 2019, la plupart des données disponibles correspondaient au rôle des pseudouridilations des ARNr dans les cancers, aucune donnée n'était disponible pour les modifications de bases et très peu pour la 2' Ome des ARNr. Aujourd'hui, le nombre d'études dans le domaine a très largement augmenté. Afin de bien contextualiser mon projet de thèse centré sur l'altération de la 2' Ome dans les cancers, je décrirai de manière la plus complète possible les données de la littérature disponible à ce jour pour les pseudouridilations et les modifications de bases dans les cancers, mais je ne présenterai ici que les données sur la 2' Ome des ARNr disponibles au début de mon projet de thèse.

C.2.1 Méthylation de bases des ARNr

Alors que plusieurs modifications de bases des ARNr ont été décrites, seules les méthylations de bases des ARNr ont été étudiées dans le contexte du cancer, certainement car les méthyl-transférases des ARNr sont bien décrites, souvent spécifiques des ARNr et relativement faciles à manipuler expérimentalement.

Chez l'Homme, la méthyltransférase NSUN5 est responsable de la méthylation du carbone en position 5 de la cytosine représentant le nucléotide 3782 de l'ARNr 28S (28S_m5C3782). Jiang *et al* ont constaté à partir des données de la base de données publique TCGA et de leurs données de microarray que NSUN5 est surexprimée dans les tissus de cancer du côlon par rapport aux tissus normaux (Z. JIANG et al., 2020). Par ailleurs, la diminution de l'expression de NSUN5 dans des lignées cellulaires provoque un arrêt du cycle cellulaire. Cependant, l'impact de l'altération de l'expression de NSUN5 sur la méthylation 28S_m5C3782 n'a pas été analysé. En parallèle d'un potentiel rôle oncogénique dans les cancers du côlon, un rôle suppresseur de tumeur a été décrit dans les glioblastomes. Il a été montré que l'hyperméthylation du promoteur NSUN5, induisant une diminution de son expression au niveau

ARNm, est associée à un bon pronostic chez les patients atteints de glioblastome (donnée TCGA). La diminution de l'expression de NSUN5 est associée à une diminution de la méthylation de la cytosine 28S_m5C3782 et entraîne une diminution de la synthèse protéique globale, associée à une augmentation de la traduction de certains facteurs de survie dans des conditions de stress (JANIN et al., 2019). Ces données suggèrent un rôle de la méthylation de l'ARNr en position 28S_m5C3782 dans le développement tumoral via une altération de l'activité du ribosome. En effet, cette étude a montré par des simulations informatiques de la structure du ribosome que la perte de la méthylation 28S_m5C3782 pourrait impacter directement le site P du ribosome (site de formation de la liaison peptidique). Il serait intéressant de définir si la dérégulation traductionnelle est la conséquence (i) de la diminution de l'expression de la méthyltransférase NSUN5 qui pourrait réguler la méthylation d'ARNm ou l'activité de protéine impliquée dans la régulation de la traduction, (ii) de l'altération de la structure du ribosome via la perte de la méthylation 28S_m5C3782 ou (iii) de la combinaison des deux. Une telle démonstration pourrait être réalisée à l'aide d'un mécanisme de traduction *in vitro* qui permettrait d'analyser la capacité de traduction intrinsèque des ribosomes avec et sans la méthylation 28S_m5C3782. En effet, ce système permet de mettre en présence dans un tube à essai, l'ensemble des éléments nécessaires à la traduction (tRNA, mRNA rapporteur, eIFs...) purifiés à partir de réticulocyte de lapin et des ribosomes d'intérêt.

Une approche similaire a permis d'identifier le rôle de la méthylation 18S_m6A1832 dans le développement tumoral. L'expression au niveau protéique de METTL5, la méthyltransférase responsable de la méthylation au site 18S_m6A1832 localisée dans le site P et proche du centre de décodage de la petite sous-unité du ribosome, est élevée chez les patients atteints de cancer du sein (RONG et al., 2020). L'étude *in vitro* dans des lignées cellulaires de cancer du sein a montré que la perte de METTL5 réduit le nombre de ribosomes dans les fractions polysomiques, qui contiennent les ribosomes en cours de traduction, et par conséquent, une diminution de la synthèse protéique globale. De plus, la perte de METTL5 diminue la viabilité d'un panel de lignées cellulaires mammaires. En revanche, le mécanisme par lequel la méthylation 18S_m6A1832 située dans le centre de décodage du ribosome peut affecter la traduction reste à élucider.

Une dernière étude soutient le rôle des méthylations de base des ARNr dans

le cancer. Babaian *et al* se sont intéressés aux variations nucléotidiques des ARNr dans environ 10 000 patients atteints de différents types de cancers et à environ 700 échantillons normaux contrôles. Les données de séquençage RNA-seq ont été alignés sur une copie de l'ADNr. Les auteurs ont identifié une variation nucléotidique de l'ARNr 18S en position 1248, localisée dans le centre de décodage du ribosome, dans plus de 22 types de cancers, mais pas dans les échantillons contrôles. La position 1248 de l'ARNr mature 18S est connue pour subir trois modifications chimiques différentes pour former une 1-méthyl-3- α -amino- α -carboxyl-propyl pseudouridine (18S_m1acp3 Ψ 1248). Cette modification chimique perturbe l'appariement des nucléotides complémentaires lors de l'étape de la synthèse de l'ADN complémentaire par la transcriptase inverse, ce qui entraîne une mauvaise incorporation de nucléotide et donc dans une variation nucléotidique après le séquençage. Ainsi, l'identification d'une variation génétique au départ a permis l'identification de l'altération d'une modification chimique de l'ARNr. Cette étude a ainsi montré que 45,9 % des patients atteints de cancer du côlon présentait une variation du niveau de la méthylation en position 18S_m1acp3 Ψ 1248. Ces résultats ont été validés *in vitro* dans des lignées cellulaires humaines de cancer du côlon. Par ailleurs, la perte de la modification chimique 18S_m1acp3 Ψ 1248 induite par la délétion de TSR3, responsable de l'addition du groupement acp3, entraîne une altération de la traduction d'un sous-groupe d'ARNm codant pour des protéines ribosomiques comme RPL4 et RPS8 (BABAIAN *et al.*, 2020). En revanche, cette étude n'a pas montré l'effet de la délétion de TSR3 sur le phénotype cancéreux. De plus, les données RNA-seq utilisées ont peu d'information sur les ARNr qui sont souvent enlevés pour pouvoir amplifier l'information des ARNm. Cela peut induire une augmentation du nombre de faux positifs ou de faux négatifs concernant la présence ou non de la modification 18S_m1acp3 Ψ 1248.

Ces résultats accumulés au cours de ces trois dernières années soutiennent un rôle des méthylations de bases des ARNr dans la tumorigenèse, certainement via une altération de la traduction de certains ARNm particuliers.

C.2.2 Pseudouridilation des ARNr

Les premiers liens entre pseudouridylation des ARNr et cancer ont été établis dès la fin des années 90 quand des mutations du gène de la dyskérine (DKC1), responsable de la pseudouridilation des ARNr, ont été identifiés

dans la ribosomopathie de type dyskératose congénitale (HEISS et al., 1998). La génération de modèle murin mimant les caractéristiques cliniques de la dyskératose congénitale par introduction d'une mutation de la pseudouridylylase DKC1 a permis de montrer le rôle des pseudouridyliations dans la régulation traductionnelle (cf chapitre 1.3). Par ailleurs, elle supporte aussi un rôle des pseudouridyliations des ARNr dans les cancers. En effet, outre que la dyskératose congénitale est associée à une forte susceptibilité au cancer, l'expression de DKC1 est diminuée dans différents types de cancer y compris dans les cancers du sein, du poumon, de la prostate ou de la tête et du cou (PENZO et Lorenzo MONTANARO, 2018). Souvent, cette diminution d'expression de DKC1 est corrélée avec la progression tumorale (ALAWI et al., 2010; MONTANARO et al., 2006) et un mauvais pronostic (LIU et al., 2012; PENZO, LUDOVINI et al., 2015). En revanche, sachant le rôle de DKC1 dans la pseudouridylation des ARNr et dans la maintenance des télomères, à l'heure actuelle, il est difficile de savoir si l'activité oncogénique de DKC1 repose sur les ribosomes ou sur les télomères.

En plus des données supportant un rôle de la diminution globale de la pseudouridylation des ARNr dans les cancers, plusieurs études ont commencé à s'intéresser aux rôles des snoARN à boîte H/ACA qui guident la pseudouridylation de l'ARNr dans les cancers. Ces approches permettent d'envisager le rôle de la pseudouridylation dans les cancers à certains sites spécifiquement, comme pour la méthylation de base des ARNr décrite précédemment. Je vais décrire brièvement deux exemples récents où la diminution d'expression d'un snoARN H/ACA altère la pseudouridylation de l'ARNr à des sites spécifiques et contribue à l'acquisition de phénotype tumoral.

En 2019, McMahon *et al* ont analysé l'expression des snoARN H/ACA dans différents modèles de cancer (lignées cellulaires, souris, échantillons de patients atteints de cancer). Il s'est avéré que l'expression de certains snoARN est affectée lors de la réponse cellulaire à des signaux de croissance caractéristiques des cancers. Un de ces snoARN est le SNORA24 qui guide deux modifications de type pseudouridine de l'ARNr 18S : 18S_PsiU609 et 18S_PsiU863. Ils ont constaté que dans un modèle de souris, la perte de l'expression de SNORA24 est associée à une diminution de la pseudouridylation à ces deux sites et favorise le développement du cancer de foie induit par l'oncogène RAS. La comparaison de 91 échantillons humains de cancer hépatocellulaire

(CHC) à des tissus hépatiques non tumoraux appariés a permis de confirmer la sous-expression de SNORA24 dans les CHC, qui est aussi associée à un mauvais pronostic (MCMAHON et al., 2019). Les auteurs se sont ensuite intéressés aux ribosomes dépourvus des modifications chimiques guidées par SNORA24 et à l'impact sur la traduction. Comme détaillé précédemment (chapitre C.1.2), alors que l'absence des deux pseudouridilations n'affecte pas la biogenèse des ribosomes ou la synthèse protéique globale, une augmentation de défauts de fidélité de traduction est observée (incorporation des mauvais acides aminés et mauvaise lecture des codons stop). À noter que ces altérations surviennent dans des conditions où l'expression de SNORA24 est diminuée. En revanche, comment ces défauts de traduction expliquent le rôle de SNORA24 dans la tumorigenèse hépatocellulaire reste à explorer.

De la même façon, dans une étude plus récente, Faucher-Giguère *et al* ont utilisé une technique de séquençage haut-débit pour comparer l'abondance des snoARNs dans les carcinomes séreux de haut grade (HGSC), les tumeurs dites borderlines (TB) de bas grade et les tissus sains. Ils ont identifié trois snoARN H/ACA (SNORA81, SNORA19 et SNORA56) capables de discriminer les HGSC, TB et les tissus sains. En particulier, SNORA81 est plus exprimé dans les HGSC comparé au TB et son niveau d'expression corrèle avec le niveau de pseudouridilation de son site 28S_Ψ4636. La sous-expression de SNORA81 réduit le niveau de pseudouridilation au site 28S_Ψ4636 et réduit la prolifération et la migration cellulaire (FAUCHER-GIGUÈRE et al., 2022). Même si l'altération de la traduction n'a pas été étudiée en réponse à l'altération de la pseudouridilation, ces résultats soutiennent un lien entre pseudouridilation et tumorigenèse.

Mis à part ces études ayant clairement affichées leur volonté d'analyser le rôle des pseudouridilations dans les cancers, d'autres données soutiennent cette hypothèse. Notamment, la dérégulation de l'expression des snoARN H/ACA qui guident la pseudouridilation des ARNr a été constatée dans plusieurs cancers solides et hématologiques, soutenant l'association entre pseudouridilation et développement tumoral. Un résumé des snoARN dérégulés dans différents types de cancers est montré dans la figure 2.3. L'ensemble de ces études soutient une potentielle utilisation en clinique des snoARN

H/ACA, en particulier comme biomarqueurs. De plus la modulation de l'expression de nombreux snoARN H/ACA affecte l'acquisition de caractéristiques cancéreuses.

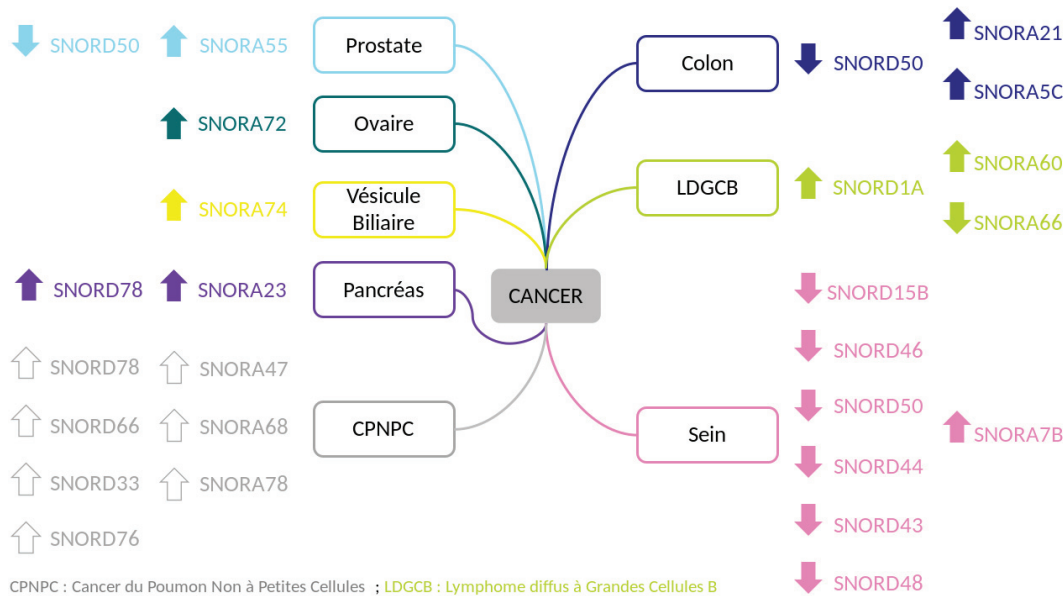


FIGURE 2.3 – Liste des snoARN dérégulés dans les cancers

L'ensemble de ces données soutient l'hypothèse qu'une altération de la pseudouridilation des ARNr contribue à la tumorigénèse en modulant la traduction de certains ARNm. En revanche, il reste aujourd'hui à montrer par quels mécanismes moléculaires et cellulaires, le ribosome dépourvu de certaines pseudouridilations participe à la tumorigénèse. En revanche, il ne faut pas oublier que de nombreux snoARN ont des activités non-canoniques et pourraient contribuer à la tumorigénèse via une régulation de l'expression génique indépendante du ribosome et qu'une simple démonstration par modulation de l'expression des snoARN limite la démonstration du rôle des modifications chimiques des ARNr dans le phénotype.

C.2.3 La 2^{Ome} des ARNr

Tout comme pour la pseudouridilation, la variation d'expression de snoARN C/D soutient une variabilité de la 2^{Ome} des ARNr dans les cancers. Une liste non exhaustive des snoARN C/D altérés dans différents cancers est présentée dans la figure 2.3. Les premières évidences d'un lien entre 2^{Ome} et cancer proviennent de l'étude de lignées cellulaires humaines du cancer du sein réalisées au sein de mon laboratoire d'accueil. En 2009, Belin *et al* ont

montré que des cellules cancéreuses mammaires présentant un phénotype agressif induit par la surexpression d'Arl2 avaient une altération de la 2'Ome des ARNr à certains sites seulement (BELIN et al., 2009). Même si ces résultats suggéraient que le phénotype agressif des cellules cancéreuses serait associé à une altération de la 2'Ome des ARNr, le lien fonctionnel n'avait pas été abordé.

Plus tard, mon équipe a montré que la surexpression de FBL, la méthyltransférase responsable des 2'Ome des ARNr, induit par l'inactivation de p53, entraîne une altération (i) de la 2'Ome des ARNr à certains sites seulement, (ii) de la traduction (cf chapitre A.4), et (iii) de certains phénotypes tumoraux (Virginie MARCEL, GHAYAD et al., 2013). En effet, la sur-expression de FBL augmente la prolifération cellulaire et la résistance à la doxorubicine. Par ailleurs, la sur-expression de FBL favorise la traduction d'IGF1R via la présence d'IRES dans la région 5'UTR de son ARNm. De manière intéressante, il s'avère que l'augmentation de la prolifération cellulaire induite par la sur-expression de FBL est abolie en présence d'un inhibiteur d'IGF1R démontrant le lien entre activation d'IGF1R et le phénotype tumoral induit par FBL. Par ailleurs, d'autres études ont démontré que FBL est surexprimée dans les tumeurs mammaires et que cette surexpression est associée à un mauvais pronostic (LONG et al., 2022; Virginie MARCEL, GHAYAD et al., 2013; SU et al., 2013; J. ZHANG et al., 2020). Ces données suggéraient indirectement l'existence d'altérations de la 2'Ome dans les cancers, mais aussi d'un rôle de la 2'Ome des ARNr dans l'acquisition de phénotype tumoral.

Avant le début de ma thèse, la totalité des études soutenant l'existence d'une altération de la 2'Ome dans les cancers provenait de l'étude de lignées cellulaires. Plusieurs études avaient en effet analysé la 2'Ome des ARNr dans des lignées cellulaires d'origine cancéreuses comme les cellules HeLa (cancer du col de l'utérus), HCT116 (cancer du côlon) ou PA-1 (tératome d'ovaire) (ERALES et al., 2017; KROGH, JANSSON et al., 2016; TAOKA, NOBE, YAMAKI, SATO et al., 2018; ZHU, PIRNIE et CARMICHAEL, 2017). Dans ces différentes lignées cellulaires, au moins un tiers des sites méthylés de l'ARNr présentaient un taux de 2'Ome partiel (c.-à-d. inférieur à 1), suggérant la présence d'un mix d'ARNr porteur d'une 2'Ome et d'ARNr non porteur d'une 2'Ome dans ces lignées cellulaires. Cette variabilité au sein d'une même population de cellules est définie comme une intra-variabilité de la 2'Ome. Par ailleurs, certains sites présentaient des différences significatives de niveau de 2'Ome

entre deux lignées cellulaires (KROGH, JANSSON et al., 2016). Dans ce cas, on parlera d'inter-variabilité de la 2'Ome.

Plusieurs questions étaient en suspens dans le domaine de la 2'Ome des ARNr en cancérologie. Existe-t-il une hétérogénéité de la 2'Ome dans les tissus humains? Est-ce que cette variabilité de la 2'Ome impacte tous les sites ou certains sites spécifiquement? Est-ce que la variabilité est aléatoire ou reflète des caractéristiques tumorales particulières?

L'analyse de la diversité de la 2'Ome des ARNr et de leur rôle dans les cancers a été largement freinée par le manque de méthodologies de quantification du taux de méthylation sur l'ensemble de l'ARNr. Dans le prochain chapitre, nous allons voir comment le développement de méthodes de séquençage haut-débit de transcriptomes et leurs utilisations peut permettre de quantifier et identifier les altérations de 2'Ome.

3 Détection et quantification de la 2'Ome via les approches de séquençage haut-débit

Comme nous l'avons vu dans les chapitres I et II, les ARNr ne sont pas de simples molécules enchaînant les nucléotides A, C, G et U présentant une activité passive. Les ARNr sont hautement structurés et riches en modifications post-transcriptionnelles (PTM) qui sont fondamentales pour leur activité ribozyme impliquée dans la catalyse de la synthèse protéique dans la cellule. Aujourd'hui, il est crucial de développer des technologies pour analyser les modifications des ARNr dans leur globalité, par des approches de type OMICS. En effet, l'arrivée des techniques de séquençage dans les années 1970 a ouvert de nouvelles portes pour améliorer la compréhension de l'expression génique et en particulier dans le domaine de la cancérologie. Initialement développée pour déterminer la suite de nucléotides d'une molécule d'ADN ou d'ARN, l'évolution des techniques a permis de s'intéresser directement à des différences de niveau d'expression d'ARNm, et plus récemment de répondre à de nombreuses questions quant à la maturation des ARNm par exemple. Dans ce chapitre, nous allons discuter la technique de RNA-seq, un séquençage nouvelle génération de transcriptome, et comment cette technique est utilisée pour identifier et quantifier la 2'-O-ribose méthylation des ARNr.

A Séquençage du transcriptome par RNA-seq

Le RNA-seq permet d'étudier les ARN en général sans avoir des connaissances *a priori* sur la séquence à analyser. Initialement dédiées à l'étude du transcriptome des cellules, de nombreuses adaptations de la préparation des échantillons en amont ont permis d'analyser d'autres aspects liés aux ARN,

comme les interactions ARN : protéines (RIPseq, CLIPseq...), la structuration des ARN (PARS, FragSeq...) et plus récemment les modifications chimiques des ARN (RiboMethSeq).

A.1 Matériel biologique

Comme son nom l'indique, le séquençage de type RNA-seq se focalise sur les ARN. Dans une cellule, on distingue trois types d'ARN : (i) les ARN messagers qui codent pour des protéines (taille supérieure à 100 nucléotides), (ii) les ARN long non codants comme les ARNr (taille supérieure à 100 nucléotides); et (iii) les petits ARN non codants (ARNnc) comme les microARN ou les snoARN (taille entre 20 et 300 nucléotides). Ces ARN peuvent être séquencés par RNA-seq, mais avec des adaptations particulières dans la préparation des bibliothèques en fonction de leur taille essentiellement.

Après extraction et purification de l'ARN total, il est important d'effectuer des mesures de contrôle qualité afin de vérifier la pureté et l'intégrité de l'échantillon avant de préparer les bibliothèques de séquençage. En effet, les analyses de RNA-seq restent coûteuses et cette étape permet de ne séquencer que les échantillons analysables. Lors de la préparation de l'échantillon, l'ARN peut tout d'abord être contaminé par des produits chimiques (phénol...) ou biologiques (protéine ou ADN génomique). L'évaluation de la pureté d'un échantillon se fait par spectrométrie en ultraviolet visible (UV). L'ARN présente un pic d'absorbance à 260 nm alors que ceux des contaminants sont différents. Par exemple, pour vérifier les contaminations par des protéines qui ont un pic d'absorbance à 280 nm, on utilise le ratio d'absorbance 260 nm/280 nm. Dans une solution d'ARN total pure, on s'attend à avoir un ratio proche de 2. En ce qui concerne l'intégrité de l'échantillon, les conditions de conservation et la présence de ribonucléases peuvent dégrader les molécules d'ARN et ainsi altérer le séquençage. Afin de quantifier la dégradation de l'ARN, on utilise souvent une électrophorèse capillaire en système microfluidique sur puce qui permet d'établir un profil électrophorétique de l'échantillon. À partir de ce profil, un RNA Integrity Number (RIN) est calculé sur la base de la dégradation de l'ARN le plus présent dans la cellule, l'ARNr (80 % des ARN totaux). Le RIN d'un échantillon varie de 1 à 10 : un RIN de 10 correspond à une très bonne intégrité et un RIN de 1 à une dégradation totale. Pour une utilisation classique du RNA-seq, il faudrait un $RIN \geq 7$. Cependant, dans certains cas des RIN plus faibles sont acceptés et

donnent des résultats concluants, comme pour des ARN extraits de tissus fixés au formol et inclus en paraffine (FFPE).

La suite d'une expérience de RNA-seq dépend de la question biologique et de l'ARN d'intérêt (Fig : 3.3). En général, il est courant de s'intéresser à l'expression différentielle des ARNm qui représentent environ 3 % de l'ARN total. Il faut alors éliminer les ARN majoritaires comme les ARNr afin d'éviter la perte d'information. Pour cela il existe deux stratégies communément utilisées. La première consiste à éliminer l'ARN majoritaire qui est l'ARNr (80 % de l'ARN total) pour permettre de séquencer tous les autres ARN minoritaires, y compris les ARNt et les ARNnc. Néanmoins, l'élimination complète de l'ARNr est difficile à atteindre et une partie des lectures de séquençage proviendra des ARNr. La deuxième stratégie consiste à sélectionner uniquement les ARN possédant une queue poly A, caractéristique commune de la plupart des ARNm et caractéristique absente chez les ARNr et les ARNt. Avec cette approche, il faut noter que l'information est perdue pour certains ARNnc qui ne possèdent pas de queue poly A, mais qui jouent néanmoins un rôle important dans l'expression génique d'une cellule, comme les lncRNAs (STATELLO et al., 2020). Dans certains cas, seule l'analyse des petits ARN (snoARN ou miARN) est requise. Cette analyse comprend alors une sélection sur la taille des ARN par l'addition d'une étape de purification sur gel d'agarose. On parle alors de small RNA-seq. Nous allons voir dans la seconde partie de ce chapitre, que pour certaines approches basées sur le RNA-seq et qui s'intéressent en particulier aux ARNr, cette étape d'enrichissement de l'ARN d'intérêt n'est pas nécessaire.

A.2 Séquençage RNA-seq Illumina

Il existe différentes plateformes de séquençage. Actuellement, le séquençage Illumina, basé sur la fluorescence pour lire les bases d'une séquence de nucléotides, est le plus utilisé (jusqu'à 90 % des données de séquençage dans le monde¹).

C'est cette technique qui est utilisée pour cette thèse et que je vais détailler.

1. <https://emea.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html>

Une approche alternative au séquençage Illumina est un séquençage Ion Torrent qui a été commercialisé pour la première fois en 2010 et qui, comme Illumina, utilise un séquençage par synthèse (Fig : 3.1). Cependant, le système de détection utilisé n'est pas optique (fluorophore) mais chimique. La méthode est basée sur la détection d'ions hydrogène (H⁺) libérés lors de l'ajout d'un nucléotide par l'ADN polymérase. Les quatre types de nucléotides (A, C, T ou G) sont rajoutés à des temps différents. La libération des ions H⁺ entraîne un changement de pH qui est détecté par des semi-conducteurs (Fig : 3.1). L'intensité du signal est proportionnelle au nombre de nucléotides incorporés (Fig : 3.1C). Ainsi, il est possible de déterminer la séquence pour chaque fragment contenu dans le puits.

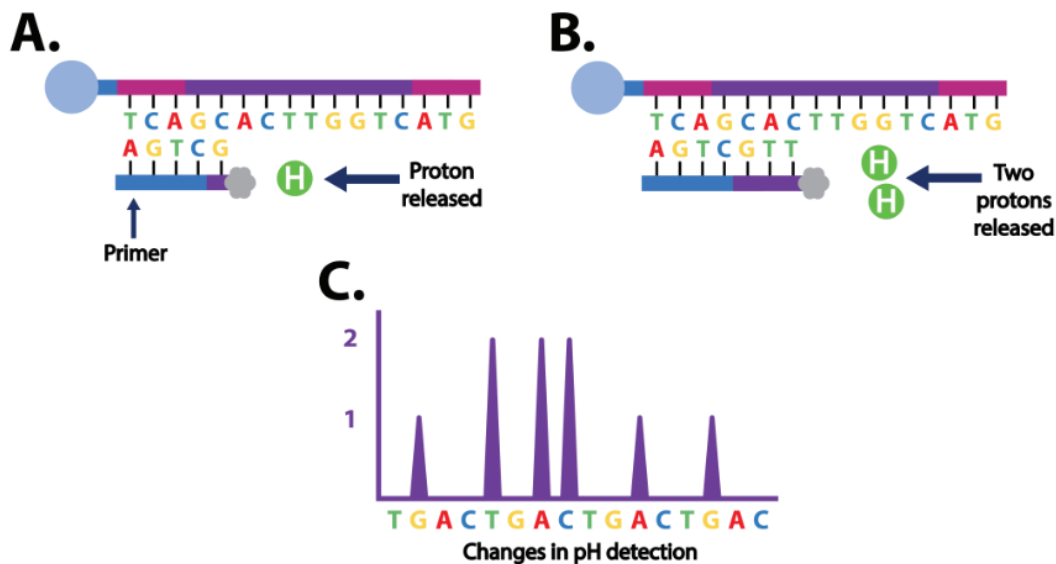


FIGURE 3.1 – Séquençage Ion Torrent : A. Ajout d'un nucléotide de guanine et libération d'un ion H⁺. B. Ajout de deux nucléotide thymines et libération de deux ions H⁺ C. Détection des changements de pH dans l'ordre

Ici, je vais présenter le séquençage Illumina qui implique 3 grandes étapes :

- (i) Préparation de la librairie de séquençage
- (ii) Amplification de la librairie
- (iii) Séquençage

A.2.1 Préparation de la librairie de séquençage

Après la sélection des ARN d'intérêt, les ARN sont fragmentés pour atteindre des tailles usuellement inférieures à 800 nucléotides (Fig : 3.2). Néanmoins, la

taille des fragments dépend du type de séquençage (single-end ou pair-end) et de la question biologique. Dans le séquençage single-end, le séquençage se fait seulement par une des extrémités des fragments d'ARN et une région de la séquence d'ARN ne sera couverte que par une lecture. En revanche, dans le séquençage en pair-end, celui-ci s'effectue par les deux extrémités de l'ARN. L'utilisation du séquençage pair-end est plus onéreux et laborieux comparé au single-end. Néanmoins, il apporte un avantage notamment dans les analyses de l'épissage alternatif, des réarrangements génétiques (insertion, délétion ...), ou d'assemblage de transcriptome de novo lorsque le transcriptome de référence est manquant. Pour conduire une analyse d'expression génique, le séquençage single-end est suffisant, en particulier quand le transcriptome ou le génome de référence est disponible et bien annoté. La fragmentation peut être réalisée par des méthodes différentes : enzymatique (endonucléase), chimique (tampon de Tris-Magnésium) ou mécanique (sonication). Cette étape permet d'éviter la formation de structure secondaire très fréquente chez les ARN et de faciliter l'étape suivante de rétro-transcription. En effet, les fragments d'ARN sont transformés en ADN complémentaire (ADNc) à l'aide d'une enzyme de reverse transcriptase, dont l'efficacité réside dans l'absence de structuration de l'ARN et une taille d'ARN souvent < 1kb. Ensuite, des adaptateurs sont fixés aux extrémités des ARN. Ils sont composés de plusieurs éléments : (i) des séquences positionnées aux extrémités des adaptateurs en 5' et 3' permettant de fixer les fragments à la plaque de séquençage (ou flowcell), (ii) un site de liaison aux amorces de séquençage, nécessaire pour initier l'étape d'élongation, (iii) et d'un barcode ou index qui permet de distinguer les fragments provenant de différents échantillons. On peut ainsi séquencer plusieurs échantillons sur une même flowcell.

A.2.2 Amplification

Grâce aux adaptateurs, les ADNc simples brins issus de la préparation de la librairie vont s'attacher à différents endroits de la flowcell, qui est remplie de petits fragments de nucléotides de séquences complémentaires aux adaptateurs. Ainsi un ADNc est relié par ses deux extrémités à la flowcell et forme un « pont ». Après fixation, une étape d'amplification en pont ou « bridge amplification » commence, étape pendant laquelle de multiples cycles de dénaturation/amplification se suivent (Fig : 3.2). Cette étape a pour but de former un groupe de milliers de molécules de séquences identiques à la

molécule d'origine, appelé « cluster ». La production de ce cluster permet d'avoir un signal suffisamment fort pour être détectable lors du séquençage. Au même moment dans la flowcell, il y a formation de millions de clusters, théoriquement un pour chaque fragment d'ARN d'intérêt.

A.2.3 Séquençage

Les plateformes Illumina utilisent le séquençage par synthèse (c.-à-d., un brin complémentaire est créé à partir de l'ARN d'origine). Ce type de séquençage nécessite des amorces qui se fixent aux adaptateurs via les sites de liaison; une enzyme d'ADN polymérase qui réalise l'élongation sur laquelle repose le séquençage; et des dNTP marqués par fluorescence, chaque nucléotide étant lié à un fluorophore différent afin de les différencier. À chaque cycle de séquençage, un nucléotide marqué est ajouté par l'ADN polymérase dans le brin en cours de synthèse. La présence du fluorophore bloque la réaction de polymérisation, de sorte qu'à la fin de chaque cycle seulement un nucléotide soit incorporé dans le brin en cours de synthèse et que la lecture de la couleur émise par le fluorophore permette d'identifier la base nouvellement incorporée pour en déduire la séquence nucléotidique (Fig : 3.3). Une fois la base identifiée, le fluorophore est clivé à l'aide d'un agent chimique pour permettre l'incorporation du nucléotide suivant. Ce processus est réalisé en parallèle pour chaque cluster du flowcell. La séquence finale obtenue est appelée une lecture (ou read). Le nombre de lectures dépend de la question biologique, de la quantité de matériel biologique de départ et du système de séquençage Illumina utilisé. Certains systèmes comme le système NovaSeq 6000 peuvent générer jusqu'à 6000 Gigaoctet de données avec un nombre maximal de 20 milliards de lectures par flowcell.

A.3 Planification expérimentale

Comme je l'ai mentionné plusieurs fois dans les paragraphes précédents, plusieurs choix doivent être fait dans le cas d'un séquençage RNA-seq, ce qui est vrai pour toutes les technologies omics en général (Fig : 3.3). Ces choix sont essentiels à la réussite d'une analyse de type RNA-seq. Pour cela, il est important de concevoir un bon plan expérimental (CONESA et al., 2016; KOCH et al., 2018; PEREIRA, IMADA et GUEDES, 2017). Le plan expérimental est guidé par les questions biologiques. Ces questions biologiques vont

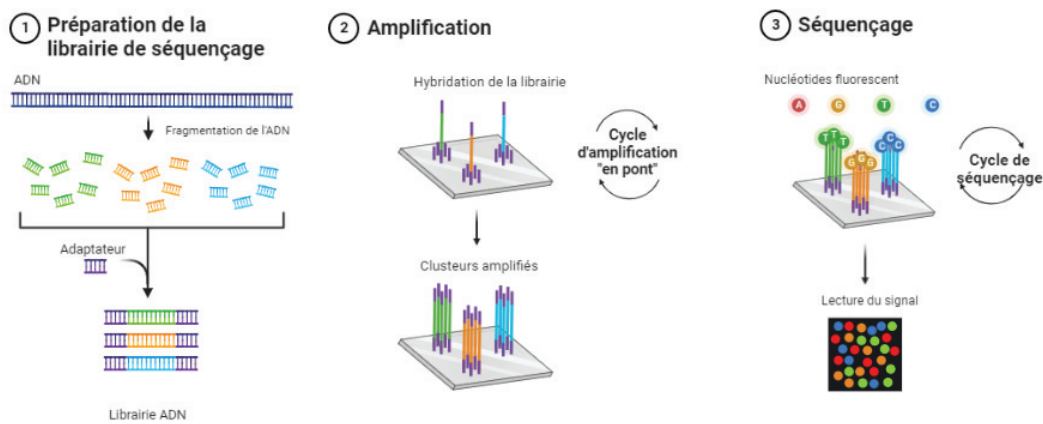


FIGURE 3.2 – Principe du séquençage Illumina. Schéma général des étapes de préparation des bibliothèques de séquençage Illumina qui sont décrites dans le texte. (Adapté de www.biorender.com)

influencer la quantité d'ARN de départ, le nombre et type de réplicats techniques ou biologiques, la préparation des bibliothèques, la couverture et la profondeur ou encore la longueur des lectures. Une analyse plutôt qualitative (annotation de transcriptome ou de sites méthylés) diffère d'une analyse quantitative (identifier les gènes différentiellement exprimés ou les sites différentiellement méthylés). Certains paramètres ne dépendent pas des questions biologiques comme le coût, le temps et la disponibilité des échantillons. Des outils comme RNASeqPower (HART et al., 2013) ou RnaSeqSampleSize (SHILIN ZHAO AND CHUNG-I LI AND YAN GUO AND QUANHU SHENG AND YU SHYR, 2020) permettent de concevoir un plan expérimental en prenant en compte plusieurs de ces paramètres. La capacité à bien interpréter les résultats dépend de la planification expérimentale, de la présence de contrôles et de l'utilisation appropriée des outils bioinformatiques et d'analyses de données.

Un facteur important à prendre en compte lors de la conception du plan expérimental est l'effet batch surtout quand l'analyse comprend un grand nombre d'échantillons qui vont être traités dans des lots différents. L'effet batch est une source de variation d'origine technique qui peut avoir diverses origines comme des temps ou lieux différents d'expérimentation, des lots de réactifs différents ou même des manipulateurs différents. Le plan expérimental peut aider à diminuer et mieux contrôler l'effet batch en incluant des contrôles, une gestion et un traitement aléatoire des échantillons. Cependant,

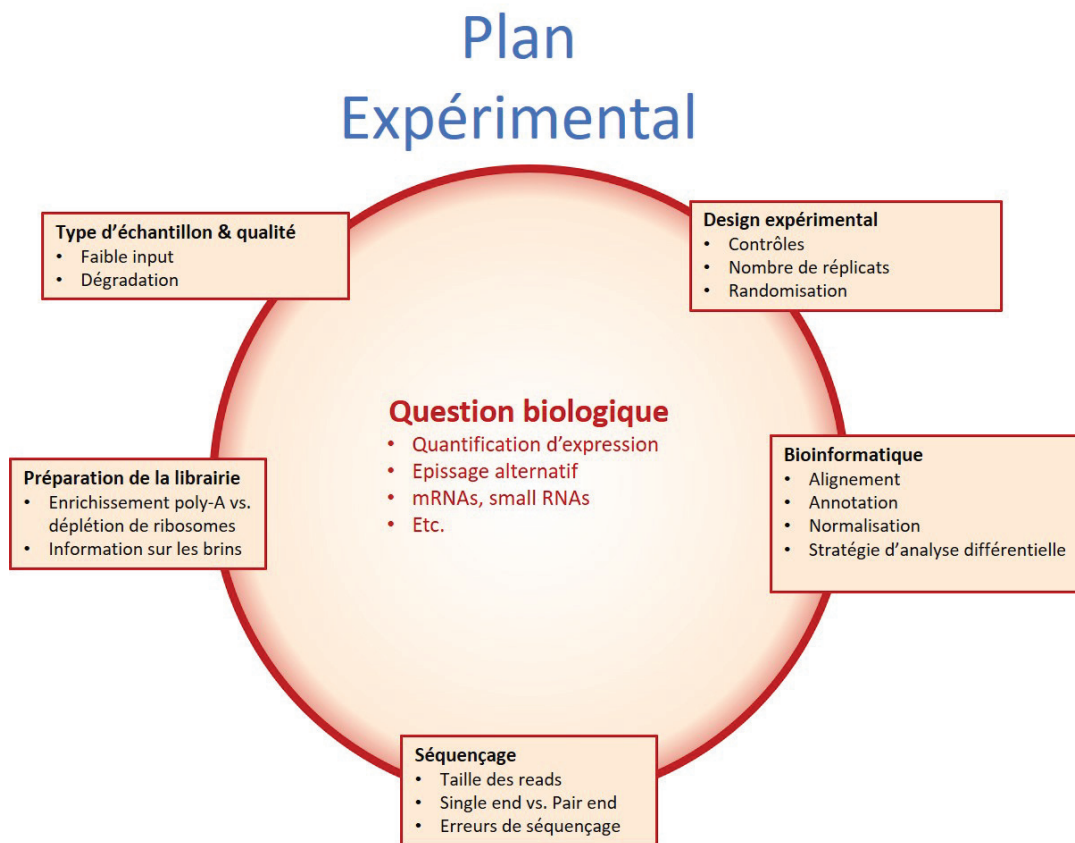


FIGURE 3.3 – Plan expérimental. Chaque étape d'une analyse OMICS dépend de la question biologique d'intérêt. Inspiré de Friederike Dündar²

il est important de noter qu'en pratique, dans les données omics les effets batch sont presque inévitables (GOH, W. WANG et WONG, 2017). Ces effets vont influencer les analyses en aval. Par exemple certains gènes identifiés comme différentiellement exprimés entre les conditions pourraient être issus de différences techniques plutôt que biologiques (faux positifs). À l'inverse certains gènes qui sont réellement différentiellement exprimés pourraient ne pas être détectés (faux négatifs). Il est donc nécessaire de pouvoir ajuster cet effet batch.

Il existe plusieurs algorithmes d'ajustement des effets batch (BECAs pour Batch Effect Correction Algorithms). Un des plus populaires est ComBat (JOHNSON, C. LI et RABINOVIC, 2006) utilisé lorsque la source de l'effet batch est connue. Pour l'effet batch de source inconnu, SVASEq (LEEK, 2014) et RUVSeq (RISSO et al., 2014) sont couramment utilisés. Lors d'une analyse différentielle réalisée avec EdgeR (ROBINSON, MCCARTHY et SMYTH, 2009) ou DESeq2 (MICHAEL

LOVE, 2017), qui ont pour but d'identifier les gènes différentiellement exprimés, il est possible de tenir compte de l'effet batch en incluant la variable responsable de l'effet batch dans le modèle linéaire utilisé par ces méthodes.

Plus récemment, Zhang et collègues ont proposé une amélioration de l'algorithme original de ComBat appelé ComBatSeq (J. ZHANG et al., 2020). ComBatSeq utilise un modèle de régression binomial négatif qui permet de mieux modéliser les données de comptage RNA-seq contrairement à l'algorithme de ComBat qui assume une distribution gaussienne des données. De plus, les données corrigées de ComBatSeq sont rendues sous la forme de comptage et conserve ainsi la nature de nombre entier des données.

A.4 Analyse des données de séquençage RNA-seq

L'analyse de données RNA-seq présente beaucoup de variantes qui dépendent des objectifs et des questions biologiques. Je vais décrire brièvement les trois étapes principales qui sont communes à tout type d'analyse : le contrôle qualité, l'alignement et la quantification. Je finirai par quelques approches permettant d'interpréter les résultats.

A.4.1 Contrôle qualité

La première étape à réaliser après l'obtention des données brutes du séquenceur est le contrôle qualité. Cette étape est très importante et permet tout d'abord de contrôler la qualité des données brutes. L'outil le plus couramment utilisé est FastQC (ANDREWS et al., 2012) (associé à MultiQC (EWELS et al., 2016)) qui permet d'obtenir un rapport sur la qualité de la totalité des données pour chaque échantillon. Plusieurs paramètres sont calculés et visualisés comme la qualité moyenne des lectures, la qualité par base, le pourcentage moyen en GC, la distribution de la longueur des lectures ou encore les séquences sur-représentées dans l'échantillon. La deuxième étape concerne la filtration des données. La présence des adaptateurs par exemple peut empêcher une lecture de s'aligner et va induire une perte de l'information. Il est donc important d'éliminer les lectures ou les bases de mauvaise qualité et de tronquer les adaptateurs. Des outils comme trimmomatic (BOLGER, LOHSE et USADEL, 2014) ou cutadapt (MARTIN, 2011) sont souvent utilisés pour réaliser cette étape.

A.4.2 Alignement et quantification

Après avoir sélectionnée les données de bonne qualité, il est important de connaître la provenance de chaque lecture séquencée, c'est-à-dire l'ARN d'intérêt que reflète cette lecture. C'est l'objectif de l'étape d'alignement. Pour cela il est nécessaire d'avoir un génome ou un transcriptome de référence et un outil d'alignement. Certaines lectures peuvent provenir d'une jonction exon-exon, ce qui signifie qu'une partie de la lecture provient d'un premier exon d'un ARNm et que l'autre partie provient d'un second exon. Quand l'alignement est fait sur un génome de référence, il est donc important de choisir un outil d'alignement tel que STAR (DOBIN et al., 2012) capable d'aligner une lecture sur deux exons différents, qui peuvent être relativement éloignés dans le génome. L'alignement sur un transcriptome de référence peut se faire avec des outils comme bwa (H. LI et DURBIN, 2009) ou bowtie (LANGMEAD et al., 2009) qui ne prennent pas en compte les jonctions exon-exon. Les lectures s'alignent alors soit à une position unique sur la référence soit à plusieurs positions. Ces dernières sont dus à des séquences répétées dans la référence ou à des séquences partagées entre des gènes paralogues.

Après l'étape d'alignement, il est courant de vouloir estimer l'expression des gènes. Une première approche directe est de compter le nombre de lectures alignées sur chaque gène à l'aide d'outils comme HTSeq-count (ANDERS, PYL et HUBER, 2014) ou featureCounts (LIAO, SMYTH et W. SHI, 2013). Les comptages bruts ne peuvent pas être comparés entre les échantillons, car ils sont affectés par la profondeur (le nombre de lectures par échantillon) et les biais du séquençage. La longueur du gène doit être prise en compte pour comparer les gènes de l'échantillon entre eux (les gènes les plus longs auront plus de lectures alignées). La métrique la plus utilisée et la plus adaptée dans ce cas de figure est le TPM (Transcripts per Millions) qui prend en compte la longueur des gènes et la profondeur de séquençage (WAGNER, KIN et LYNCH, 2012). Les TPM ont une somme d'un million pour chaque échantillon, ce qui facilite les comparaisons.

Des outils comme DESeq2 (MICHAEL LOVE, 2017) ou edgeR (ROBINSON, MCCARTHY et SMYTH, 2009) ont leurs propres méthodes de normalisation des comptages brutes et sont recommandés pour comparer les niveaux d'expression entre les échantillons.

A.4.3 Aspect analytique pour l'interprétation des données de RNAseq

Un des objectifs principaux d'une analyse RNA-seq est d'identifier les gènes qui présentent une expression différente entre deux conditions. C'est l'objectif de l'analyse différentielle qui donne lieu à une comparaison statistique de l'expression de chaque gène entre deux conditions. Des analyses plus poussées ont pour but d'identifier un contexte biologique plus large et non restreint à un seul gène, mais à une fonction biologique portée par plusieurs gènes. Par exemple, l'analyse des voies (pathway analysis) a pour but de comparer les gènes issus de l'analyse différentielle avec des listes de gènes prédéfinis qui sont annotées en fonction d'un processus biologique ou d'une maladie. D'autres types d'analyses permettent d'analyser l'épissage alternatif ou la fusion des gènes dans le cas d'une maladie comme le cancer (PEREIRA, IMADA et GUEDES, 2017). La plupart des expériences RNA-seq s'arrêtent à ce niveau. Des méthodes bioinformatiques plus récentes ont été développées pour capter encore plus d'informations dans les données RNA-seq comme l'altération du nombre de copies, l'identification de contamination microbienne ou le type cellulaire (déconvolution) (THIND et al., 2021). De plus, il existe différentes variantes du protocole RNA-seq classique qui ont été adaptées pour répondre à des questions biologiques spécifiques (CELLERINO et SANGUANINI, 2018). Voici quelques exemples aujourd'hui utilisés dans le domaine de la régulation traductionnelle et du ribosome :

1. Le small RNA-seq permet d'analyser l'expression de petits ARNnc comme les snoARN ou les miARN en sélectionnant avant le séquençage, les ARN de petite taille.
2. Le « ribosome profiling » ou Ribo-seq permet d'analyser les ARNm en cours de traduction (ou translatome). Cette technique permet de séquencer les régions des ARNm qui sont protégés par les ribosomes et donc déterminer leurs niveaux de traduction, mais aussi des événements de traductions particuliers (codon initiateur, pause des ribosomes, défaut de lectures des codons stop ou readthrough...).
3. Le RIP-seq (RNA-immunoprecipitation sequencing) est une autre variante du RNA-seq qui permet d'étudier les interactions entre les ARN et une protéine d'intérêt. L'approche utilise un anticorps spécifique à la protéine d'intérêt, ce qui permet d'immunoprécipiter cette protéine et

tous les ARN associés à cette protéine. Ces ARN sont ensuite purifiés puis identifiés par RNA-seq.

4. RNA-seq dépendant des propriétés chimiques de l'ARN. Ce type d'approche permet d'étudier, en plus de l'expression génique, d'autres mécanismes de biologie moléculaire. Le CAGE-seq, par exemple, permet d'étudier l'expression génique et d'identifier les sites d'initiation de la transcription. D'autres techniques utilisent des protocoles spécialisés qui reposent sur les propriétés physico-chimiques des nucléotides des ARN ayant subi des modifications post-transcriptionnelles. Ces techniques permettent d'identifier et même de quantifier les variations du niveau de modifications post-transcriptionnelles des ARN comme les méthylations des bases (m6A-seq, m1A-seq...), les pseudouridilations (HydraPsiSeq, Psi-seq...) ou les 2'-O-ribose méthylation (Nm-seq, RiboMethSeq...).

Dans la seconde partie de ce chapitre, je vais décrire les techniques de séquençage haut-débit basées sur le RNA-seq qui permettent d'étudier la 2'Ome des ARNr.

B Détection et quantification de la 2'-O-ribose méthylation des ARN

La 2'Ome apporte à l'ARN des propriétés physico-chimiques particulières qui le distingue d'un ARN non 2'Ome (KROGH et NIELSEN, 2019; MOTORIN et MARCHAND, 2018). Ces propriétés ont été exploitées dès les années 1960 pour détecter la présence de 2'Ome, et ce, de manière site-à-site. Au cours des sept dernières années, ces techniques ont été conjuguées avec les technologies de séquençage haut-débit, pour permettre la détection et la quantification de la 2'Ome sur l'ensemble des sites en un seul séquençage. Ces techniques de séquençage sont en principe utilisables pour tous les ARN, mais leur preuve de concept a été réalisée sur l'ARN le plus abondant dans la cellule, l'ARNr. Voici les quatre propriétés physico-chimiques utilisées pour la détection de la 2'Ome :

- (i) La présence du groupement CH_3 en position 2' du ribose réduit considérablement la réactivité de l'oxygène sur lequel il est fixé. Cela empêche le clivage de la liaison phosphodiester entre deux nucléotides

voisins, par exemple lors d'une hydrolyse alcaline ou d'un clivage par des nucléases (RNase T2 et RNase H) (B. LANE et TAMAOKI, 1967; MADEN, 2001).

- (ii) Dans certains cas, lorsqu'un nucléotide 2'Ome est à l'extrémité 3' de l'ARN, le groupement CH₃, volumineux et hydrophobe, empêche la réaction avec des agents oxydants bidentés (ex : IO₄⁻) et protège ainsi le ribose terminal d'une dégradation (MOTORIN et MARCHAND, 2018).
- (iii) Le groupement CH₃ stabilise l'ARN en donnant une conformation particulière au ribose appelée C3'-endo qui favorise l'appariement des bases et le repliement de l'ARN (PRUSINER, YATHINDRA et SUNDARALINGAM, 1974).
- (iv) Comme dit précédemment, la 2'Ome n'empêche pas l'appariement des bases, mais la présence du groupement CH₃, qui est plus volumineux que l'atome d'hydrogène H, crée un encombrement stérique. Cet encombrement peut empêcher la transcriptase inverse d'avancer lors de l'extension de l'amorce, particulièrement en présence de faible concentration de désoxynucléotides triphosphates (dNTPs) (MADEN, 2001).

Dans cette partie, je vais décrire les techniques les plus utilisées à l'heure actuelle en me focalisant en particulier sur le RiboMethSeq qui est au cœur de mon projet de thèse.

B.1 Techniques basées sur la propriété d'encombrement stérique induite par la présence d'une 2'Ome (propriété iv)

B.1.1 Primer extension (extension d'amorces)

La technique historique utilisée dans le domaine de l'ARNr pour détecter la 2'Ome est la technique de primer extension (ou extension d'amorce) (Fig : 3.4A). À l'origine, elle a été développée pour séquencer et identifier correctement les extrémités 5' de l'ARN. Dans le monde du ribosome, elle est aussi utilisée pour détecter les clivages et les modifications chimiques de l'ARNr. Dans le cas d'une utilisation pour détecter la 2'Ome, le principe de la technique est basé sur l'encombrement stérique induit par le groupement CH₃ sur l'avancée de la transcriptase inverse (ou RT pour reverse transcriptase) en condition de faible concentration de dNTPs (propriété iv). Très brièvement, l'amorce anti-sens (c.-à-d., complémentaire à l'ARN cible) étant marquée par

des éléments radioactifs ou fluorescents en 5', s'hybride à la séquence d'intérêt. Ensuite, la réaction de RT est réalisée en présence, soit d'une forte, soit d'une faibles concentrations de dNTPs. En effet, la transcription inverse n'est pas impactée par la présence de nucléotide 2'Ome sauf à de faible concentration de dNTP qui cause un arrêt de la RT (D. J. LANE et al., 1985; MADEN, 2001). En présence d'une forte concentration de dNTPs, l'enzyme RT génère des fragments d'ADNc jusqu'à ce que la RT atteigne l'extrémité 5' de l'ARN. En présence d'une faible concentration de dNTP, l'enzyme RT génère des fragments d'ADNc jusqu'à ce que la RT atteigne une 2'Ome. Dans ce cas, l'extrémité 5' des fragments générés va représenter la position des nucléotides 2'Ome sur l'ARN (n-1). Les fragments d'ADNc générés sont visualisés par imagerie fluorescente ou autoradiographie après une résolution sur un gel de séquençage, grâce à l'amorce anti-sens marquée (BIRKEDAL, KROGH et al., 2016).

En parallèle de cette RT, on réalise un séquençage Sanger de l'ARN cible. Ainsi, en comparant l'emplacement sur le gel du signal des fragments d'ADNc générés sous différentes concentrations de dNTPs et par séquençage, il est possible de connaître la position du nucléotide exact qui a causé l'arrêt de la RT. Lors d'un arrêt avant l'extrémité 5' de l'ARN et en condition de faible concentration de dNTPs, c'est indicateur d'un nucléotide portant une 2'Ome.

Il est possible de détecter plusieurs nucléotides 2'Ome par cette technique si la distance qui les sépare n'est pas importante. La technique d'extension d'amorce est très sensible et est souvent utilisée pour détecter de nouveaux sites de 2'Ome. Cependant, elle peut générer des signaux dits faux positifs (c.-à-d., détection d'une 2'Ome inexistante) ou faux négatifs (c.-à-d., absence de détection d'une 2'Ome existante). Cela peut être causé par la présence d'une structure secondaire de l'ARN qui induit un arrêt de la RT mais aussi par la présence d'autres types de modifications chimiques de l'ARN comme les méthylations de bases m6A ou m1acp3 qui empêchent le bon appariement des bases et l'arrêt de la RT. De plus, l'approche est laborieuse expérimentalement, nécessite beaucoup de matériel biologique, prend beaucoup de temps et ne permet pas d'avoir une vue globale de la 2'Ome, cette technique ne permettant d'analyser que 1 à 3 sites à la fois. De plus, la quantification du niveau de la 2'Ome reste difficile par cette technique, même couplée à une amplification PCR en temps réel (RT-qPCR). C'est d'ailleurs cette approche

de primer extension couplée à une RT-qPCR qui a permis de réaliser les premières analyses de quantification relative du niveau de 2'Ome des ARNr entre deux conditions biologiques, mais pour une vingtaine de sites seulement (BELIN et al., 2009).

B.1.2 2Ome-seq

L'approche de 2Ome-seq a été publiée par Incarnato *et al* en 2017 (INCARNATO et al., 2016). Cette technique est basée sur le principe de l'extension d'amorce (c.-à-d, arrêt de la reverse transcription sur les nucléotides 2'Ome sous condition de concentration limitante de dNTP) couplée à un séquençage (Fig : 3.4C). Les bibliothèques de séquençages Illumina sont préparées à partir des fragments d'ADNc issus d'extension d'amorce aléatoire réalisée sous des concentrations de dNTPs élevées (1 mM) et faibles (0,004 mM). Le séquençage Illumina single-end peut être utilisé. Les lectures issues du séquençage suivent un traitement bio-informatique spécifique comprenant : l'étape de QC (élimination des adaptateurs et des lectures de mauvaise qualité) ; l'alignement sur un ARNr de référence ; et le comptage des extrémités 5' des lectures alignées pour chaque position de l'ARNr de référence.

À partir de ces données, les auteurs ont utilisé deux mesures différentes pour analyser les sites de 2'Ome de l'ARNr. Le premier, le score de 2'Ome, reflète un enrichissement des extrémités 5' des lectures séquencées dans la bibliothèque à faible concentration de dNTPs par rapport à celles provenant de la bibliothèque à forte concentration de dNTPs pour chaque position de l'ARNr. Cela permet de détecter les sites de 2'Ome et d'éliminer les faux positifs qui peuvent être dus à une forte structuration de l'ARNr induisant un arrêt de la transcription inverse. La deuxième mesure, le ratio de 2'Ome, représente la fraction des extrémités 5' des lectures, issues de l'échantillon à faible concentration de dNTPs, à une position par rapport au total des lectures couvrant cette position. Cette seconde mesure se rapproche d'une quantification du niveau de 2'Ome des ARNr.

Cette méthode de 2Ome-seq est très performante pour détecter des sites de 2'Ome (KROGH et NIELSEN, 2019). Cependant, en termes de quantification du niveau de 2'Ome, la méthode a montré des résultats différents des autres techniques (c.-à-d. RiboMethSeq, Nm-seq ou RibOxi-seq – sections

suivantes). En effet, non seulement la métrique utilisée (ratio de 2'Ome) varie beaucoup plus que les métriques équivalentes issues des autres méthodes disponibles et répertoriées ci-dessous, mais en plus une majorité des sites ont des valeurs de 2'Ome faibles. Par exemple, l'analyse de la 2'Ome des ARNr de la lignée cellulaire cancéreuse HeLa donne un ratio de 2'Ome moyen de 0.292 [max : 0.767, min : 0.005] pour l'ARNr 18S, et de 0.175 [max : 0.69, min : 0] pour l'ARNr 28S. Avec les autres techniques, la plupart des sites des ARNr 18S et 28S sont presque complètement méthylés avec un score proche de 1. Il est ainsi difficile de comparer les ratios de 2'Ome issus du 2Ome-seq aux autres techniques. Par ailleurs, comme pour le RibOxi-seq et le Nm-seq décrits ci-dessous, la détection de la 2'Ome de deux nucléotides proches est réduite par 2Ome-seq. Enfin, en termes de quantité de matériel biologique nécessaire pour l'analyse d'un échantillon, il faut extraire $2 \times 2 \mu\text{g}$ d'ARN total (soit $4 \mu\text{g}$), ce qui restreint les domaines d'application de la méthode.

B.2 Techniques basées sur la propriété de résistance à l'oxydation et à la β -élimination induit par la présence d'une 2'Ome (propriété ii)

Après fragmentation aléatoire des ARN totaux, les fragments générés sont soumis à une oxydation/ β -élimination (Fig : 3.3D). En effet, les nucléotides 2'Ome qui se retrouvent en extrémité 3' d'une molécule d'ARN fragmentée sont résistants à l'oxydation et à la beta-élimination (propriété ii), et conservent leurs hydroxyles en 3'. En revanche, les nucléotides non 2'Ome en extrémité 3' d'une molécule d'ARN sont oxydés par le périodate de sodium (NaIO_4) qui transforme ainsi les cis-diols du ribose (hydroxyles en 2' et 3') en dialdéhydes. Cette réaction permet (1) d'éliminer le nucléotide dialdéhydé par beta-élimination et (2) empêche la ligation des adaptateurs nécessaires pour la préparation des librairies de séquençage. Finalement, cette approche permet un enrichissement des fragments d'ARN ayant en extrémité 3' un nucléotide 2'Ome.

Ce principe est utilisé par deux techniques qui ont été publiées au même moment par deux groupes différents : le RibOxi-seq (ZHU, PIRNIE et CARMICHAEL, 2017) et le Nm-seq (DAI et al., 2017) :

- (i) Le RibOxi-seq utilise l'endonucléase de benzonase pour fragmenter l'ARN

de façon aléatoire. Cela génère des fragments d'ARN ayant des extrémités 3' avec des riboses hydroxylés en 2' et 3', qui sont ensuite soumis à une étape d'oxydation et de b-élimination de la base présente en extrémité 3'. Parmi les fragments générés, il y aura 2 cas de figure. S'il n'y a pas de 2'Ome au nucléotide précédent l'extrémité 3' du fragment, l'étape d'oxydation/ β -élimination génère des fragments d'ARN sans hydroxyle en 3'. Si une 2'Ome est présente sur le nucléotide précédent l'extrémité 3' du fragment, l'étape d'oxydation/b-élimination n'impactera pas les fragments d'ARN qui conservent alors leurs hydroxyles en 3'. En effet, la 2'Ome protège la liaison phosphodiester située en 3' du nucléotide 2'Ome. Lors de la préparation de la librairie, seuls les fragments ayant un hydroxyle en extrémité 3' (reflétant la présence d'une 2'Ome) sont disponibles pour la ligation des adaptateurs nécessaires pour la construction des librairies de séquençage. Il y aura donc un enrichissement en fragments d'ARN qui portaient une 2'Ome en extrémité 3'. Après le séquençage, les lectures sont alignées sur une séquence d'ARNr de référence. L'extrémité 3' des lectures alignées est comptée pour chaque nucléotide de l'ARNr de référence. En parallèle, des échantillons contrôles n'ayant pas subis d'oxydation sont séquencés. Les données de comptage des extrémités 3' des lectures provenant des échantillons oxydés et non-oxydés sont ensuite normalisées, comparées et analysées comme des données de comptage de gènes RNA-seq à l'aide de l'outil DESeq2. Cela permet de détecter les sites méthylés en utilisant les métriques de « log2 fold change » et de « p-valeur ».

- (ii) Le Nm-seq repose sur le principe utilisé pour le RibOxi-seq avec néanmoins quelques différences. La première différence intervient dans la fragmentation de l'ARN. Le Nm-seq utilise une fragmentation chimique aléatoire de l'ARN, au lieu d'une fragmentation enzymatique comme le RibOxi-seq. L'étape de fragmentation est ensuite suivie, comme le RibOxi-seq, d'une étape d'oxydation/b-élimination. Mais contrairement au RibOxi-seq, cette étape n'est pas réalisée une seule fois, mais plusieurs fois afin d'exposer les nucléotides 2'Ome présents en 3' de l'ensemble des fragments générés. Ainsi, dès qu'un nucléotide 2'Ome est rencontré, le processus d'élimination des nucléotides s'arrête car, comme expliqué précédemment, l'absence des cis-diols de ribose disponibles

(CH₃ en place de l'hydroxyle en 2') empêche l'oxydation du nucléotide en dialdéhydes. Ces étapes aboutissent aussi à un enrichissement en fragments ayant un nucléotide 2'Ome en extrémité 3'. La méthode d'analyse des données de séquençage est similaire à celle utilisée pour le RibOxi-seq.

Ces deux techniques permettent une détection sensible et précise des nucléotides 2'Ome et sont donc efficaces pour identifier des sites 2'Ome. Cependant, plusieurs limites existent, en particulier pour l'ARNr :

- (i) Comme nous l'avons vu dans le premier et deuxième chapitre de cette thèse, certains sites 2'Ome présentent une intra-variabilité. Ce qui signifie que dans la population d'ARNr étudiée, il existe des molécules d'ARNr avec une 2'Ome et d'autres molécules d'ARNr sans 2'Ome à un site donné. Or, les techniques de RibOxi-seq et Nm-seq ne rendent compte que de la présence d'une 2'Ome. Ainsi, ces deux techniques ne fournissent pas d'information quantitative (c.-à-d., niveau de 2'Ome à un site donné) mais seulement une information qualitative (détection ou non d'un site 2'Ome).
- (ii) Il faut noter que différentes techniques ont été largement utilisées jusqu'à aujourd'hui pour cartographier la quasi-totalité des sites 2'Ome des ARNr chez l'Homme, la levure et la bactérie. L'ensemble de ces techniques de biochimie (primer extension, spectrométrie de masse...) a permis d'identifier un nombre de sites 2'Ome qui varie peu (entre 106 et 113), suggérant que peu de nouveaux sites sont encore à découvrir chez ces organismes. Cela montre que l'utilisation du RibOxi-seq ou du Nm-seq pour l'analyse de la 2'Ome de l'ARNr humain semble être peu enrichissante. En effet, même si certains sites encore inconnus peuvent être présents seulement dans certaines conditions spécifiques, d'autres techniques peu onéreuses d'identification de nouveau site peuvent être utilisées comme la primer extension. Cependant, les deux techniques peuvent être utiles lors de la cartographie des sites 2'Ome pour des ARNr encore peu étudiés, comme chez la souris, ou pour des ARN de nature différentes, comme les ARNm, ARNt ou ARN non-codants.
- (iii) Une autre limite est la proximité de deux sites méthylés. Dans ce cas, seulement un site sur les deux sera détecté. Par exemple, le site 28S_Um4226 ne sera pas détecté si le site suivant 28S_Um4227 est méthylé. En effet,

c'est le dernier nucléotide 2'Ome en extrémité 3' du fragment qui sera révélé par cette approche d'oxydation/b-élimination. Cela peut empêcher d'identifier des sites ayant un intérêt biologique.

- (iv) La quantité d'ARN nécessaire pour réaliser ces analyses semble aussi être une limite non négligeable. En effet, la quantité de matériel biologique nécessaire peut-être un élément très limitant dans certains cas, comme c'est le cas pour des projets en lien avec la santé humaine. Le matériel provenant des échantillons humains est précieux et souvent limité en quantité. Les auteurs de RibOxi-seq et Nm-seq ont utilisé respectivement 7,5 µg et 10 µg d'ARN total pour analyser la 2'Ome de l'ARNr, qui pour rappel représente jusqu'à 80 % de l'ARN total. Ces quantités ne sont pas compatibles avec des applications dans le domaine de la santé humaine.

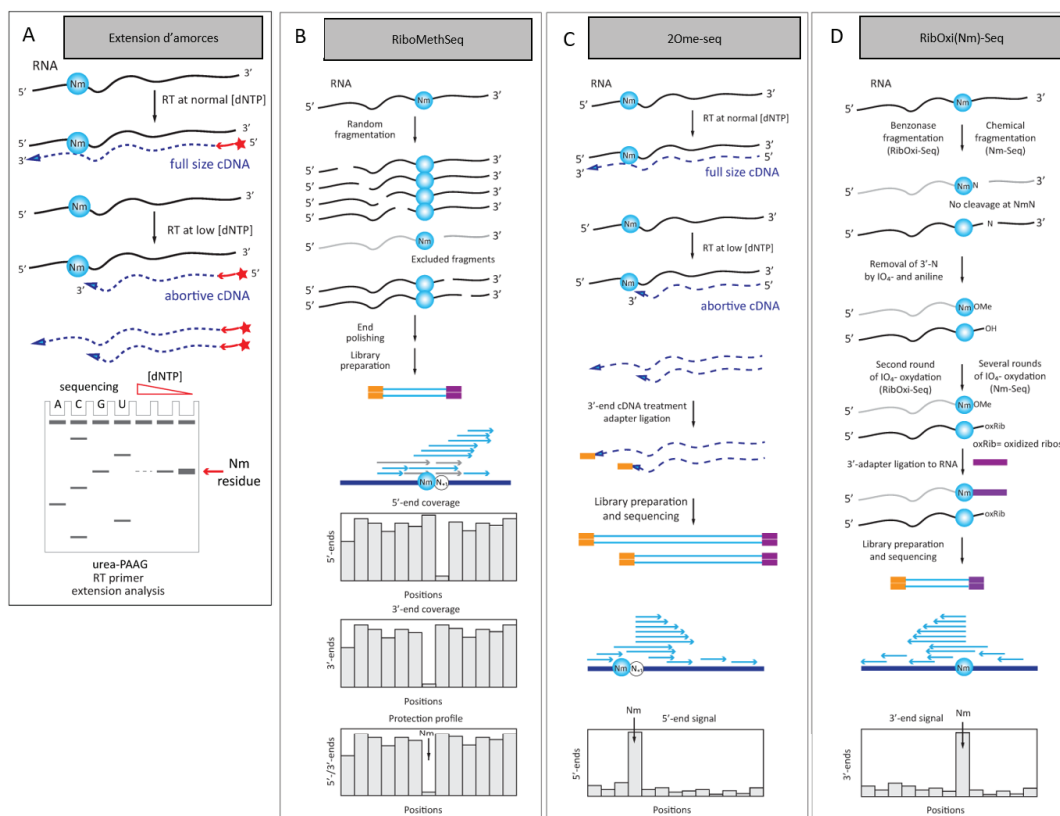


FIGURE 3.4 – Représentation schématique de quatre méthodes (A. Primer extension, B. RiboMethSeq, C. 2Ome-seq, D. RibOxi(Nm)-seq) utilisées pour la détection et la quantification de la 2'Ome décrites dans le texte. (MOTORIN et MARCHAND, 2018)

B.3 Technique basée sur la propriété de résistance à l'hydrolyse alcaline induite par la présence d'une 2'Ome (propriété i) : le RiboMethSeq

B.3.1 Principe du RiboMethSeq

Le RiboMethSeq est une autre technologie de séquençage haut débit disponible permettant de mesurer le taux de 2'Ome sur la totalité des sites de l'ARN en un seul séquençage (Fig : 3.4B, 3.5). La méthode est basée sur le principe que la présence d'un nucléotide 2'Ome rend la liaison phosphodiester avec le nucléotide voisin en 3' résistante au clivage dans des conditions alcalines (propriété i). Après extraction, l'ARN est ainsi fragmenté par une hydrolyse alcaline partielle. Puisque la liaison phosphodiester située en 3' d'un nucléotide i 2'Ome est protégée de l'hydrolyse alcaline, les fragments générés ne peuvent ni (i) finir en 3' par le nucléotide i et (ii) commencer en 5' par le nucléotide $i+1$. Ainsi, ces fragments commençant en position $i+1$ ou finissant en position i seront moins présents dans la population d'ARN fragmentés.

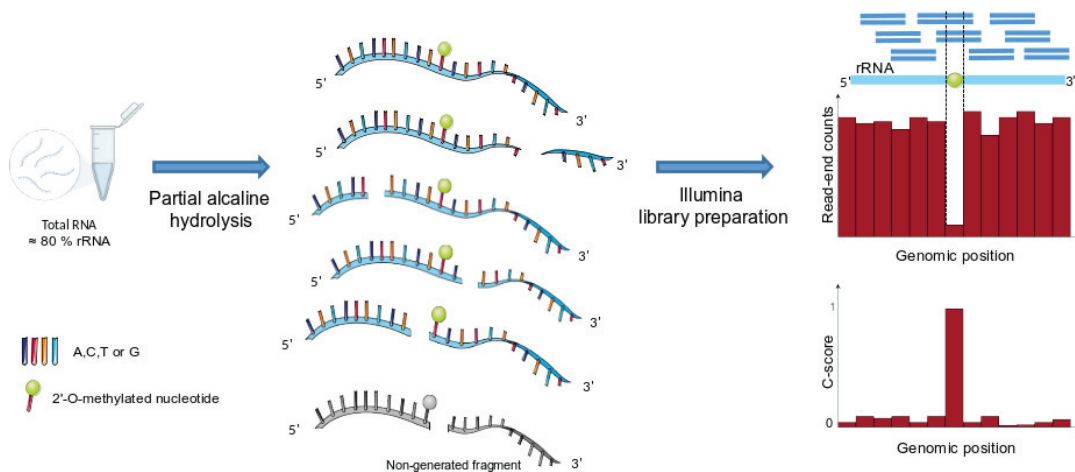


FIGURE 3.5 – Principe du RiboMethSeq. L'ARN total est soumis à une hydrolyse alcaline partielle. Le nucléotide 2'Ome rend la liaison phosphodiester avec le nucléotide voisin en 3' résistante au clivage. Les fragments générés ne peuvent ni (i) finir en 3' par le nucléotide 2'Ome i ni commencer en 5' par le nucléotide 2'Ome $i+1$. Après traitement bioinformatique, les comptages des extrémités 5' ou 3' sont transformés en taux de 2'Ome appelé Cscore.

La librairie de séquençage est construite à partir des fragments issus de l'hydrolyse alcaline par ajout des adaptateurs nécessaires. Le séquençage est ensuite réalisé en single ou pair-end. Les lectures obtenues sont nettoyées, alignées sur l'ARN de référence et les extrémités 5' et 3' des lectures sont comptées pour chaque position de l'ARN. Enfin, un taux de 2'Ome est calculé en normalisant les comptages d'une position donnée par rapport aux comptages des positions voisines. Pour avoir une analyse robuste du niveau de 2'Ome, il est nécessaire d'avoir une couverture de lecture élevée. Par conséquent, l'ARN cible doit être très abondant dans la population d'ARN total, ce qui est le cas pour l'ARNr qui représente jusqu'à 80 % de l'ARN total. Dans la littérature, il existe deux protocoles de RiboMethSeq distincts qui utilisent l'hydrolyse alcaline partielle pour fragmenter les ARN, mais des méthodes de séquençage haut débit différentes pour quantifier la 2'Ome des ARNr.

B.3.2 RiboMethSeq et technologie Ion Torrent

Le premier protocole de RiboMethSeq a été publié en 2015 par Birkedal et collègues, une équipe danoise (BIRKEDAL, CHRISTENSEN-DALSGAARD et al., 2014). La méthode décrite précédemment présente quelques différences, car le séquençage des lectures repose sur un séquençage par semi-conducteur ionique de type Ion Torrent, mais aussi sur des enzymes peu disponibles dans le commerce. La technologie Ion Torrent étant peu répandue et les réactifs peu accessibles, ce protocole est aujourd'hui peu utilisé. Ce premier protocole comprend plusieurs étapes. Les fragments d'ARN issus de l'hydrolyse alcaline sont tout d'abord purifiés sur gel pour ne conserver que les fragments de 20 à 40 nucléotides. Ces fragments sont ensuite ligaturés à des adaptateurs (Fig : 3.4B). La particularité de ce protocole est que les adaptateurs liant l'extrémité 5' des fragments sont produits par transcription in vitro au laboratoire en utilisant le ribozyme HDV (Hepatitis Delta Virus). Par ailleurs, la ligation entre les adaptateurs et les fragments d'ARN est réalisée par une ARN ligase mutée. L'utilisation de ce type de ligase nécessite des quantités importantes de matériel biologique (1 à 10 µg). Après le séquençage, une analyse bioinformatique est réalisée. Les adaptateurs sont tronqués et les lectures de longueur inférieure à 15 nucléotides sont éliminées à l'aide du logiciel Cutadapt v1.2 (MARTIN, 2011). Le reste des lectures est aligné sur une séquence d'ARNr de référence à l'aide de l'outil bowtie2, qui est approprié pour l'alignement des lectures courtes. Les lectures alignées seulement

une fois sur la séquence de l'ARNr de référence sont prises en compte pour la suite de l'analyse. Pour chaque position de l'ARNr, le nombre de lectures qui commencent en 5' ou finissent en 3' à cette position est compté. Comme décrit précédemment, la présence d'une 2'Ome induit une chute du nombre de ces comptages. Ces comptages sont donc utilisés pour calculer un taux de 2'Ome à une position donnée et correspond à une fréquence/taux. Ce taux est appelé Cscore et représente le ratio entre les comptages à la position i et la couverture locale. La couverture locale est calculée comme étant le comptage moyen pondéré dans une fenêtre glissante de 12 nucléotides autour de la position i (six nucléotides à gauche et six nucléotides à droite).

$$Cscore = \max \begin{cases} 1 - \frac{n_i}{\frac{1}{2} \left(\frac{\sum_{j=i-\delta}^{i-1} \omega_j n_j}{\sum_{j=i-\delta} \omega_j} + \frac{\sum_{j=i+1}^{i+\delta} \omega_j n_j}{\sum_{j=i-1}^{i+\delta} \omega_j} \right)} \\ 0 \end{cases}$$

- n_i représente le comptage des lectures 5'-end à la position i
- δ représente la longueur de la fenêtre glissante (6 dans notre cas)
- ω : le poids que le comptage des lectures 5'-end représente. Les comptages 5'-end proches de la position i ont un poids plus important que ceux qui se trouvent loin.

Dans une situation idéale, le RiboMethSeq produirait un profil de comptage des extrémités 5' ou 3' homogènes, sans beaucoup de différence dans ces comptages entre les différentes positions de l'ARNr (Fig : 3.6). En absence d'une 2'Ome et ainsi d'une fragmentation des ARNr par hydrolyse alcaline, aucune chute du nombre de comptages ne sera observée et ainsi la plupart des positions auront un Cscore proche de 0 avec le comptage à la position i proche de la couverture locale. En revanche, dans le cas de la présence d'une 2'Ome et donc d'une résistance à l'hydrolyse alcaline, les sites 2'Ome auront un Cscore proche de 1, résultant d'un comptage à la position i beaucoup plus faible que la moyenne locale (Fig : 3.6). Finalement, les sites présentant une intra-variabilité et étant en conséquence partiellement méthylés dans une population d'ARN auront un Cscore entre 0 et 1 (mix d'ARNr avec ou sans une 2'Ome à la position i). Ces définitions sont en réalité plus complexes, en particulier pour les positions non 2'Ome (Cscore proche de 0). En effet, le bruit de fond du Cscore varie et dépend de la profondeur de séquençage. Ainsi,

pour certaines positions non 2'Ome avec une faible couverture, un Cscore peut être supérieur à 0. Cette problématique est très limitante pour détecter de nouveaux sites. Cependant, une fois les sites 2'Ome cartographiés et confirmés par d'autres méthodes (i.e spectrométrie de masse, primer extension ou identification d'un SNORD), cette problématique n'empêche pas la quantification du niveau de 2'Ome des sites connus. Ainsi, les Cscores de faible valeur à des positions non connues pour être 2'Ome sont exclus de l'analyse.

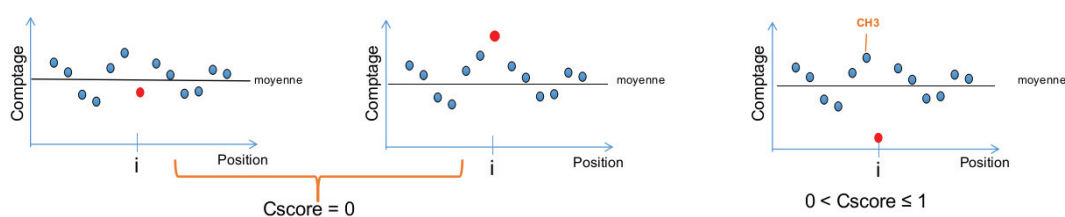


FIGURE 3.6 – Calcul du Cscore. Quand le comptage de la position i est proche de la moyenne locale ou plus élevée, le Cscore sera de 0. Si le comptage de la position d'intérêt i est inférieur à la moyenne locale, le Cscore va varier entre 0 et 1.

Ce protocole de RiboMethSeq a été appliqué pour la première fois à l'analyse de l'ARNr issu de la levure *S.cerevisiae* qui est un organisme modèle pour l'étude de la 2'Ome dans le domaine du ribosome. Malgré les limites de la technique, notamment en termes de détection de sites de 2'Ome, en utilisant un seuil de Cscore de 0.8, les auteurs ont identifié 56 sites 2'Ome, dont 53 sites 2'Ome sur les 55 connus et seulement trois faux positifs. Les deux sites non détectés, 18S_Am100 et 25S_Um2347, présentaient un Cscore au-dessous du seuil, avec des valeurs de 0.77 et 0.61 respectivement. Depuis, le taux de méthylation partiel de ces deux sites a été confirmé par d'autres techniques (RiboMethSeq, spectrométrie de masse, chromatographie en phase liquide à haute performance en phase inverse (RP-HPLC)) (MARCHAND, BLANLOEIL-OILLO et al., 2016; TAOKA, NOBE, YAMAKI, YAMAUCHI et al., 2016; YANG et al., 2017). Ces résultats soutiennent l'hypothèse de l'existence d'une hétérogénéité du ribosome due à l'absence ou à la présence de la 2'Ome à certains sites. Ils ont aussi démontré que le RiboMethSeq, une technologie de séquençage haut débit, permet de capturer cette hétérogénéité. Cependant, l'utilisation d'une procédure de ligation particulière nécessitant la synthèse *in vitro* des adaptateurs 5' et l'utilisation d'un ARN ligase mutée rendent la

technique difficilement utilisable en routine, notamment sur les plateformes de séquençage. De plus, les kits de séquençage les plus utilisés sont actuellement fournis par Illumina (LAHENS et al., 2017), tandis que la technologie Ion Torrent est moins utilisée.

B.3.3 RiboMethSeq et technologie Illumina

Peu de temps après, en 2016, une équipe française (Marchand et collègues) basée à Nancy a proposé une adaptation du protocole original de RiboMethSeq afin de préparer les échantillons pour un séquençage Illumina, permettant à de nombreux laboratoires de s'emparer de cette technique (MARCHAND, BLANLOEIL-OILLO et al., 2016).

Le nouveau protocole utilise le même principe de fragmentation de l'ARN total sous condition alcaline. En revanche, les fragments générés ne sont pas purifiés sur gel et la taille des fragments peut donc être supérieure à 40 nucléotides. Cependant, comme le séquençage repose sur des lectures de 50 à 75pb, les fragments de cette taille doivent être enrichis et le temps d'hydrolyse a été ajusté au matériel biologique utilisé dans ce second protocole. Les extrémités des fragments sont ensuite traitées avec une phosphatase afin de faciliter la ligation avec les adaptateurs en suivant une préparation standard de la librairie de séquençage Illumina. Le remplacement de l'enzyme de ligation mutée de Birkedal *et al.* par des protocoles utilisés en routine disponibles dans des kits standardisés pour le séquençage des ARN, a permis de réduire fortement la quantité de matériel biologique nécessaire. Alors que le protocole original de Birkedal *et al.* nécessite de 1 à 10 µg d'ARN totaux, le protocole adapté de Marchand et collègues nécessite seulement 250 ng d'ARN totaux. La diminution de la quantité d'ARN initiale nécessaire a permis d'envisager de nouvelles applications, notamment dans le domaine de la santé humaine avec des échantillons provenant de patients, mais aussi pour l'étude de la 2'Ome d'autres types d'ARN.

En ce qui concerne le traitement bioinformatique des données de séquençage, les deux protocoles sont très similaires. Marchand *et al.* utilisent le logiciel trimmomatic (BOLGER, LOHSE et USADEL, 2014) au lieu de cutadapt, afin de tronquer les adaptateurs et de garder les lectures supérieures à 17 nucléotides. Néanmoins, l'utilisation de différents logiciels pour cette étape

de contrôle qualité ne devrait pas avoir d'influence sur les résultats. Le Cscore est calculé essentiellement comme décrit par Birkedal *et al* (ratio entre les comptages à la position i et la couverture locale qui correspond au comptage moyen dans une fenêtre glissante de 12 nucléotides autour de la position i). En revanche, la pondération n'est pas réalisée, donnant ainsi le même impact relatif à tous les nucléotides voisins.

Dans leur globalité, les résultats issus de l'analyse de la 2'Ome chez la levure à partir des deux protocoles sont similaires. En effet, à partir de ce nouveau protocole, les auteurs ont aussi analysé la 2'Ome chez la levure *S.cerevisiae*. Sous l'hypothèse d'une 2'Ome complète de toutes les molécules d'ARNr (Cscore > 0.9), les auteurs ont identifié 53 sites 2'Ome, dont 50 sites sur un total de 55 sites connus et trois faux positifs. Par rapport à la technologie Ion Torrent, seuls trois sites connus n'ont pas été détectés, certainement lié à l'utilisation d'un seuil plus élevé.

De plus, Marchand *et al.* ont montré que le RiboMethSeq est une technique robuste et reproductible pour la quantification de la 2'Ome de l'ARNr. Tout d'abord, l'analyse de trois réplicats techniques montre une faible dispersion des valeurs de Cscore, avec des valeurs moyennes de l'écart-type (SD) de 1,35 % pour l'ARNr 18S et de 1,8 % pour l'ARNr 25S; et des valeurs maximales de 3,9 % et de 6,7 % respectivement. En ce qui concerne l'analyse de réplicats biologiques correspondant à 4 souches de levures différentes portant des délétions similaires dans des gènes non essentiels, la dispersion des Cscore est un peu plus élevée que les réplicats techniques, mais reste très raisonnable (SD moyen : 2,1 % pour l'ARNr 18S et 2,0 % pour l'ARNr 25S). De plus, le taux de 2'Ome de deux sites a été mesuré par spectrométrie de masse pour comparaison de la valeur absolue obtenue avec la méthode biochimique et la méthode de RNA-seq. Ainsi le site 25S_Cm2197 présente 61 % de 2'Ome par spectrométrie de masse et un Cscore de 0.71 en RiboMethSeq. Quant au site 25S_Am2220, il présente 96 % de 2'Ome en spectrométrie de masse et un Cscore de 0.96 en RiboMethSeq. Ces valeurs, bien qu'issues de technologies différentes, sont cohérentes.

Enfin, il s'avère que les deux protocoles de RiboMethSeq permettent d'identifier des sites partiellement 2'Ome. Par exemple, le site 18S_Am100 est considéré comme partiellement 2'Ome par les deux protocoles de RiboMethSeq. Ces résultats sont soutenus par une étude de Buchhaupt en 2014 qui a montré par chromatographie en phase liquide couplée à de la spectrométrie de masse

que le 18S_Am100 est partiellement méthylé (68 %) chez la levure (BUCHHAUPT et al., 2014).

Au début de ma thèse, aucune méthodologie de contrôle qualité dédiée au RiboMethSeq n'était disponible. De plus, les analyses de la 2'Ome des ARNr réalisées correspondaient à des analyses simples : identification des sites 2'Ome présentant un Cscore < 0.8 au sein d'une même lignée cellulaire suggérant la possibilité d'une intra-variabilité. Dans un second temps, quelques études avaient comparé les Cscores entre lignées cellulaires avec une approche statistique essentiellement basée sur des comparaisons de moyennes. Afin de déterminer s'il existe une hétérogénéité de la 2'Ome dans les cancers, j'ai utilisé des méthodes statistiques plus poussées explicitées dans la prochaine partie de ce chapitre.

C Méthodes d'analyses utilisées pour exploiter les données de RiboMethSeq issues d'échantillons humains

C.1 Analyses exploratoires des données

Les analyses exploratoires permettent d'étudier, d'explorer et d'analyser un jeu de données afin d'identifier des structures dans les données, des relations entre les observations (c.-à-d. un échantillon, un gène ou un site 2'Ome) et d'évaluer la qualité de ces données. Ces méthodes n'ont pas besoin de connaissances *a priori* sur les données et sont précurseurs des hypothèses biologiques et des tests statistiques. Je vais décrire brièvement trois de ces méthodes que j'ai utilisées au cours de cette thèse : une visualisation univariée sous forme de boxplot, une analyse de réduction de dimension (analyse en composante principale) et une analyse de clustering (regroupement hiérarchique).

C.1.1 Visualisation des données quantitative avec boxplot

Le boxplot (boîte à moustache) est une des méthodes les plus utilisées pour visualiser la distribution de données quantitatives (Fig : 3.7). Il permet de visualiser 5 chiffres clés d'un ensemble de données :

- (i) le minimum : la valeur la plus basse de l'ensemble des données sans prendre en compte les valeurs aberrantes (outlier)
- (ii) le premier quartile Q1 : la valeur dont un quart de l'ensemble des données est inférieure
- (iii) la médiane : la valeur médiane de l'ensemble des données
- (iv) le troisième quartile Q3 : la valeur dont trois quarts de l'ensemble des données est inférieure
- (v) le maximum : la valeur la plus haute de l'ensemble des données sans prendre en compte les valeurs aberrantes (outlier)

Le graphique est visualisé par un simple rectangle qui est limité par les valeurs Q1 et Q3 et qui est coupé par la valeur médiane (Fig : 3.7). La taille du rectangle est définie comme étant le rang interquartile IQR (la distance entre les valeurs Q1 et Q3). Du rectangle sortent deux droites dont les extrémités représentent les valeurs minimale et maximale. En général, la valeur minimale est définie comme $Q1 - 1,5 * IQR$ et la valeur maximale comme $Q3 + 1,5 * IQR$. Les droites peuvent être de longueurs inégales puisque les valeurs minimale et maximale représentent des valeurs observées parmi l'ensemble des données. Les valeurs au-delà des limites de ces deux droites sont considérées comme des outliers et sont représentées par des points ou cercles. Tout au long de cette thèse, l'IQR sera utilisé comme une mesure de la dispersion des Cscores.

C.1.2 Analyse en composantes principales

L'analyse en composantes principales (ACP) est une méthode d'analyse de données exploratoire non supervisée, parmi les plus utilisées quand il s'agit de données de grandes dimensions (multidimensionnelles). Elle fait partie de la famille des méthodes d'analyses multivariées et n'a pas besoin d'hypothèse nulle à tester (méthode descriptive). Le jeu de données multi-dimensionnelles de départ se présente sous forme d'un tableau ou d'une matrice à dimension $n \times p$ où n est le nombre d'individus et p le nombre de variables explicatives quantitatives.

Dans le cas des données transcriptomiques par exemple, on mesure l'abondance de milliers de gènes pour plusieurs échantillons en même temps. Ces informations sont souvent stockées dans une matrice qui représente en ligne

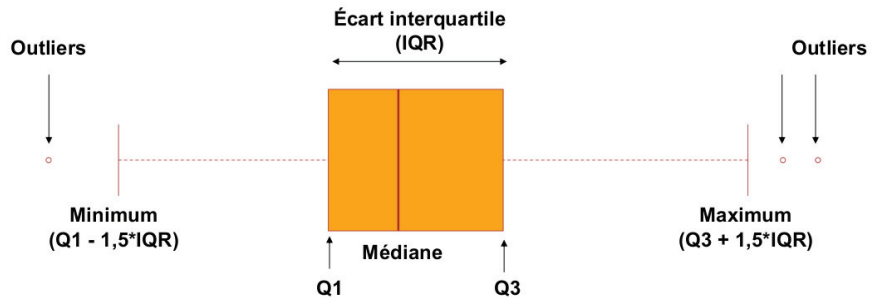


FIGURE 3.7 – Visualisation des données quantitatives par box-plot.

les échantillons et en colonne les gènes. Plusieurs questions sur les échantillons et les gènes peuvent être posées. Souvent, on s'intéresse à savoir si les échantillons ont un profil d'expression globale similaire ou différent : Quels échantillons se ressemblent et lesquels se différencient ? Quels sont les gènes responsables de ces caractéristiques de ressemblances ou de différences ? Il est donc difficile de répondre à ces questions en analysant les gènes un à un ou sans avoir une vue d'ensemble.

L'objectif de l'ACP est de réduire les dimensions du jeu de donnée de départ en gardant la structure et les informations les plus importantes. La méthode construit de nouvelles variables qui captent le maximum de variation du jeu de données tout en étant décorréliées entre elles. Ces nouvelles variables sont appelés composantes principales (PC). Les PC sont les résultats de combinaisons linéaires des variables d'origines (dans le cas RNA-seq, des gènes). Ces PC synthétisent l'information du jeu de données multi-dimensionnelles de départ : si certains gènes sont fortement corrélés entre eux comme c'est souvent le cas dans les données transcriptomiques, ils peuvent être synthétisés par l'information d'une seule composante principale. C'est pour cela qu'on parle de réduction de dimension.

Après identification des composantes principales, elles sont ordonnées en fonction de la portion de variance captée par chacune. Ainsi, par définition

la PC1 est la première composante principale qui explique le plus de variance dans le jeu de données, PC2 explique la deuxième plus grande partie de la variance et ainsi de suite.

Il est maintenant possible de visualiser les PC par paires (PC1 vs PC2 ou PC2 vs PC3) sur des graphiques de type nuage de points. Un point représente un échantillon. Dans le cas où les deux premières PC expliquent une grande partie de la variance totale, le reste des PC peuvent ne pas être prises en compte et un seul graphique peut suffire pour résumer les données sans perte d'information. Il existe d'autres cas où il est nécessaire de visualiser plusieurs combinaisons de PC afin de mieux explorer les différentes relations ou structures qui peuvent exister dans les données.

L'ACP peut être réalisée sur la totalité des données ou un sous-groupe de variable (c.-à-d. gènes) afin de mieux identifier les relations entre les échantillons. Par exemple, on peut sélectionner les gènes les plus variants à travers les échantillons ou un sous-groupe de gènes d'intérêt.

Dans le cadre de cette thèse, nous allons nous intéresser au taux de 2'Ome (Cscore) pour les différents sites de l'ARNr chez l'Homme. La matrice de départ contiendra alors le Cscore de chaque site (colonne) et chaque échantillon (ligne). Voici un exemple d'ACP conçue sur les Cscore des 106 sites de l'ARNr chez l'Homme (Fig : 3.9). À gauche, le graphique représente le pourcentage de variance capturé par chaque PC (Fig : 3.8). On remarque que PC1 explique 49.1 % de la variance totale, largement supérieur aux autres PC. La figure (Fig : 3.9) montre une représentation en nuage de points de PC1 en abscisses et PC2 en ordonnées. Deux échantillons sont proches s'ils présentent des profils de Cscore similaires (FMG40, FMG25 et FMG33 par exemple). PC1 sépare les échantillons en deux groupes. Des analyses complémentaires et la traçabilité de l'expérimentation ont montré que la variabilité capturée par PC1 est due à un effet batch. En plus d'identifier des effets techniques ou biologiques, l'ACP permet d'identifier des échantillons aberrants qui présentent des profils de 2'Ome très différents des autres. Ces échantillons sont généralement exclus de l'analyse.

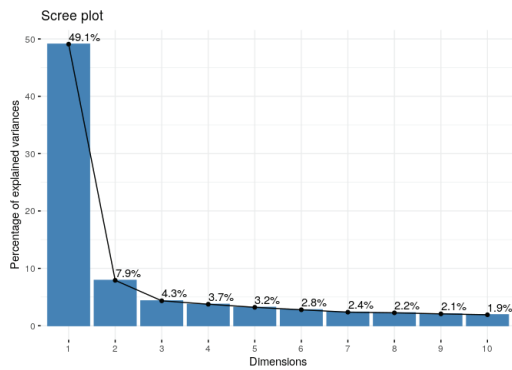


FIGURE 3.8 – Pourcentage de variance expliqué par chaque composante principale

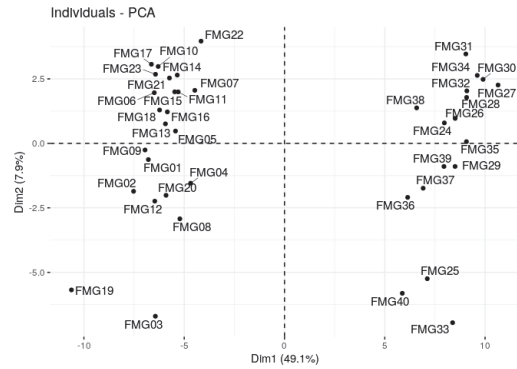


FIGURE 3.9 – Exemple de visualisation de l'ACP. La figure montre les deux premiers axes de l'ACP. Chaque point représente un échantillon.

C.1.3 Regroupement hiérarchique (hierarchical clustering)

Le regroupement hiérarchique (HC pour hierarchical clustering), comme l'ACP, est une méthode d'analyse exploratoire et non supervisée. Il a pour but d'identifier des groupes ou « clusters » dans un jeu de données. Les observations appartenant à un même groupe doivent être similaire entre elles, mais différentes de celles appartenant à un autre groupe. Dans le cas des données d'expression issues de RNA-seq, on s'intéresse à identifier les échantillons qui ont des profils d'expression similaire, mais aussi à déterminer quels sont les gènes qui s'expriment de façon similaire entre les échantillons. Le jeu de données de départ est organisé de la même façon que pour l'ACP : une matrice à dimension $n \times p$ où n est le nombre d'individus et p le nombre de variables explicatives quantitatives. Au cours de cette thèse, les variables explicatives quantitatives seront les taux de 2'Ome (Cscore) mesurés à l'aide du RiboMethSeq.

La première étape de HC est d'évaluer la similarité ou la dissimilarité entre les observations (échantillons ou sites 2'Ome). Pour cela, il est tout d'abord nécessaire de comparer les observations deux à deux et de calculer une distance. Cette distance sera utilisée par la suite pour regrouper les observations qui sont similaires. Il existe différentes mesures de distance comme les distances de corrélation euclidienne, Manhattan. Cette dernière mesure de

distance sera souvent utilisée au cours de cette thèse, car elle est moins influencée par des valeurs aberrantes et décrit ainsi mieux les données de Cscore. Deux observations sont considérées comme étant similaires quand la distance est faible et comme étant très différentes quand la distance est élevée.

Dans un deuxième temps, il faut définir une méthode d'agrégation (linkage method) qui permet de calculer cette fois-ci une distance entre les groupes/clusters. Les deux groupes les plus proches seront regroupés à chaque itération de l'algorithme jusqu'à n'obtenir qu'un seul groupe. Là aussi, différentes métriques existent : lien simple (single linkage), lien complet (complete linkage), lien moyen (average linkage) et celui de Ward qui est le plus utilisé. Contrairement aux autres méthodes qui mesurent une distance directe entre les groupes, la méthode de Ward prend en compte la variance intra-groupe.

La visualisation des résultats d'une analyse HC se fait souvent sous forme d'un arbre appelé dendrogramme (Fig : 3.10). La partie basse de l'arbre est constituée de chaque observation appelée feuille (dans le cas de la figure 3.10, les observations sont les échantillons). Avec la hauteur de l'arbre, les feuilles fusionnent pour donner des branches. Plus cette fusion se passe tôt, plus les feuilles/branches sont similaires. À l'inverse, quand les branches fusionnent tardivement, les groupes sont différents.

Le dendrogramme peut être coupé à différents niveaux, ce qui donne des groupes ou clusters de tailles différentes. Les groupes peuvent être comparés pour établir les différences et les similarités, ou encore comparés *a posteriori* avec des données techniques, biologiques ou cliniques. Le dendrogramme de la figure 3.10 a été coupé de façon à obtenir deux groupes colorés en rose et bleu. Ces deux groupes correspondent à l'effet batch déjà vu par l'ACP.

Le dendrogramme est souvent accompagné d'une heatmap qui dans notre cas représente le Cscore (Fig : 3.11). La couleur de la heatmap est liée à la valeur du Cscore : rouge pour les sites ayant un Cscore proche de 1 et bleu pour les sites ayant un Cscore proche de 0. L'ordre des colonnes (échantillons) est défini par l'analyse HC. Dans la figure 3.11, le dendrogramme des échantillons est le même que celui de la figure 3.10. On remarque que la heatmap possède un deuxième dendrogramme sur les lignes qui représentent les sites 2'Ome de l'ARNr permettant ainsi d'identifier des groupes de sites 2'Ome au comportement similaire ou dissemblable. L'HC est complémentaire de l'ACP

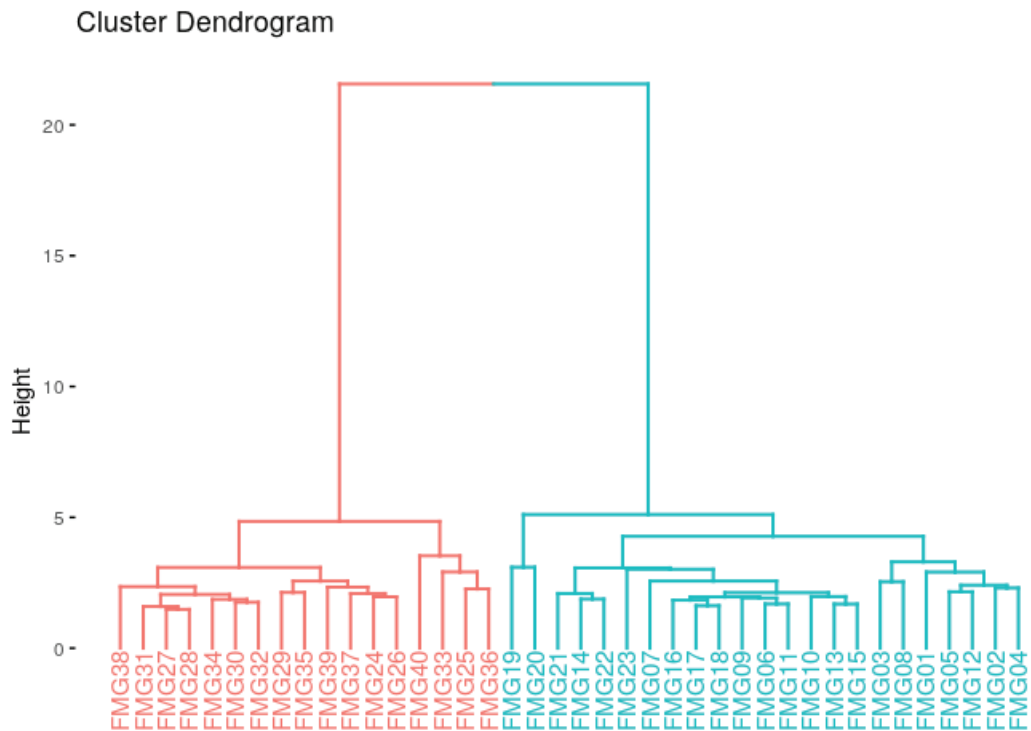


FIGURE 3.10 – Exemple de dendrogramme. Chaque observation est un échantillon. L'arbre est coupé afin d'avoir deux groupes colorés en rouge et bleu.

et permet d'identifier des profils de 2'Ome similaires ou distincts entre des échantillons, les sites 2'Ome de l'ARNr ainsi que les échantillons outliers, ou encore des effets batch.

C.2 Comparaison du taux de 2'Ome entre deux ou plusieurs groupes

Dans les données transcriptomiques, il est courant de comparer deux ou plusieurs groupes (ou conditions biologiques) différents et d'identifier les gènes responsables de ces différences. De la même façon, dans les données de RibomethSeq, il est intéressant d'identifier les sites dont la 2'Ome est différente entre les conditions. Pour décider si un site est différentiellement 2'Ome entre les groupes, il est important de réaliser un test statistique. Le test va être confronté à une hypothèse nulle qui est souvent du type « il n'existe pas de différence entre les groupes ». La significativité du test est mesurée en calculant la probabilité d'obtenir les valeurs observées si l'hypothèse nulle est vrai : la valeur p (p-value). L'hypothèse nulle est rejetée quand la p-value est

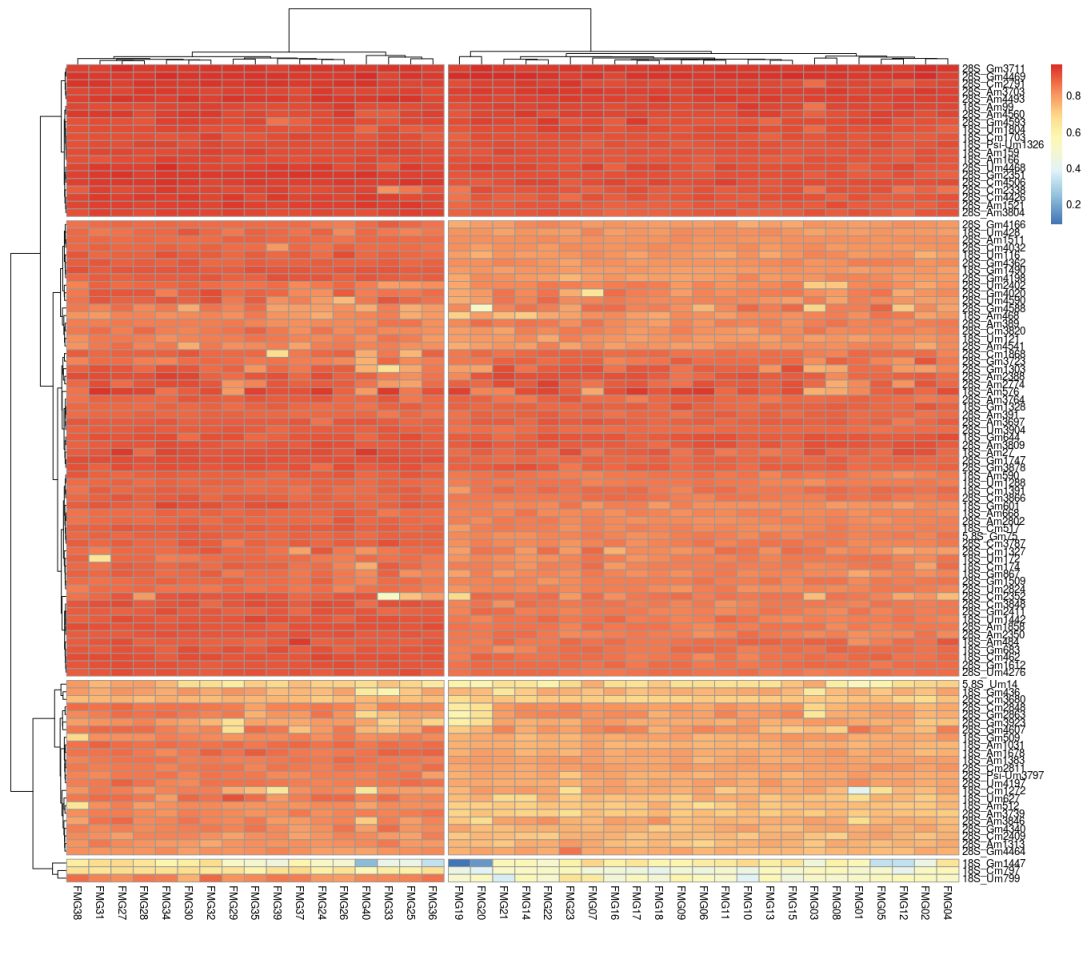


FIGURE 3.11 – Exemple de heatmap. En colonne est représenté HC des échantillons et ligne celle des sites 2’Ome. La couleur du heatmap représente la valeur du Cscore : rouge pour un Cscore proche de 1 et bleu proche de 0.

inférieure au risque de première espèce α qui peut-être fixée à 5 % (0.05). On parle alors de test significatif. Plus la p-valeur est proche de 0, plus les valeurs observées s’éloignent de l’hypothèse nulle. Le risque alpha permet de contrôler le taux de faux positifs (rejeter l’hypothèse nulle et déclarer un site comme différentiellement 2’Ome alors qu’il ne l’est pas). Il existe aussi un risque de deuxième espèce β qui correspond au fait de ne pas rejeter l’hypothèse nulle alors que cette hypothèse est fautive (ne pas détecter un site réellement différentiellement 2’Ome). Le risque β n’est pas fixé *a priori*, contrairement au risque α et est difficile à estimer.

Quand le test statistique est répété plusieurs fois de suite pour un même jeu de données, un risque alpha de 0.05 est attendu à chaque test. Ainsi, plus il

Il y a de tests, plus la probabilité de déclarer une différence significative au hasard augmente. Par exemple, il existe environ 106 sites 2'Ome de l'ARNr chez l'Homme. Réaliser un test statistique pour chaque site avec un risque alpha de 0.05 augmente ainsi la probabilité de déclarer un site comme différentiellement 2'Ome par hasard à $0.9956 (1 - (1 - 0.05)^{106})$ ou autrement dit 99,56 % de chance de se tromper sur au moins un site. Pour parer à ce problème, différentes méthodes ont été développées. La méthode de Benjamini et Hochberg est une des plus répandues et permet de contrôler le nombre de faux positifs parmi ceux déclarés comme significatifs (BENJAMINI et HOCHBERG, 1995). C'est cette méthode qui sera utilisée dans cette thèse.

Les tests statistiques qui ont pour objectifs de détecter s'il y a une différence significative entre deux ou plusieurs conditions, se séparent en deux catégories : les tests paramétriques et non-paramétriques.

Les tests paramétriques se basent sur l'hypothèse d'une distribution supposée des données. Par exemple, le test t de Student pour comparer deux moyennes de deux groupes d'échantillons suppose que les données suivent une distribution normale. Les tests non-paramétriques n'ont pas besoin d'une hypothèse sur la distribution des données et ne se basent pas sur la moyenne ou la variance, mais plutôt sur le rang. À noter qu'en contrepartie les tests non-paramétriques sont moins puissants et peuvent ne pas détecter certaines différences. N'ayant pas beaucoup d'information sur la distribution du Cs-core à un site donné et un nombre d'échantillons limité, nous avons privilégié les tests non-paramétriques dans cette thèse.

4 Hypothèse et objectifs

Pendant longtemps, le ribosome a été considéré comme un simple effecteur de la synthèse protéique. Or, des études récentes ont montré que des ribosomes de compositions différentes participent à une reprogrammation traductionnelle qui pourrait contribuer à la mise en place de phénotypes particuliers (GENUTH et BARNA, 2018; VENEZIA et al., 2019). Mon équipe d'accueil a démontré pour la première fois, à l'aide de la technique de primer extension couplée à une RTqPCR, que la 2'O-méthylation des riboses (2'Ome) des ARN ribosomiques (ARNr) est altérée à certains sites seulement au cours de l'initiation tumorale mammaire dans des modèles cellulaires, et que cette altération site-spécifique affecte directement la traduction de certains ARNm codant des oncogènes comme IGF1R ou c-myc (ERALES et al., 2017; Virginie MARCEL, GHAYAD et al., 2013). Par ailleurs, mon équipe d'accueil a montré par une étude de 3500 tumeurs que l'expression de la méthyltransférase des ARNr, FBL, est altérée dans les cancers du sein et représente un marqueur indépendant de mauvais pronostic (LONG et al., 2022). Ces données suggèrent que la 2'Ome des ARNr est altérée dans les cancers du sein et induit une dérégulation de la traduction favorisant ainsi la tumorigenèse mammaire et la progression de la maladie. Au début de ma thèse, il restait à établir que la 2'Ome des ARNr varie dans les cancers et à caractériser l'association entre la 2'Ome des ARNr et des caractéristiques biologiques et cliniques particulières afin de comprendre leur rôle en cancérologie.

Aborder une telle thématique n'était pas réalisable avec les techniques disponibles qui nécessitaient beaucoup de matériel biologique et de temps. Fin 2015, une approche innovante basée sur le RNA-seq, le RiboMethSeq, a été développée afin d'analyser le taux de 2'Ome de l'ensemble des sites méthylés de l'ARNr chez la levure en un seul séquençage (BIRKEDAL, CHRISTENSEN-DALSGAARD et al., 2014). Cette approche repose sur la fragmentation des ARN par une hydrolyse alcaline ménagée, qui est inhibée en présence d'une 2'Ome, et qui permet ainsi de calculer la fréquence de 2'Ome à un site donné.

Dès 2017, mon équipe d'accueil a contribué à transférer la technologie de RiboMethSeq chez l'Homme et a initié le développement d'outils bioinformatiques dédiés à l'analyse des données de RiboMethSeq. Grâce à cette approche, l'analyse de la 2'Ome des ARNr dans les tumeurs était réalisable à mon arrivée en thèse, même si des contrôles qualité ou des méthodologies de normalisation étaient manquantes et nécessaires pour l'analyse de grandes cohortes.

Ainsi, sur la base de la littérature et de nos données préliminaires, l'hypothèse sur laquelle est basée mon projet de thèse est que les ribosomes présentent des profils de 2'Ome différents chez l'Homme, plus particulièrement dans les cancers, et que ces profils sont associés à des caractéristiques biologiques et cliniques distinctes. Mon projet de thèse a pour objectifs d'étudier la diversité de la 2'Ome des ARNr dans les cancers et d'identifier les sites d'intérêt en cancérologie. En particulier, mon projet de thèse a pour objectifs de :

1. Développer des outils bioinformatiques et des méthodologies innovantes dédiés au RiboMethSeq permettant l'analyse de séries d'échantillons humains
2. Identifier et décrire les différents profils de 2'Ome survenant dans les cancers (cancers du sein et gliomes)
3. Déterminer l'association de ces différents profils de 2'Ome avec les caractéristiques biologiques et cliniques des tumeurs

5 Résultats

A Riboscore : an R package for RiboMethSeq data analysis

Ce package correspond à la première version d'un package, dont je suis deuxième auteur, est en cours d'écriture par un ingénieur bioinformaticien avec lequel je travaille étroitement pour l'aider à la conception de cet outil qui regroupe l'ensemble des méthodes que j'ai développées au cours de ma thèse. La vignette du package est incluse après cette page introductive.

A.1 Contexte

Mon projet de thèse se focalise sur une modification chimique particulière des ARN : la 2'-O-ribose méthylation (2'Ome). Comme expliqué dans l'introduction et dans la revue, la 2'Ome des ARNr participe à l'hétérogénéité de composition du ribosome. En effet, de nombreuses études ont montré que le niveau de 2'Ome des ARNr varie dans les lignées cellulaires mais aussi dans les organismes vivants. Plusieurs laboratoires de recherche dans le monde s'intéressent aujourd'hui à la 2'Ome avec pour objectifs d'identifier les différents contextes biologiques dans lesquels le niveau de 2'Ome varie et de comprendre si cette variation permet aux ribosomes d'acquérir des activités de régulation de la traduction et de spécifier des phénotypes particuliers.

Ces études reposent principalement sur l'approche de RiboMethSeq, publiée pour la première fois en 2015. Cette technique basée sur le RNA-seq permet de quantifier le taux de 2'Ome à tous les sites connus de l'ARNr en un seul séquençage. Au début de ma thèse, il n'existait aucun outil dans la littérature dédié aux données de RiboMethSeq. Nous avons voulu intégrer les différentes méthodes d'analyses et de visualisation de données utilisées pendant ma thèse dans un unique package R pour faciliter l'utilisation du RiboMethSeq mais aussi standardiser les analyses en vue d'homogénéiser les données

et d'ainsi pouvoir, un jour, envisager de comparer des études provenant de différents laboratoires.

A.2 Résultats principaux

Ce package est dédié à l'analyse de la 2'Ome de l'ensemble des ARN et calcule le Cscore sur ensemble des positions génomiques des ARN d'intérêt avant extraction des Cscores aux positions biologiquement pertinentes. Pour les ARNr humains, ce package fournit une liste annotée des sites décrits dans la littérature. Au niveau fonctionnel, ce package R permet de :

- (i) Lire les fichiers de comptages des extrémités 5'/3' des lectures obtenues à partir des données de séquençage
- (ii) Ajouter une description des échantillons à analyser pour faciliter l'analyse et la visualisation
- (iii) Calculer le Cscore en offrant la possibilité de choisir sa méthode de normalisation (moyenne ou médiane par rapport à la couverture locale) ainsi que la taille de la couverture locale
- (iv) Ajuster les effets batchs en utilisant l'algorithme de Combat-seq intégré dans le package, par normalisation du comptage des extrémités 5'/3' des lectures
- (v) Apporter une annotation basée sur la littérature des sites 2'Ome de l'ARNr humain ou offrir la possibilité d'utiliser ses propres annotations de sites 2'Ome dans le cas d'étude de l'ARNr provenant d'autres organismes ou d'autres types d'ARN
- (vi) Réaliser et visualiser des analyses de type non supervisées comme l'ACP
- (vii) Exporter les résultats sous différents formats de tableaux

A.3 Ma contribution

J'ai contribué à la conception du package, l'écriture du code, la mise en place des visualisations et des méthodes d'analyses non-supervisées ainsi que des tests statistiques permettant d'identifier les altérations de la 2'Ome entre différentes conditions. Par ailleurs, j'ai proposé et mis en place une méthode d'ajustement des effets batch. Enfin, j'ai participé à l'écriture de la vignette et du guide d'utilisation associés au package.

A.4 Vignette du package R

Riboscore: package to analyze RiboMethSeq data

Théo COMBE, Hermes PARAQUINDES, Janisse KIELBASSA, Emilie THOMAS,
Anthony FERRARI and Virginie MARCEL

10/27/2022

Abstract

RiboMethSeq is an innovative RNaseq-based approach, which has been developed in 2015 to analyze 2'O-ribose methylation (2'Ome) at all the sites of the ribosomal RNAs (rRNA) at once, in yeast (Birkedal et al. 2014). RiboMethSeq approach was then transferred to human using the Illumina technology (Marchand et al. 2016; Erales et al. 2017; Marcel et al. 2021). Briefly, presence of 2'Ome protects the phosphodiester bond located at the 3' of the 2'Ome nucleotide from alkaline hydrolysis. Thus, the presence of 2'Ome at the given nucleotide n induces under-representation of RNA fragments starting at the nucleotide $n+1$ and an over-representation of RNA fragments ending at the nucleotide $n-1$, allowing to calculate a 2'Ome level at the corresponding nucleotide position n (or *C-score*) varying from 0 to 1 (Birkedal et al. 2014). The Riboscore package can be used for any kind of RNA and organisms. This package provides a set of easy-to-use functions to compute *C-scores* from RiboMethSeq read end counts as input, adjust batch effect with ComBat-Seq, visualize the data and provide a table with the annotated human rRNA sites and their *C-scores*. Processing of the raw data to obtain read end counts from sequencing data were previously described (Marchand et al. 2016).

Contents

Installation	2
Usage	2
Help, bug reports and suggestions	2
Acknowledgements	3
Funding	3
Data	3
RiboClass	3
Loading data	3
Data	4
Metadata	4
RNA names	5
C-score calculation	5
What is C-score	5
C-score computation when loading data	5
C-score computation during the analysis process	6
Quality control	6
Identification of batch effect	6
Batch effect adjustment	7

Sample manipulation	8
Keep or remove samples	8
RNA annotation	9
RNA manipulation	9
Annotation of RNA 2'Ome sites	9
Annotate RNA sites	10
Data visualization with PCA	10
Export data	12
Export as a dataframe	12
Export as a ggplot-ready dataframe	13
Export as a dataframe by condition	13
Sessioninfo	14
Reference	15

Installation

The latest version of Riboscore package can be installed from CRAN with :

```
install.packages("riboscore")
```

Usage

```
library(Riboscore)

ribo <- load_ribodata(
  count_path = "/path/to/your/csvfiles/directory/",
  metadata = "path/to/metadata.csv",
  metadata_key = "filename",
  metadata_id = "samplename")

# Compute the c-score using different parameters,
# including calculation of the local coverage using the mean instead of the median
ribo <- compute_cscore(ribo, method = "mean")

# If necessary, adjust any technical biases using ComBat-Seq.
# Here, as an example, we use the "library" column in metadata.
ribo <- adjust_bias(ribo,"library")

# Plot a Principal Component Analysis (PCA) whose colors depend on the "condition" column in metadata
plot_PCA(ribo,"condition")
```

Help, bug reports and suggestions

To report a bug or any suggestion to improve the package, please let us know by opening a new issue on: *github issue link coming soon!*

Acknowledgements

We would like to thank all our collaborators for their advices and suggestions.

Funding

This project has been funded by the French Cancer Institute (INCa, PLBIO 2019-138 MARACAS), the SIRIC Program (INCa-DGOS-Inserm_12563 LyRICAN) and Synergie Lyon Cancer Foundation.

Data

RiboClass

RiboClass is the main class of the package that allows to store both the data matrices (counts and c-scores) and the associated metadata. It is automatically created when calling `load_ribodata` (see Loading data).

It is a list containing three main elements, individually described below:

1. **Data**: a list of dataframe, containing for each sample the 5' and/or 3' read-end counts provided by the user, and the calculated c-score.
2. **Metadata**: a dataframe, containing all the information related to the samples that can be provided by the user.
3. **RNA_names**: a dataframe, reporting the names of the RNA used in Data.

Some main function calls' parameters (such as the normalization method used for c-score computation) are also kept in the RiboClass object as a reminder.

Loading data

Here is an example, where all `load_ribodata` parameters are shown :

```
path <- system.file("extdata", package="Riboscore")

ribo <- load_ribodata(
  #data & metadata files path
  count_path = paste0(path, "/miniglioma"),
  metadata = paste0(path, "/metadata.csv"),
  # data & metadata files separator
  count_sep = "\t",
  metadata_sep = ",",
  # count data parameters :
  count_header = FALSE,
  count_value = 3,
  count_rnaid = 1,
  count_pos = 2,
  # Metadata parameters :
  metadata_key = "filename",
  metadata_id = "samplename",
  # c-score parameters :
  flanking = 6,
  method = "median",
  ncores = 1)

#> [SUCCESS] Your data have been imported and the following RiboClass has been created :
#> a RiboClass with 14 samples and 4 RNA(s) :
```

```
#> Name : NR_023363.1_5S, length : 121
#> Name : NR_046235.3_5.8S, length : 157
#> Name : NR_046235.3_18S, length : 1869
#> Name : NR_046235.3_28S, length : 5070
```

Data

To use this package, the user should provide one csv file with the 5', 3' or 5'/3' read end counts resulting from RiboMethSeq data per sample. The folder structure containing the csv files is not important, as long as either the directory and its sub-directories contain the necessary csv files.

1. The **name of the RNA** on which the read end counting has been performed.
2. The **number of the position** on the RNA.
3. The **value of the read end counts** at the position.

Here is an example :

RNA	Position on RNA	read end count
18S	123	3746
18S	124	345
18S	125	324
18S	126	789
18S	127	1234

Note 1: it is not necessary to provide an header in the count files, because column index can be used in the function `load_ribodata`, using `count_value`, `count_rnaid` and `count_pos` parameters.

Note 2: if no metadata is specified, Riboscore will try to fetch any csv files in the folder specified in `count_path` and its subfolders.

Metadata

Two columns are mandatory for the metadata dataframe:

1. **filename**: name of the csv file on disk containing the read end counts described above. Do not modify it unless the filename has changed on disk.
2. **samplename**: rename the samples, which will be analyzed and shown on the plots. This column could be modified, as long as the sample names are unique.

Following these two mandatory columns, provide as many columns as needed for the analysis.

Here is an example of metadata for three samples:

filename (mandatory)	samplename (mandatory)	biological condition (optionnal)
sample1.csv	sample 1	condition 1
sample2.csv	sample 2	condition 1
sample3.csv	sample 3	condition 2

Note: if no metadata are provided in `load_ribodata`, an empty metadata will be created with pre-completed "filename" and "samplename" columns. The "samplename" column will be identical to "filename", but can be modified by the user.

Here is an example of auto-generated metadata:

filename	samplename
sample1.csv	sample1.csv
sample2.csv	sample2.csv
sample3.csv	sample3.csv

RNA names

RNA names are stored in an auto-generated dataframe when loading data. It contains two columns :

1. **original_name**: original name of each RNA (e.g NR_023363.1).
2. **current_name**: current name of each RNA, reflecting any user's change with *rename_rna* function (see Rename RNA).

This dataframe is used to keep track of the original name, which often contains NCBI's accession ID.

Here is an example:

original_name	current_name
NR_023363.1_5S	5S
NR_046235.3_5.8S	5.8S
NR_046235.3_18S	18S
NR_046235.3_28S	28S

C-score calculation

What is C-score

The C-score corresponds to the 2'Ome level at a RNA position. The C-score represents a drop in the end read coverage at a given position compared to the environmental coverage, as described by Birkedal et al. (2014). The C-score can be of 0 (i.e., no RNA molecule is 2'Ome at the position of interest), of 1 (i.e., all the RNA molecules are 2'Ome at the position of interest) and of]0:1[(i.e., a mix of un-methylated and methylated RNA molecules).

Different C-scores can be determined to obtain robust estimation of 2'Ome level depending on the parameters used to compute the local coverage. In particular, the computation method and the size of the local coverage can be changed.

By default, the computation method of the local coverage corresponds to the **median** and the size of the local coverage corresponds to a **flanking region of 6** (i.e., 6 nucleotides downstream the nucleotide n and 6 upstream the nucleotide n). This package offers the possibility to change these two parameters either when loading the data or during the analysis process.

C-score computation when loading data

When the function *load_ribodata* is used, a C-score is automatically computed for all genomic positions of the RNA. The C-score is computed using either the default parameters of the *load_ribodata* function or the user's parameters as followed:

```
load_ribodata(count.path = "/path/to/csv",
              metadata = "/path/to/metadata.csv",
              # everything below is linked to c-score computation
              flanking = 6, # flanking region size
```

```
method = "median", # use mean or median on flanking region's values
ncores = 8 # number of CPU cores to use for computation
)
```

C-score computation during the analysis process

During the analysis, the parameters to compute the C-score can be modified using the `compute_cscore` function, which will automatically update the C-score in the RiboClass.

In the following example, both the flanking region's size of the local coverage and the computation method have been modified:

```
ribo <- compute_cscore(ribo,
                      flanking = 4,
                      method = "mean")
```

Note: this function will override the previous c-score of the RiboClass.

Quality control

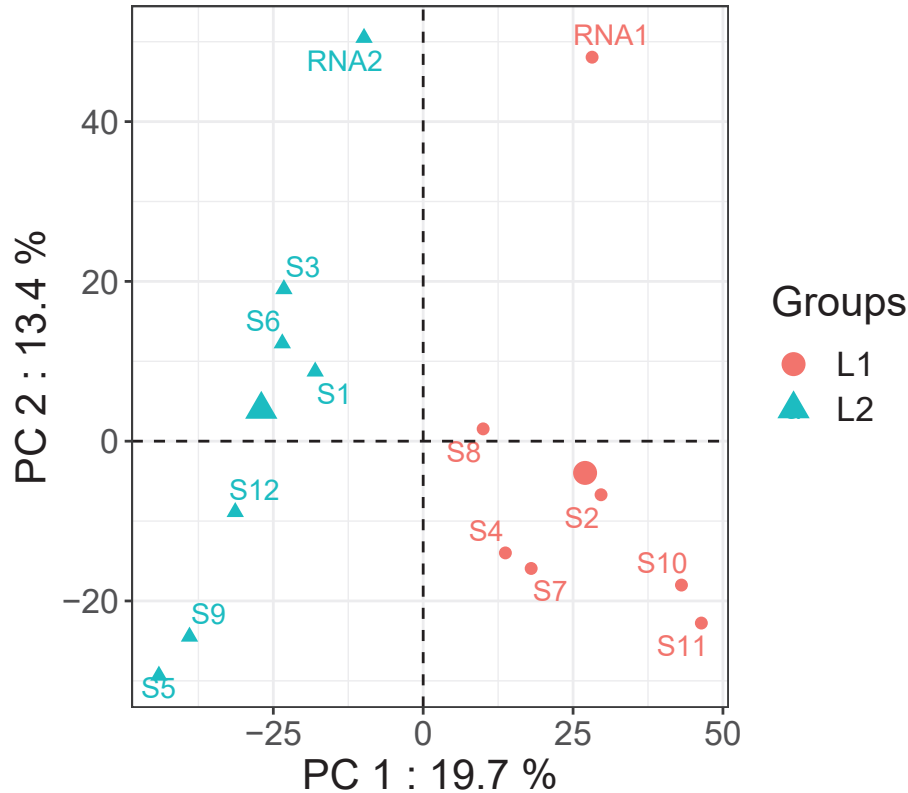
Identification of batch effect

Technical bias (i.e., batch effect) can be identified by plotting C-scores at all the genomic positions of the RNA for each sample on a PCA (see also [Visualization with PCA] for more usages).

Here is an example:

```
plot_PCA(ribo = ribo,
         color_col = "run")
```

PCA of Cscore for 7161 sites 14 samples



In this example, the technical replicates RNA1 and RNA2 included in library 1 and 2 respectively, are distant from each other on the PC1 axis. Moreover, the samples should not be grouped by library or batch. The following section will resolve this batch effect.

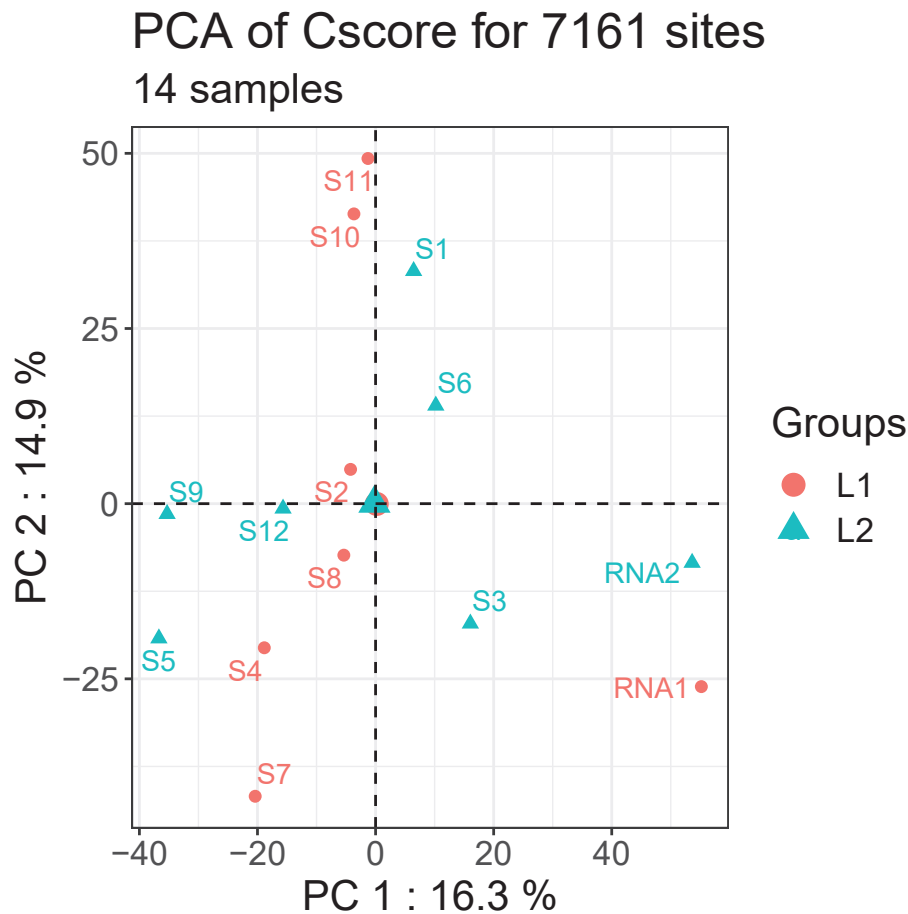
Batch effect adjustment

Batch effect of RiboMethSeq data can be adjusted using the ComBat-seq method (Paraquindes et al, in preparation; Zhang, Parmigiani, and Johnson (2020)) . The Riboscore package includes a wrapper (`adjust_bias`) to perform ComBat-seq adjustment that will return a new RiboClass with adjusted read end count values and C-scores automatically recomputed with the same setup parameters.

```
ribo_adjusted <- adjust_bias(ribo, batch = "run")
#> Found 2 batches
#> Using null model in ComBat-seq.
#> Adjusting for 0 covariate(s) or covariate level(s)
#> Estimating dispersions
#> Fitting the GLM model
#> Shrinkage off - using GLM estimates for parameters
#> Adjusting the data
#> Recomputing c-score with the following parameters :
#> - C-score method : median
#> - Flanking window : 6
```

Batch effect adjustment can be verified using the `plot_PCA` function using the new RiboClass:


```
plot_PCA(ribo_adjusted,"run")
```



After batch effect adjustment using ComBat-seq method, the two technical replicates RNA1 and RNA2 show reduced dispersion, and the samples are separated on the PCA axes independently of the library groups.

Sample manipulation

Keep or remove samples

A sample subset can be easily analyzed by indicating the samples to keep or to remove. The user can thus create a new RiboClass object containing the data and metadata of the samples of interest. In both cases, only the remaining samples' metadata are kept in the RiboClass object, so no manual updating is required.

Here is an example to generate a new RiboClass by keeping two samples of interest ("S1" and "S2"):

```
ribo_2samples <- keep_ribo_samples(ribo_adjusted,c("S1","S2"))
print(ribo_2samples)
#> a RiboClass with 2 samples and 4 RNA(s) :
#> Name : NR_023363.1_5S, length : 121
#> Name : NR_046235.3_5.8S, length : 157
#> Name : NR_046235.3_18S, length : 1869
#> Name : NR_046235.3_28S, length : 5070
```

Here is an example to generate a new RiboClass by removing two samples ("S1" and "S2"):

```

ribo_removed_samples <- remove_ribo_samples(ribo,c("S1","S1"))
print(ribo_removed_samples)
#> a RiboClass with 13 samples and 4 RNA(s) :
#> Name : NR_023363.1_5S, length : 121
#> Name : NR_046235.3_5.8S, length : 157
#> Name : NR_046235.3_18S, length : 1869
#> Name : NR_046235.3_28S, length : 5070

```

In both cases, only the remaining samples' metadata are kept in the RiboClass object. You do not need to update it by hand.

RNA annotation

RNA manipulation

Remove RNA

A RNA subset can be easily analyzed by indicating the RNA to remove. The user can thus create a new RiboClass object containing the data of the RNAs of interest, without affecting the samples' metadata.

Here is an example where the RNA 5S is removed:

```

ribo_adjusted <- remove_rna(ribo, rna_to_remove = "NR_023363.1_5S")
print(ribo_adjusted)
#> a RiboClass with 14 samples and 3 RNA(s) :
#> Name : NR_046235.3_5.8S, length : 157
#> Name : NR_046235.3_18S, length : 1869
#> Name : NR_046235.3_28S, length : 5070

```

Rename RNA

Annotation of rRNA 2'Ome sites using the lists provided by this package requires the usage of specific RNA names.

Here is an example to check whether the RNA names provided by the user in the RiboClass match the ones used by this package :

```

data("human_methylated")
cat("human_methylated's rna names: ", unique(human_methylated$rRNA), "\n")
#> human_methylated's rna names: 5.8S 18S 28S
cat("ribo's rna names: ", as.character(ribo_adjusted$rna_names$current_name))
#> ribo's rna names: NR_046235.3_5.8S NR_046235.3_18S NR_046235.3_28S

```

In this example, the names are different and need to be updated before annotation.

The function `rename_rna` updates automatically the rRNA names that are given by rRNA size order:

```

ribo_adjusted <- rename_rna(ribo_adjusted,
                           new_names = c("5.8S", "18S", "28S"))
# from the shortest RNA in our RiboClass to the longest.

```

Annotation of RNA 2'Ome sites

This package computes a C-score for each genomic position of the RNAs. The C-scores of the RNA 2'Ome sites of interest are then extracted using a list of annotated sites and the `annotate` function.

Annotation included: Human 2'Ome rRNA sites

By default, this package includes two dataframes including the positions and the annotations of the human rRNA 2'Ome sites:

- `human_methylated`: a dataframe, containing the 112 known 2'Ome sites for the human rRNAs.
- `human_suspected`: a dataframe, containing the 17 sites that are putative 2'Ome sites for the human rRNAs, as described in the literature.

Customize 2'Ome sites annotations

Instead of using the list of human rRNA 2'Ome sites included with this package, the user can provide its own list and annotation using the function `annotate_site`.

A dataframe with three mandatory columns should be provided :

- **RNA name**: the name of the RNA, matching the RNA name of the RiboClass.
- **Position on RNA**: the number of the position on the RNA.
- **Nomenclature**: the name given to the site of interest.

You can see an example below :

Position	rRNA	Nomenclature
15	5.8S	Um14
76	5.8S	Gm75
28	18S	Am27

Annotate RNA sites

The 2'Ome sites of interest can be annotated using the `annotate` function with either the provided annotations or customized ones.

Here is an example with the provided human methylated annotations:

```
ribo_adjusted <- annotate_site(ribo_adjusted,
                             annot = human_methylated,
                             anno_rna = "rRNA",
                             anno_pos = "Position",
                             anno_value = "Nomenclature")
```

Note: if an issue occurs during the annotation process, the user should check the Rename RNA section.

This vignette has also some explanations on how to create your own sites annotation dataset with Customize 2'Ome sites annotations.

Data visualization with PCA

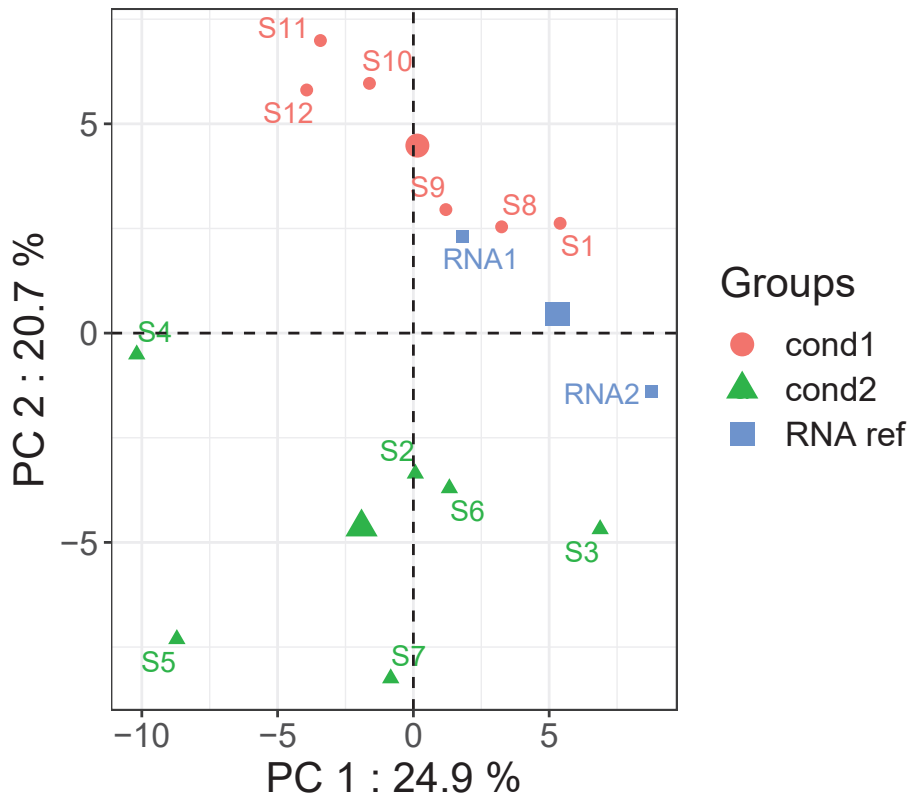
The function `plot_PCA`, which returns a ggplot, has been implemented with several parameters for more flexibility:

1. **only_annotated**: plot samples based on the annotated RNA 2'Ome sites only to determine whether samples clustered depending on their rRNA 2'Ome profile.

Here is an example comparing samples reflecting different biological conditions based on the rRNA 2'Ome profile of the provided `human_methylated` list:

```
plot_PCA(ribo_adjusted,
         color_col = "condition",
         only_annotated = TRUE)
```

PCA of Cscore for 112 sites 14 samples

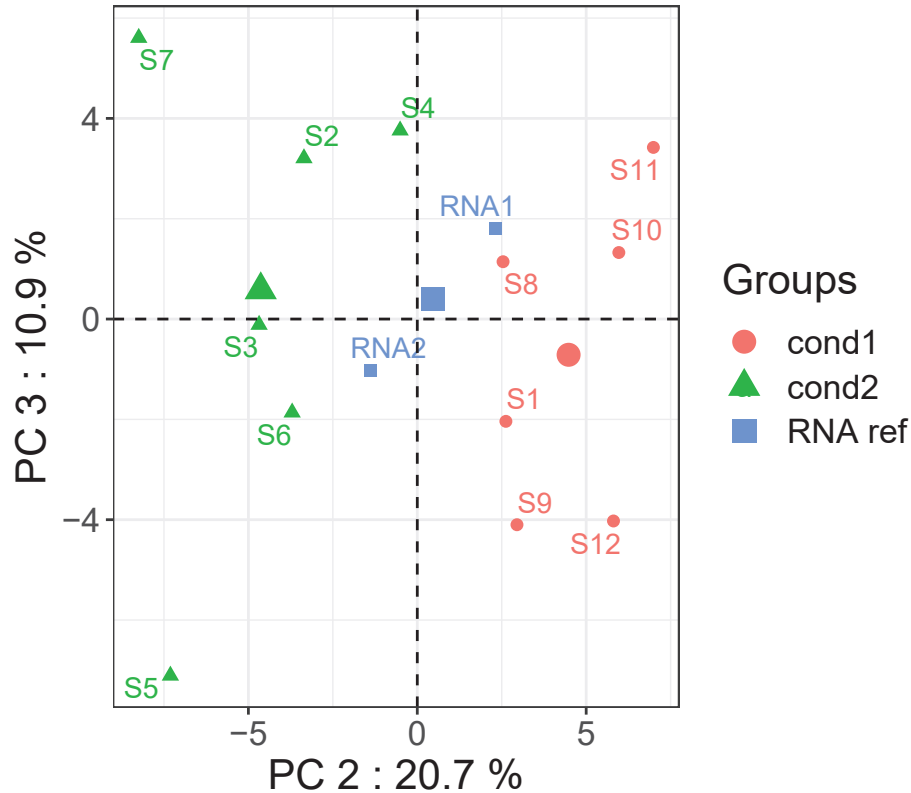


2. **axes**: by default, PC1 and PC2 axes are plotted. However, the user can choose the PCA axes of interest using the “axes” parameter.

```
plot_PCA(ribo_adjusted,
         color_col = "condition",
         axes = c(2,3), #PC2 and PC3 will be plotted
         only_annotated = TRUE)
```

PCA of Cscore for 112 sites

14 samples



3. `pca_object_only`: return the full `dudi.pca` object by setting `pca_object_only` to `True`:

```
pca <- plot_PCA(ribo_adjusted,  
  color_col = "condition",  
  only_annotated = TRUE,  
  pca_object_only = TRUE)
```

Export data

Data can be exported as two different objects.

Export as a dataframe

The user can export data as a dataframe using the function `extract_data`.

By default, it will export c-score for all the genomic RNA positions.

```
ribo_df <- extract_data(ribo_adjusted,  
  col = "cscore")
```

The user can export data related to the subset of annotated RNA 2'Ome sites by setting the `only_annotated` parameter to `True`.

```
ribo_df <- extract_data(ribo_adjusted,  
  col = "cscore",  
  only_annotated = TRUE)
```

```
only_annotated = TRUE)
```

Table 6: excerpt from the output dataframe. S1, S2 and S3 are samples.

site	S1	S2	S3
18S_Am27	0.9783333	0.9745514	0.9810412
18S_Am99	0.9680204	0.9646470	0.9724150
18S_Um116	0.9276274	0.9470968	0.9407083
18S_Um121	0.9630216	0.9684459	0.9691576
18S_Am159	0.9629730	0.9620986	0.9686766
18S_Am166	0.9809492	0.9744627	0.9744998
18S_Um172	0.9510189	0.9400922	0.9517981
18S_Cm174	0.9119119	0.8668456	0.8806886
18S_Um354	0.9709273	0.9722054	0.9756174
18S_Um428	0.9268668	0.9276018	0.9504224

Export as a ggplot-ready dataframe

The user can export data as a ggplot-ready dataframe using the function `format_to_plot`.

By default, it will export c-score for all the genomic RNA positions. The user can export additional information present in the metadata by indicating the name of the column of interest. The user can export information related to the subset of annotated RNA 2'Ome sites by setting the `only_annotated` parameter to `True`.

Here is an example of ggplot-ready dataframe including the C-scores of all the genomic RNA positions as well as the condition related to the particular sample of interest:

```
ggplot_table <- format_to_plot(ribo_adjusted,"condition")
```

Table 7: Excerpt from the output ggplot-ready dataframe

	sampleID	Cscore	condition
501	RNA1	0	RNA ref
502	RNA1	0.282380396732789	RNA ref
503	RNA1	0.110590988882387	RNA ref
504	RNA1	0.615841584158416	RNA ref
505	RNA1	0	RNA ref
506	RNA1	0	RNA ref
507	RNA1	0	RNA ref
508	RNA1	0.507692307692308	RNA ref
509	RNA1	0.118110236220472	RNA ref
510	RNA1	0.0637554585152839	RNA ref

Export as a dataframe by condition

The user can export a dataframe compiling the mean C-scores of each position by specific conditions provided in the metadata dataframe using the `mean_samples_by_conditon` function.

By default, it will export for all the genomic RNA positions the name of the position, the mean and sd of the C-scores. The user can also export mean and sd of read end count. The user can export information related to the subset of annotated RNA 2'Ome sites by setting the `only_annotated` parameter to `True`.

Here is an example of dataframe including the mean C-scores per conditions for all the genomic RNA positions:

```
mean_tb <- mean_samples_by_conditon(ribo_adjusted,
                                   value = "cscore",
                                   metadata_condition = "condition",
                                   only_annotated = TRUE)
```

Table 8: Excerpt from the output dataframe by condition

site	condition	mean	sd
18S_Am1031	cond1	0.9703699	0.0032725
18S_Am1031	cond2	0.9723714	0.0057723
18S_Am1031	RNA ref	0.9749922	0.0047461
18S_Am1383	cond1	0.9765212	0.0027083
18S_Am1383	cond2	0.9779684	0.0028759
18S_Am1383	RNA ref	0.9796876	0.0035939
18S_Am159	cond1	0.9609920	0.0054882
18S_Am159	cond2	0.9623362	0.0065239
18S_Am159	RNA ref	0.9684126	0.0022025
18S_Am166	cond1	0.9791230	0.0027282

Sessioninfo

```
sessionInfo()
#> R version 4.2.1 (2022-06-23)
#> Platform: x86_64-pc-linux-gnu (64-bit)
#> Running under: Ubuntu 20.04.5 LTS
#>
#> Matrix products: default
#> BLAS: /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.9.0
#> LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.9.0
#>
#> locale:
#> [1] LC_CTYPE=fr_FR.UTF-8          LC_NUMERIC=C
#> [3] LC_TIME=fr_FR.UTF-8          LC_COLLATE=fr_FR.UTF-8
#> [5] LC_MONETARY=fr_FR.UTF-8     LC_MESSAGES=fr_FR.UTF-8
#> [7] LC_PAPER=fr_FR.UTF-8        LC_NAME=C
#> [9] LC_ADDRESS=C                 LC_TELEPHONE=C
#> [11] LC_MEASUREMENT=fr_FR.UTF-8  LC_IDENTIFICATION=C
#>
#> attached base packages:
#> [1] stats      graphics  grDevices  utils      datasets  methods   base
#>
#> other attached packages:
#> [1] Riboscore_1.0.0
#>
#> loaded via a namespace (and not attached):
#> [1] Biobase_2.56.0      httr_1.4.3          edgeR_3.38.2
#> [4] tidyrr_1.2.0        splines_4.2.1       bit64_4.0.5
#> [7] carData_3.0-5       assertthat_0.2.1    highr_0.9
#> [10] stats4_4.2.1        blob_1.2.3          GenomeInfoDbData_1.2.8
#> [13] yaml_2.3.5          ggrepel_0.9.1       lattice_0.20-45
#> [16] factoextra_1.0.7    pillar_1.8.0        RSQLite_2.2.15
```



```

#> [19] backports_1.4.1      limma_3.52.2         glue_1.6.2
#> [22] digest_0.6.29       ggsignif_0.6.3      XVector_0.36.0
#> [25] colorspace_2.0-3    Matrix_1.4-1        htmltools_0.5.3
#> [28] XML_3.99-0.10      pkgconfig_2.0.3     broom_1.0.1
#> [31] genefilter_1.78.0  zlibbioc_1.42.0     purrr_0.3.4
#> [34] xtable_1.8-4        scales_1.2.0        BiocParallel_1.30.3
#> [37] tibble_3.1.8        annotate_1.74.0      mgcV_1.8-40
#> [40] KEGGREST_1.36.3    generics_0.1.3     farver_2.1.1
#> [43] IRanges_2.30.0     car_3.1-0           ggplot2_3.3.6
#> [46] ellipsis_0.3.2     ggpubr_0.4.0       cachem_1.0.6
#> [49] BiocGenerics_0.42.0 cli_3.3.0           survival_3.2-13
#> [52] magrittr_2.0.3     crayon_1.5.1        memoise_2.0.1
#> [55] evaluate_0.15      fansi_1.0.3         nlme_3.1-157
#> [58] MASS_7.3-58        rstatix_0.7.0       tools_4.2.1
#> [61] matrixStats_0.62.0 lifecycle_1.0.1     stringr_1.4.0
#> [64] S4Vectors_0.34.0  locfit_1.5-9.6      munsell_0.5.0
#> [67] AnnotationDbi_1.58.0 Biostrings_2.64.0   ade4_1.7-19
#> [70] compiler_4.2.1     GenomeInfoDb_1.32.2 rlang_1.0.4
#> [73] grid_4.2.1         RCurl_1.98-1.8     bitops_1.0-7
#> [76] labeling_0.4.2     rmarkdown_2.15     gtable_0.3.0
#> [79] codetools_0.2-18  abind_1.4-5         DBI_1.1.3
#> [82] R6_2.5.1           knitr_1.39          dplyr_1.0.9
#> [85] fastmap_1.1.0     bit_4.0.4           utf8_1.2.2
#> [88] stringi_1.7.8     parallel_4.2.1     sva_3.44.0
#> [91] Rcpp_1.0.9         vctrs_0.4.1        png_0.1-7
#> [94] tidyselect_1.1.2  xfun_0.31

```

Reference

- Birkedal, Ulf, Mikkel Christensen-Dalsgaard, Nicolai Krogh, Radhakrishnan Sabarinathan, Jan Gorodkin, and Henrik Nielsen. 2014. “Profiling of Ribose Methylations in Rna by High-Throughput Sequencing.” *Angewandte Chemie International Edition*, November, n/a–n/a. <https://doi.org/10.1002/anie.201408362>.
- Erales, Jenny, Virginie Marchand, Baptiste Panthu, Sandra Gillot, Stéphane Belin, Sandra E. Ghayad, Maxime Garcia, et al. 2017. “Evidence for rRNA 2′-O-Methylation Plasticity: Control of Intrinsic Translational Capabilities of Human Ribosomes.” *Proceedings of the National Academy of Sciences* 114 (49): 12934–9. <https://doi.org/10.1073/pnas.1707674114>.
- Marcel, Virginie, Janice Kielbassa, Virginie Marchand, Kundhavai S Natchiar, Hermes Paraqindes, Flora Nguyen Van Long, Lilia Ayadi, et al. 2021. “Ribosomal RNA 2′ O-methylation as a novel layer of inter-tumour heterogeneity in breast cancer.” *NAR Cancer* 3 (1). <https://doi.org/10.1093/narcan/zcab006>.
- Marchand, Virginie, Florence Blanloeil-Oillo, Mark Helm, and Yuri Motorin. 2016. “Illumina-based RibomethSeq approach for mapping of 2′-O-Me residues in RNA.” *Nucleic Acids Research* 44 (16): e135–e135. <https://doi.org/10.1093/nar/gkw547>.
- Zhang, Yuqing, Giovanni Parmigiani, and W Evan Johnson. 2020. “ComBat-Seq: Batch Effect Adjustment for Rna-Seq Count Data.” *NAR Genomics and Bioinformatics* 2 (3). <https://doi.org/10.1093/nargab/lqaa078>.

B Article I : snoDB 2.0 : An enhanced interactive database, specializing in human snoRNAs

Ce travail, dont je suis second auteur, a été publié en septembre 2022 dans le journal Nucleic Acid Research (NAR). L'article et les figures supplémentaires sont inclus après cette page introductive.

B.1 Contexte

Les petits ARN nucléolaires (snoRNA) sont des ARN non-codants de 60 à 300 nucléotides qui se retrouvent principalement dans le nucléole. Chez l'Homme, la majorité des snoRNA provient des introns de gènes codants appelés «gènes hôtes». Ils sont divisés en deux grandes classes, les snoRNA de type C/D et les snoRNA de type H/ACA, dont la fonction canonique est de guider respectivement la méthylation en 2' du ribose (2'Ome) ou la pseudo-uridylation des ARN ribosomiques (ARNr). Au cours des deux dernières décennies, plusieurs fonctions non canoniques dans divers mécanismes moléculaires impliqués dans la régulation de l'expression génique ont été identifiées, mais seulement pour une minorité de snoARN qui peuvent interagir par exemple avec des protéines ou des ARNm.

Les snoARNs restent encore peu étudiés. Cela est dû en partie à des difficultés techniques pour détecter et quantifier ces petits ARN à partir des données de transcriptomique. En effet, les snoARNs sont présents en un grand nombre de copies dans le génome et sont hautement structurés. En outre, les informations de qualité utiles sur les snoARN sont souvent absentes ou difficiles à trouver dans les ressources publiques, générales ou spécialisées. Malgré l'intérêt biologique croissant, une grande partie des informations concernant les snoARN sont désorganisées et dispersées, dans différentes bases de données accessibles en ligne ou dans la littérature. Jusqu'en 2018, trois ressources dédiées aux snoARN chez l'Homme étaient disponibles : snoRNA-Base (LESTRADE, 2006), snOPY (YOSHIHAMA, NAKAO et KENMOCHI, 2013) et snoRNA Atlas (JORJANI et al., 2016). En plus d'informations strictement liées aux séquences des ARNr, snoRNABase contient des informations sur le gène hôte, la séquence des snoARN et les sites cibles des snoARN sur les

ARNr. snOPY se focalise principalement sur l'orthologie entre les espèces eucaryotes et snoRNA Atlas fournit des informations sur la localisation chromosomique, la conservation et l'abondance moyenne des snoARN. Néanmoins, ces bases de données ne regroupent pas l'ensemble des informations utiles, de nombreux snoARN humains ne sont pas répertoriés dans ces bases, et il y a des manques concernant leurs gènes hôtes et leurs cibles qui ont été récemment annotés. Par exemple, snoRNABase n'a pas été mise à jour depuis 2007 et, comme snoRNA Atlas, ne respecte plus les normes de sécurité et de confidentialité définies par les navigateurs récents. Il existe donc un besoin urgent d'unifier les données actuellement disponibles sur les snoARN humains afin d'optimiser les efforts de recherche futurs.

En 2018, l'équipe de Michelle Scott de l'université de Sherbrooke, Canada, a développée snoDB, une base de données interactive spécialisée dans les snoARN humains (BOUCHARD-BOURELLE et al., 2019). snoDB offre des informations actualisées obtenues à partir d'autres bases de données, des résultats de la littérature et des données expérimentales sur l'abondance des snoRNA générées par le protocole TGIRT-seq qui a été spécifiquement développé pour la quantification des petits ARN très structurés, comme les snoARNs (BOIVIN et al., 2018; DESCHAMPS-FRANCOEUR et al., 2019; NOTTINGHAM et al., 2016).

B.2 Résultats principaux

En 2020, nous avons initié une collaboration avec l'équipe de Michelle Scott avec pour objectif de partager nos connaissances respectives sur les snoRNA afin de mieux comprendre leurs fonctions canoniques et non canoniques. Lors de cette collaboration, j'ai eu l'occasion de participer à la mise à jour de la base de données snoDB. snoDB 2.0 inclut entre autres :

- (i) des snoRNA nouvellement découverts,
- (ii) des données d'expression dans 9 tissus et 5 lignées cellulaires,
- (iii) des prédictions d'interactions snoRNA-protéines,
- (iv) des prédictions de la structure secondaire ainsi que les motifs des snoRNA (boîte H/ACA et C/D) et de la localisation des régions fonctionnelles,
- (v) des informations sur les différentes copies de chaque snoRNA avec une correction des duplicats,

- (vi) une toute nouvelle section interactive sur les interactions canoniques snoRNA-ARNr, qui permet non seulement de trouver facilement les sites de modification pour chaque snoRNA (ou au contraire, les snoRNA qui modifient un site spécifique), mais aussi de visualiser de manière interactive les données obtenues par de récentes analyses à haut débit du niveau de modification des bases (2'-O-méthylation ou pseudouridilation) dans les ARNr (Fig. 1 de l'article).

La totalité des données peut être téléchargée. snoDB 2.0 est accessible sur le site : <https://bioinfo-scottgroup.med.usherbrooke.ca/snoDB/>

En conclusion, la mise à jour de cette base de données a permis de centraliser et de mettre à jour l'ensemble des informations concernant les snoARNs. De plus, cette nouvelle version contient maintenant des informations et des outils clés pour l'analyse des modifications chimiques des ARNr, dont la 2'Ome des ARNr. Ainsi, cette base de données incluant des informations sur la 2'Ome des ARNr permettra de diffuser nos savoirs-faires, connaissances et données sur la 2'Ome des ARNr issus de nos différents projets.

B.3 Ma participation

Une des grandes nouveautés de cette deuxième version de snoDB concerne l'addition des modifications chimiques de l'ARNr (i.e. pseudo-uridilation et 2'Ome) qui est visible dans la toute nouvelle section « rRNA Chemical Modification ». Cette section a bénéficié de tout le travail que j'avais réalisé au cours de mon Master 2 et de ma première année de thèse, et qui était nécessaire pour mettre en place le pipeline de méthodologie et d'analyse décrit en section A et A.1.

J'ai tout d'abord fourni une liste à jour des sites de modifications chimiques des ARNr que j'avais compilées à partir de la base snoRNAbase et de la littérature (onglet « rRNA modification »). Lors de cette étude, je me suis aperçu que dans la littérature, les études dédiées à l'analyse le taux de méthylation ou de pseudouridilation des ARNr utilisent différentes séquences de référence des ARNr (Table S1). Par conséquent, la nomenclature des sites peut différer d'une étude à l'autre. J'ai donc fourni les différentes séquences de référence des ARNr utilisées dans la littérature (onglet « rRNA sequence »). L'utilisateur peut ainsi visualiser ces différentes séquences sur lesquelles les nucléotides 2'Ome ou pseudouridilés ont été positionnés. J'ai également créé

une table de conversion qui comprend ces différentes versions de séquences de référence des ARNr (onglet « Conversion table »). Cette table contient la correspondance entre toutes les positions des séquences de référence des ARNr mais aussi les positions modifiées chimiquement. Une fonctionnalité a été ajoutée afin de pouvoir rechercher une position d'ARNr d'intérêt. Ce tableau permet donc de palier les problèmes de nomenclature des sites modifiés entre les différentes études. Enfin, j'ai fourni les données de Cscore des 106 sites de 2'Ome des ARNr pour 195 tumeurs mammaires et 4 lignées cellulaires de cancer du sein. Ces données proviennent de l'étude sur la variabilité de la 2'Ome dans les cancers du sein (cf. C). Les données peuvent être visualisées de manière interactive par l'utilisateur (Fig 1 et S2 de l'article).

En plus de fournir les données pour enrichir cette base, j'ai participé aux discussions afin de définir les besoins côté utilisateurs et de faire un retour sur la partie dédiée aux snoARN.

B.4 Article

snoDB 2.0: an enhanced interactive database, specializing in human snoRNAs

Danny Bergeron¹, Hermes Paraqindes^{2,3,4}, Étienne Fafard-Couture¹, Gabrielle Deschamps-Francoeur¹, Laurence Faucher-Giguère⁵, Philia Bouchard-Bourelle¹, Sherif Abou Elela⁵, Frédéric Catez^{2,3,4,6}, Virginie Marcel^{2,3,4,6} and Michelle S. Scott^{1,*}

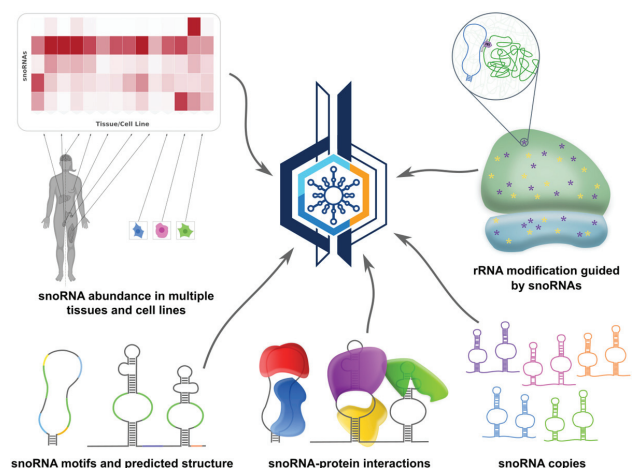
¹Département de biochimie et génomique fonctionnelle, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec J1E 4K8, Canada, ²Inserm U1052, CNRS UMR5286 Centre de Recherche en Cancérologie de Lyon, F-69000 Lyon, France, ³Centre Léon Bérard, F-69008 Lyon, France, ⁴Université de Lyon 1, F-69000 Lyon, France, ⁵Département de microbiologie et d'infectiologie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec J1E 4K8, Canada and ⁶Institut Convergence PLASCAN, F-69373 Lyon, France

Received August 09, 2022; Revised September 06, 2022; Editorial Decision September 12, 2022; Accepted September 16, 2022

ABSTRACT

snoDB is an interactive database of human small nucleolar RNAs (snoRNAs) that includes up-to-date information on snoRNA features, genomic location, conservation, host gene, snoRNA–RNA targets and snoRNA abundance and provides links to other resources. In the second edition of this database (snoDB 2.0), we added an entirely new section on ribosomal RNA (rRNA) chemical modifications guided by snoRNAs with easy navigation between the different rRNA versions used in the literature and experimentally measured levels of modification. We also included new layers of information, including snoRNA motifs, secondary structure prediction, snoRNA–protein interactions, copy annotations and low structure bias expression data in a wide panel of tissues and cell lines to bolster functional probing of snoRNA biology. Version 2.0 features updated identifiers, more links to external resources and duplicate entry resolution. As a result, snoDB 2.0, which is freely available at <https://bioinfo-scottgroup.med.usherbrooke.ca/snoDB/>, represents a one-stop shop for snoRNA features, rRNA modification targets, functional impact and potential regulators.

GRAPHICAL ABSTRACT



INTRODUCTION

Small nucleolar RNAs (snoRNAs) are small non-coding RNAs ranging in size from 60 to 300 nucleotides and encoded in all eukaryotic genomes annotated so far. Based on their characteristic RNA motifs and protein-binding partners, they are classified into two groups, the box C/D snoRNAs and the box H/ACA snoRNAs, best characterized for their role in guiding, respectively, the 2'-O-ribose methylation and the pseudouridylation of ribosomal RNAs (rRNAs) and small nuclear RNAs (snRNAs). However, in the last two decades, a handful of snoRNAs have been shown to be involved in the regulation of a multitude of other cellular processes such as alternative splicing (1–3), 3' end processing (4), exosome recruitment (5) and chromatin

*To whom correspondence should be addressed. Tel: +1 819 821 8000 (Ext 72123); Email: michelle.scott@usherbrooke.ca

remodelling (6) through either RNA or protein interactions. Moreover, a plethora of studies found that snoRNAs are dysregulated in many cancer types and other diseases (7–11).

The study of snoRNAs and their complex family relationships involves several technical challenges. These include the very stable structure of snoRNAs and, more specific to mammals, their high copy numbers in genomes and their genomic location in introns of longer genes referred to as their host genes (12–14). These challenges are compounded by other factors related to our limited knowledge of these small RNAs and the few reliable datasets detecting them and their targets accurately. For example, the functions of chemical modification in rRNAs guided by snoRNAs, while being extensively studied, are not yet fully understood [e.g. (15–20)]. Indeed, for a long time, it was accepted that these chemical modifications are present on all rRNA molecules to maintain their conserved structures. However, with the emergence of advanced techniques to detect the level of chemical modification in rRNAs, such as RiboMethSeq (21) or HydraPsiSeq (22), it was recently shown that some positions are only partially modified and that the level of modification could vary in different cell lines, conditions or in certain pathologies (15,17,20,22–24). Furthermore, although there is currently a massive amount of publicly available RNA-seq datasets, few provide a reliable and realistic quantification for snoRNAs. TGIRT-seq, a sequencing method using a thermostable reverse transcriptase, is one such method that performs well for highly structured RNAs such as transfer RNAs and snoRNAs (12,25,26).

For the study of human snoRNAs, four dedicated databases are currently publicly available. snoRNABase (27), available since 2005, provides information about snoRNA sequences, rRNA and snRNA targets as well as their host genes, but is no longer regularly updated and is lacking many human snoRNAs present in other resources. snoRNA Atlas provides information about chromosomal location, conservation, average abundance and targets (28), while snOPY is focused on snoRNA orthology relationships across eukaryotes (29). As a complementary resource, we developed in 2018 snoDB (30), an interactive specialized human snoRNA database offering up-to-date information obtained from general and snoRNA-specific resources as well as our low structure bias TGIRT-seq datasets. We now provide snoDB 2.0, a second version with updated identifiers and updated links to other resources, correction for duplicate entries and most importantly diverse new data presented in an interactive format. snoDB now includes snoRNA–protein interaction data, information on the different copies of each snoRNA, snoRNA motifs and domains, secondary structure prediction, TGIRT-seq abundance data for 9 tissues and 5 cell lines (for a total of 37 samples) and a brand new section on rRNA modifications guided by snoRNAs.

DATABASE CONTENT

To ensure that the information is as up to date as possible, we obtained data from the latest versions of all the resources that were originally considered for the first version of snoDB. One of the problems we faced with the

original version was the redundancy of some entries (Supplementary Figure S1). Indeed, some overlapping snoRNAs having as few as one nucleotide difference in 5', 3' or both were present as two or more distinct entries in snoDB (as well as in other resources), with their own sets of IDs and missing information. To solve this problem in version 2.0, we merged overlapping entries, ensuring that one genomic location was associated with only one snoRNA. Because we had no way of knowing the true beginning and end of a snoRNA, we prioritized information (e.g. genomic location, gene name, etc.) from some resources over others. We arbitrarily prioritized, in the following order, Ensembl (V104) (31), RefSeq (V109) (32), snoRNA Atlas (July 2021) (28) and snOPY (July 2021) (29) annotations. Ensembl itself has duplicate entries for some snoRNAs in its annotation file. In these cases, we chose snoRNAs starting with 'SNOR' first; otherwise, we prioritized the longest of the entries. In all cases, we used a concatenation of the IDs (e.g. ENSG00000281910;ENSG00000206952) to help the user to easily identify their snoRNAs of interest. We also included links (all obtained in July 2021) from HGNC (33), NCBI (34), snoRNABase (27), RNAcentral (35) and Rfam (36). When possible, we used the Infernal package to add missing Rfam IDs as recommended by Rfam (36). snoDB 2.0 now contains 2123 non-redundant entries.

Along with all the features that were originally included in the first version of snoDB (30), we now have additional evolutionary conservation information, sequence annotations, a predicted secondary structure, protein interaction data and copy information for all snoRNAs in snoDB as well as the level of abundance of snoRNAs and of their host gene. snoDB now presents a conservation score that was extracted from the UCSC Genome Browser, PhastCons track for 100 vertebrates (37,38) for each snoRNA. In addition, each entry now also features a secondary structure predicted using RNAfold and RNAplot from the Vienna package (39) in which the predicted snoRNA sequence motifs and guide regions can be highlighted. snoDB already contained information on RNA interacting partners of snoRNAs, to which we have added a section on snoRNA–protein interactions, extracted from eCLIP datasets of 150 RNA-binding proteins from the Encyclopedia of DNA Elements (ENCODE) Consortium (40). We also added a section on family members (based on Rfam classification), which is important when analysing snoRNAs by PCR, for example, since such snoRNAs often have very similar sequences. Finally, the abundance of snoRNAs and their host genes was obtained from 37 datasets that use the low structure bias TGIRT-seq approach (12,22,23) (GEO entries GSE126797, GSE157846, GSE99065 and GSE209924, and SRA entries SRX1426160 and SRX1426193). For all these newly added characteristics, more details are available in the About/Help section of the database.

To improve snoRNA quantification from RNA-seq datasets, we create a simple tool called snoRupdate. Indeed, to avoid biases due to the lack of some snoRNAs from current annotation files (e.g. gene transfer format), it is important to complete these annotations before performing RNA-seq analysis. For this purpose, snoRupdate, accessible from the main page or directly from GitHub (<https://github.com/scottgroup/snoRupdate>), is designed to

add all missing snoRNAs from snoDB to the provided annotation file.

Finally, in this second version of the database, we added a completely new section on chemical modifications of rRNAs, which includes snoRNA–rRNA interaction sites, widely used rRNA sequences with conversion between them and modification levels for known modified rRNA positions. The rRNA modification sites were taken from snoRNABase (27), snoRNA Atlas (28), Krogh *et al.* (41) and Kehr *et al.* (42). Modification site annotations are dependent on rRNA sequences, and several versions of rRNAs are used in the literature (Supplementary Table S1). Lastly, rRNA modification levels were extracted from four studies using RiboMethSeq (17,23), HydraPsiSeq (22) or SILNAS (24).

WEB INTERFACE

Main page

On the main page, we made changes taking care to keep the interface similar enough to allow a smooth transition for users of the previous version. The navigation bar, on the top right panel, now includes a change log page, which allows to track changes made in future versions and to have access to old versions, a new rRNA chemical modification section (discussed further below), a link to snoRupdate (described above) and finally a sequence search tool section embedding a widget designed by RNACentral (35) that enables the comparison of an input sequence against a large collection of non-coding RNAs.

The bottom of the page presents the main snoRNA interactive table, the columns of which can be controlled using the ‘Column visibility’ button. In version 2.0, we generated unique snoDB identifiers and use them for navigation through the website. The table can be easily filtered using the multiple search fields (see the tutorial section for advanced usage). As in the previous version, all the information in the main table, or only a selection, can be exported in different formats (xlsx, tsv, bed and now in fasta). Individual rows can also be selected to export only a part of the visible data.

snoRNA abundance visualization page

snoDB now provides abundance data from a considerably larger number of samples (9 tissues and 5 cell lines consisting in total of 37 datasets). We thus dedicated an independent page to the visualization of abundance, accessible through a button at the top right of the table of the main page, employing the Clustergrammer-JS and Clustergrammer-PY libraries (43), to provide an interactive heatmap. From the main table, the abundance profiles for all entries, or only a subpart (filtered and/or selected), can be observed simultaneously. The abundance data of all 37 samples can also be easily downloaded from this page.

Individual snoRNA pages

Aside from the main page, snoDB also has an individual page dedicated to each snoRNA. In this detailed view, all the information contained in the table from the main page

on a particular snoRNA is recorded. In this updated version, it is possible to visualize important motifs (boxes and guide regions) of a snoRNA, and its predicted secondary structure by toggling two buttons in the first section. Furthermore, since snoRNA copies can often be a problem, as they are not so easy to find and are thus often overlooked, we included in this page a section listing all copies of a snoRNA. These copies can also be further investigated by clicking on ‘Inspect copies in main table’. When redirected, the information for the different entries can be exported in different formats and the level of abundance of all the snoRNA copies can be viewed in a heatmap. Next, in addition to snoRNA–RNA interactions, we also added a section dedicated to snoRNA–protein interactions. This section lists the RNA-binding proteins reported to bind to the snoRNA, the cell line in which the experiment was conducted, a *P*-value for the interaction and a score [obtained from ENCODE (40)]. Finally, instead of using a table for the level of abundance of the snoRNA and its host gene, they are now available as interactive box plots.

rRNA chemical modifications

One of the significant additions of the snoDB 2.0 new release is the rRNA modification section, which contains four subsections accessible via the different top tabs: rRNA modifications, rRNA sequences, conversion table and modification levels (Figure 1A). The rRNA modification subsection includes an interactive table of all listed snoRNA–rRNA interactions. The status of the modification (i.e. whether the modification site has been validated by one or more techniques) is also mentioned to avoid confusion due to some predicted sites not being modified. The whole, or only the desired part, can be copied or exported in csv or xlsx formats. Next, the rRNA sequence tab enables visualization of the most widely used rRNA sequences. Validated modified positions are annotated and color-coded [2'-*O*-ribose methylation (Nm), pseudouridylation (Ψ) or both], and clicking on any of these modified sites returns snoRNAs predicted to modify this position, as well as a clickable link to the detailed view for the snoRNAs. The third subsection is a conversion table between different versions of rRNAs used in different articles and references. Any position can be searched, and the table contains the position in each different rRNA reference sequence as well as the nucleotide at this specific position. The whole table or a part can also be copied or exported in csv or xlsx formats. The last tab refers to a section where one can visualize the level of modification observed by different techniques and independent groups, in different tissues and cell lines (Figure 1), with the provision that rigorous comparisons cannot be made between the datasets because different methods were employed. By selecting the modification type (2'-*O*-ribose methylation or pseudouridylation; Figure 1B), the rRNA molecule (18S, 28S or 5.8S; Figure 1C) and the study (all the studies or only a single one; Figure 1D), the user sees a box plot displaying a compilation of all the modification levels for each position (Figure 1E). From there, it is possible to search a position with the top left input field (or by clicking on the chart; Figure 1F), reorder the rRNA positions according to position, coefficient of variation or mean

rRNA chemical modifications

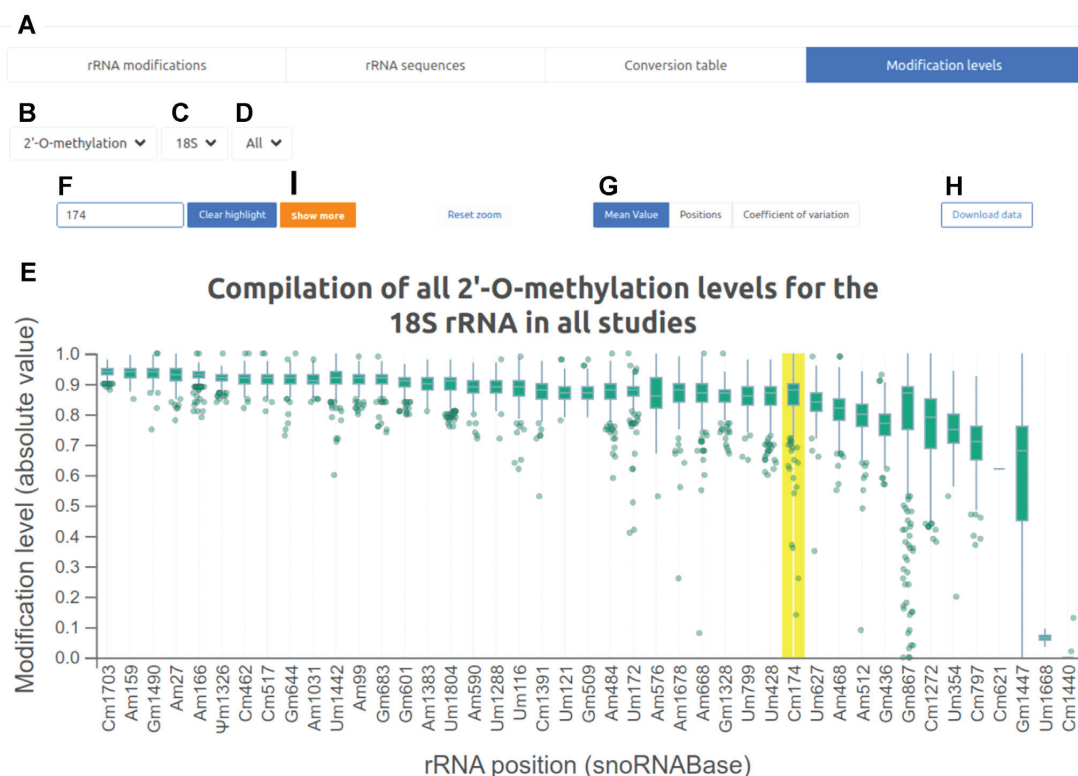


Figure 1. Screenshot of the ‘Modification levels’ tab, one of the four subsections of the new rRNA chemical modification section. (A) Four tabs corresponding to each of the new subsections: rRNA modifications, rRNA sequences, conversion table and modification levels. (B–D) Modification type (2'-O-ribose methylation or pseudouridylation), rRNA molecule (18S, 28S or 5.8S) and study (all the studies or only a single one) to choose from. (E) Box plot displaying a compilation of the levels of all modified positions. (F) Search field used to look for a specific modified position. (G) Buttons to reorder the box plot according to mean value, position or coefficient of variation. (H) ‘Download data’ button to download the raw data used to generate the chart. (I) ‘Show more’ button used to get a more detailed view of the level of modification of a single position (see Supplementary Figure S2).

value (default sorting) (Figure 1G) and change the zoom of the chart. The entire dataset used to generate the box plot can be exported using the ‘Download data’ button (Figure 1H). When a position is highlighted in the chart, an orange button appears (Figure 1I), which provides the user with a more detailed view of the level of modification, in a dot plot representation, of a single site across different previously selected cell lines and studies (Supplementary Figure S2).

CONCLUSION AND FUTURE PLANS

snoDB aims to provide the community with an efficient way to navigate through the most up-to-date snoRNA data. This second version of snoDB has kept its user-friendly interface with its main overview table and individual snoRNA pages, but on top of the consolidation and addition of missing snoRNAs, many more characteristics of snoRNAs are now available, including sequence motifs and structures, copy identification, considerably more abundance data and snoRNA–protein interactions. We also added a completely new section on rRNA modifications. Our future plans, in addition to periodic updates of the current information contained in the database, include providing the same level of information on snoRNAs from other species.

DATA AVAILABILITY

snoDB 2.0 is freely available at <https://bioinfo-scottgroup.med.usherbrooke.ca/snoDB/>.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful to members of their groups for useful discussions, to Leandro Fequino for technical support and to Yuri Motorin and Virginie Marchand for providing their ribosome modification datasets. M.S.S. and S.A.E. are members of the RNA group and the Centre de Recherche du Centre Hospitalier Universitaire de Sherbrooke (CRCHUS). V.M. is an Institut National de la Santé et de la Recherche Médicale (Inserm) fellow and F.C. is a Centre National de Recherche Scientifique (CNRS) fellow.

FUNDING

Canadian Institutes of Health Research [PJT-153171 to M.S.S. and S.A.E.]; Fonds de Recherche du

Québec—Nature et Technologies [2022-PR-298214 to M.S.S. and S.A.E.; to G.D.-F.]; Fonds de Recherche du Québec—Santé [296343 to M.S.S.]; Fondation ARC pour la Recherche sur le Cancer [20161204686 MARACAS.v0 to V.M. and F.C.]; SIRIC program [INCa-DGOS-Inserm 12563 LyRICAN to V.M. and F.C.]; Labex program [DEVCAN2UMAN to V.M. and F.C.]; Synergie Lyon Cancer Foundation [to V.M. and F.C.]; Agence Nationale de la Recherche [PLAsCAN—ANR-17-CONV-0002 and ActiMeth—ANR-19-CE12-0004 to V.M. and F.C.]; Cancéropôle Lyon Auvergne Rhône-Alpes—AAP international 2021 [MARACAS.v3.0 to V.M.]; Institut National du Cancer [PLBio 2019-138 MARACAS.v1 to V.M.]; Inserm-ITMO Cancer [non-coding RNA, REMOTE to V.M.]; Natural Sciences and Engineering Research Council of Canada [to D.B. and E.F.-C.]; La Ligue Nationale Contre le Cancer [to H.P.]; Fonds de Recherche du Québec—Santé [to L.F.-G.]. Funding for open access charge: Canadian Institutes of Health Research.
Conflict of interest statement. None declared.




REFERENCES

- Cavaillé, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.P., Brosius, J. and Hüttenhofer, A. (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl Acad. Sci. U.S.A.*, **97**, 14311–14316.
- Falaleeva, M., Pages, A., Matuszek, Z., Hidmi, S., Agranat-Tamir, L., Korotkov, K., Nevo, Y., Eyra, E., Sperling, R. and Stamm, S. (2016) Dual function of C/D box small nucleolar RNAs in rRNA modification and alternative pre-mRNA splicing. *Proc. Natl Acad. Sci. U.S.A.*, **113**, E1625–E1634.
- Scott, M.S., Ono, M., Yamada, K., Endo, A., Barton, G.J. and Lamond, A.I. (2012) Human box C/D snoRNA processing conservation across multiple cell types. *Nucleic Acids Res.*, **40**, 3676–3688.
- Huang, C., Shi, J., Guo, Y., Huang, W., Huang, S., Ming, S., Wu, X., Zhang, R., Ding, J., Zhao, W. *et al.* (2017) A snoRNA modulates mRNA 3' end processing and regulates the expression of a subset of mRNAs. *Nucleic Acids Res.*, **45**, 8647–8660.
- Zhong, F., Zhou, N., Wu, K., Guo, Y., Tan, W., Zhang, H., Zhang, X., Geng, G., Pan, T., Luo, H. *et al.* (2015) A snoRNA-derived piRNA interacts with human interleukin-4 pre-mRNA and induces its decay in nuclear exosomes. *Nucleic Acids Res.*, **43**, 10474–10491.
- He, X., Chen, X., Zhang, X., Duan, X., Pan, T., Hu, Q., Zhang, Y., Zhong, F., Liu, J., Zhang, H. *et al.* (2015) An lnc RNA (GAS5)/snoRNA-derived piRNA induces activation of TRAIL gene by site-specifically recruiting MLL/COMPASS-like complexes. *Nucleic Acids Res.*, **43**, 3712–3725.
- Liang, J., Wen, J., Huang, Z., Chen, X.-P., Zhang, B.-X. and Chu, L. (2019) Small nucleolar RNAs: insight into their function in cancer. *Front. Oncol.*, **9**, 587.
- van der Werf, J., Chin, C.V. and Fleming, N.I. (2021) SnoRNA in cancer progression, metastasis and immunotherapy response. *Biology (Basel)*, **10**, 809.
- Deogharia, M. and Majumder, M. (2018) Guide snoRNAs: drivers or passengers in human disease? *Biology (Basel)*, **8**, E1.
- Huang, Z.-H., Du, Y.-P., Wen, J.-T., Lu, B.-F. and Zhao, Y. (2022) snoRNAs: functions and mechanisms in biological processes, and roles in tumor pathophysiology. *Cell Death Discov.*, **8**, 259.
- Faucher-Giguère, L., Roy, A., Deschamps-Francoeur, G., Couture, S., Nottingham, R.M., Lambowitz, A.M., Scott, M.S. and Abou Elela, S. (2022) High-grade ovarian cancer associated H/ACA snoRNAs promote cancer cell proliferation and survival. *NAR Cancer*, **4**, zcab050.
- Boivin, V., Deschamps-Francoeur, G., Couture, S., Nottingham, R.M., Bouchard-Bourelle, P., Lambowitz, A.M., Scott, M.S. and Abou-Elela, S. (2018) Simultaneous sequencing of coding and noncoding RNA reveals a human transcriptome dominated by a small number of highly expressed noncoding genes. *RNA*, **24**, 950–965.
- Fafard-Couture, É., Bergeron, D., Couture, S., Abou-Elela, S. and Scott, M.S. (2021) Annotation of snoRNA abundance across human tissues reveals complex snoRNA–host gene relationships. *Genome Biol.*, **22**, 172.
- Bergeron, D., Laforest, C., Carpentier, S., Calvé, A., Fafard-Couture, É., Deschamps-Francoeur, G. and Scott, M.S. (2021) SnoRNA copy regulation affects family size, genomic location and family abundance levels. *BMC Genomics*, **22**, 414.
- Erales, J., Marchand, V., Panthou, B., Gillot, S., Belin, S., Ghayad, S.E., Garcia, M., Laforêts, F., Marcel, V., Baudin-Baillieu, A. *et al.* (2017) Evidence for rRNA 2'-O-methylation plasticity: control of intrinsic translational capabilities of human ribosomes. *Proc. Natl Acad. Sci. U.S.A.*, **114**, 12934–12939.
- Sloan, K.E., Warda, A.S., Sharma, S., Entian, K.-D., Lafontaine, D.L.J. and Bohnsack, M.T. (2017) Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.*, **14**, 1138–1152.
- Motorin, Y., Quinternet, M., Rhalloussi, W. and Marchand, V. (2021) Constitutive and variable 2'-O-methylation (Nm) in human ribosomal RNA. *RNA Biol.*, **18**, 88–97.
- Monaco, P.L., Marcel, V., Diaz, J.-J. and Catez, F. (2018) 2'-O-Methylation of ribosomal RNA: towards an epitranscriptomic control of translation? *Biomolecules*, **8**, E106.
- Penzo, M. and Montanaro, L. (2018) Turning uridines around: role of rRNA pseudouridylation in ribosome biogenesis and ribosomal function. *Biomolecules*, **8**, E38.
- Jaafar, M., Paraqindes, H., Gabut, M., Diaz, J.-J., Marcel, V. and Durand, S. (2021) 2'-O-Ribose methylation of ribosomal RNAs: natural diversity in living organisms, biological processes, and diseases. *Cells*, **10**, 1948.
- Marchand, V., Blanloeil-Oillo, F., Helm, M. and Motorin, Y. (2016) Illumina-based RiboMethSeq approach for mapping of 2'-O-Me residues in RNA. *Nucleic Acids Res.*, **44**, e135.
- Marchand, V., Pichot, F., Neybecker, P., Ayadi, L., Bourguignon-Igel, V., Wacheul, L., Lafontaine, D.L.J., Pinzano, A., Helm, M. and Motorin, Y. (2020) HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA. *Nucleic Acids Res.*, **48**, e110.
- Marcel, V., Kielbassa, J., Marchand, V., Natchiar, K.S., Paraqindes, H., Nguyen Van Long, F., Ayadi, L., Bourguignon-Igel, V., Lo Monaco, P., Monchiet, D. *et al.* (2020) Ribosomal RNA 2'-O-methylation as a novel layer of inter-tumour heterogeneity in breast cancer. *NAR Cancer*, **2**, zcaa036.
- Taoka, M., Nobe, Y., Yamaki, Y., Sato, K., Ishikawa, H., Izumikawa, K., Yamauchi, Y., Hirota, K., Nakayama, H., Takahashi, N. *et al.* (2018) Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.*, **46**, 9289–9298.
- Nottingham, R.M., Wu, D.C., Qin, Y., Yao, J., Hunnicke-Smith, S. and Lambowitz, A.M. (2016) RNA-seq of human reference RNA samples using a thermostable group II intron reverse transcriptase. *RNA*, **22**, 597–613.
- Qin, Y., Yao, J., Wu, D.C., Nottingham, R.M., Mohr, S., Hunnicke-Smith, S. and Lambowitz, A.M. (2016) High-throughput sequencing of human plasma RNA by using thermostable group II intron reverse transcriptases. *RNA*, **22**, 111–128.
- Lestrade, L. and Weber, M.J. (2006) snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.*, **34**, D158–D162.
- Jorjani, H., Kehr, S., Jedlinski, D.J., Gumienny, R., Hertel, J., Stadler, P.F., Zavolan, M. and Gruber, A.R. (2016) An updated human snoRNAome. *Nucleic Acids Res.*, **44**, 5068–5082.
- Yoshihama, M., Nakao, A. and Kenmochi, N. (2013) snOPY: a small nucleolar RNA orthological gene database. *BMC Res. Notes*, **6**, 426.
- Bouchard-Bourelle, P., Desjardins-Henri, C., Mathurin-St-Pierre, D., Deschamps-Francoeur, G., Fafard-Couture, É., Garant, J.-M., Elela, S.A. and Scott, M.S. (2020) snoDB: an interactive database of human snoRNA sequences, abundance and interactions. *Nucleic Acids Res.*, **48**, D220–D225.
- Howe, K.L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amodé, M.R., Armean, I.M., Azov, A.G., Bennett, R., Bhari, J. *et al.* (2021) Ensembl 2021. *Nucleic Acids Res.*, **49**, D884–D891.

32. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciuffo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D. *et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, **44**, D733–D745.
33. Gale, R.P., Hochhaus, A., Cross, N.C.P. and Harrison, C.J. (2021) HGNC nomenclature for fusion genes. *Leukemia*, **35**, 3039.
34. Schoch, C.L., Ciuffo, S., Domrachev, M., Hotton, C.L., Kannan, S., Khovanskaya, R., Leipe, D., McVeigh, R., O'Neill, K., Robbertse, B. *et al.* (2020) NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database (Oxford)*, **2020**, baaa062.
35. RNAcentral Consortium (2021) RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Res.*, **49**, D212–D220.
36. Kalvari, I., Nawrocki, E.P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., Griffiths-Jones, S., Toffano-Nioche, C., Gautheret, D., Weinberg, Z. *et al.* (2021) Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.*, **49**, D192–D200.
37. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
38. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S. *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, **15**, 1034–1050.
39. Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
40. Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K. *et al.* (2018) The Encyclopedia of DNA Elements (ENCODE): data portal update. *Nucleic Acids Res.*, **46**, D794–D801.
41. Krogh, N., Jansson, M.D., Häfner, S.J., Tehler, D., Birkedal, U., Christensen-Dalsgaard, M., Lund, A.H. and Nielsen, H. (2016) Profiling of 2'-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity. *Nucleic Acids Res.*, **44**, 7884–7895.
42. Kehr, S., Bartschat, S., Tafer, H., Stadler, P.F. and Hertel, J. (2014) Matching of soulmates: coevolution of snoRNAs and their targets. *Mol. Biol. Evol.*, **31**, 455–467.
43. Fernandez, N.F., Gundersen, G.W., Rahman, A., Grimes, M.L., Rikova, K., Hornbeck, P. and Ma'ayan, A. (2017) Clustergrammer, a web-based heatmap visualization and analysis tool for high-dimensional biological data. *Sci. Data*, **4**, 170151.

rRNA	Version/ID	Articles using this version
5.8S	U13369.1:6623-6779 , NR_046235.3:6601-6757	snoRNABase , snoRNA Atlas , snOPY , Gumienny et al. (2017) , Incarnato et al. (2017) , Zhu et al. (2017)
18S	X03205.1	snoRNABase , snoRNA Atlas
18S	NR_003286	snOPY , Gumienny et al. (2017)
18S	NR_145820.1	Incarnato et al. (2017) , Zhu et al. (2017)
28S	U13369.1:7935-12969	snoRNABase , snoRNA Atlas , snOPY , Gumienny et al. (2017)
28S	NR_003287.4	Incarnato et al. (2017) , Zhu et al. (2017)

Table S1. Most widely used rRNA versions in the literature.

136,726,104	136,726,239	Ensembl ID	RefSeq ID	HGNC ID	Rfam ID
		ENSG00000280496	-	-	RF00416
136,726,105	136,726,234	Ensembl ID	RefSeq ID	HGNC ID	Rfam ID
		ENSG00000276161	-	HGNC:32636	RF00416
136,726,104	136,726,237	Ensembl ID	RefSeq ID	HGNC ID	Rfam ID
		-	NR_002975	HGNC:32636	-



snoDB0971


136,726,104	136,726,239	Ensembl ID	RefSeq ID	HGNC ID	Rfam ID
		ENSG00000280496;ENSG00000276161	NR_002975	HGNC:32636	RF00416

Figure S1. SnoRNAs from different sources are merged to give at most one entry per chromosomal position in snoDB 2.0. Example of three independent overlapping entries from Ensembl (first two) and one from RefSeq (last one) for the snoRNA SNORA17B. The merged entry groups together all useful ids and information.

Studies Mean Value Tissue - Cell Lines Alphabetical

Compilation of 2'-O-methylation levels for the 18S rRNA in all studies for position Cm174

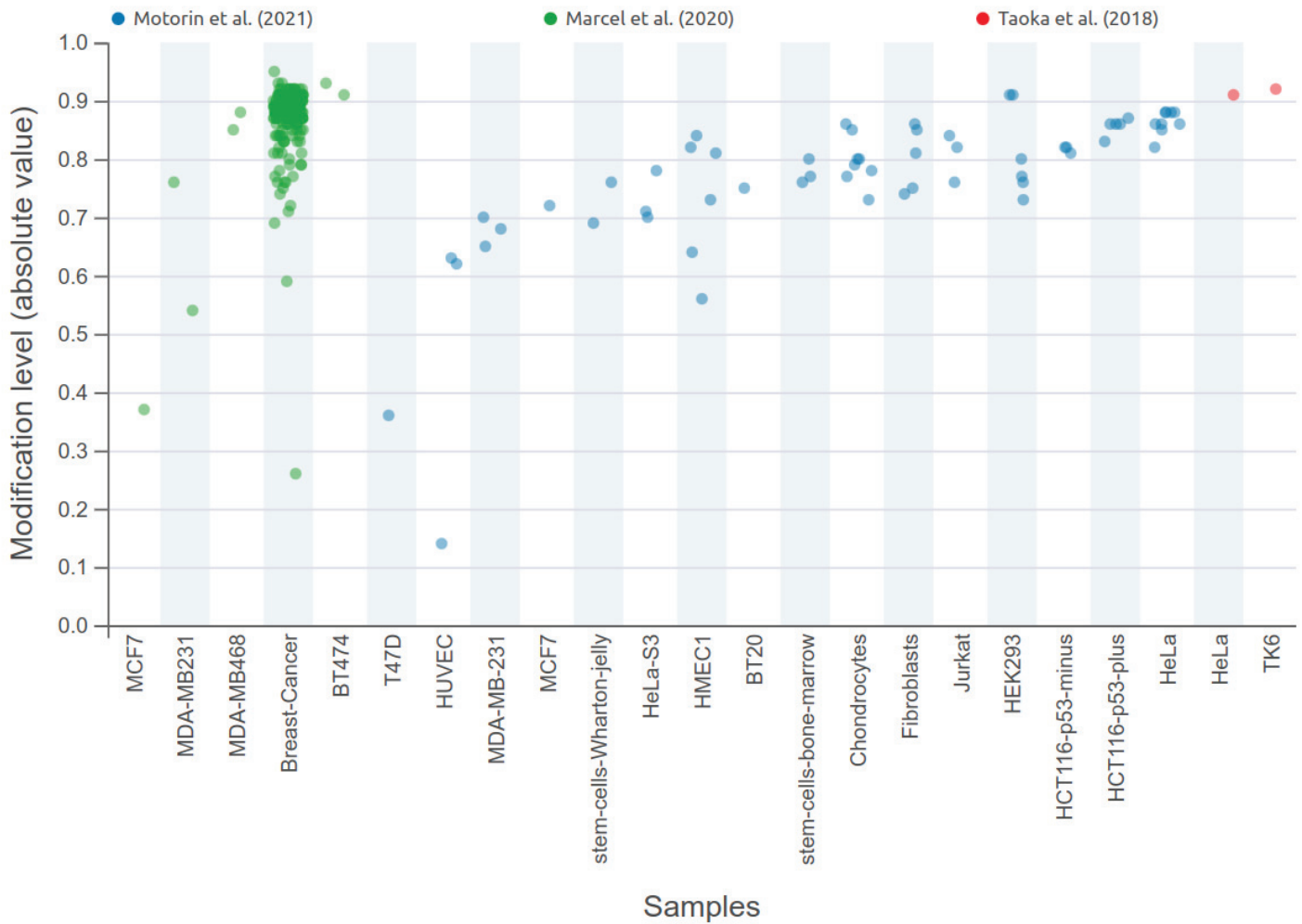


Figure S2. Screenshot of the detailed dot plot for a specific position in the "modification levels" subsection. Each value (modification level ranging from 0 to 1) from a sample is represented as a dot, and grouped by cell type. Colors represent different studies. The dot plot can also be reordered according to the mean value, tissue - cell lines, alphabetical order of the samples or by the studies (default).

C Article II : Ribosomal RNA 2'O-methylation as a novel layer of inter-tumour heterogeneity in breast cancer

Ce travail, dont je suis auteur placé en 5^{ème} position, a été publié en décembre 2020 dans le journal Nucleic Acid Research – Cancer (NAR Cancer). L'article et les figures supplémentaires sont inclus après cette page introductive. Les tableaux supplémentaires sont accessibles sur le lien suivant : <https://academic.oup.com/narcancer/article/2/4/zcaa036/6042944#supplementary-data>

C.1 Contexte

Pendant longtemps le ribosome a été considéré comme un simple effecteur de la synthèse protéique. Or, des études récentes ont démontré que des ribosomes de compositions différentes favorisent la traduction de certains ARNm, participant ainsi à la mise en place de phénotypes particuliers (GENUTH et BARNA, 2018; VENEZIA et al., 2019). La plupart des études ayant émis ce concept de « ribosome spécialisé » se sont concentrées sur la variation de composition en protéines ribosomiques. Cependant, la plus grande source de variabilité de composition du ribosome repose sur les 4 ARNr.

Mon équipe d'accueil a démontré pour la première fois que la 2'Ome des ARNr est altérée au cours de l'initiation tumorale dans des modèles cellulaires. Il s'avère que le taux de 2'Ome des ARNr varie sur certains sites seulement dans des lignées cellulaires cancéreuses mammaires. De plus, cette altération site-spécifique affecte directement la traduction de certains ARNm qui contiennent une structure IRES et codent pour des oncogènes comme IGF1R (ERALES et al., 2017; Virginie MARCEL, GHAYAD et al., 2013). Ces données suggèrent que la 2'Ome des ARNr est altérée dans les cancers du sein et induit une dérégulation de la traduction favorisant la tumorigenèse mammaire et la progression de la maladie. En revanche ces données reposent uniquement sur des modèles cellulaires.

Les objectifs de cette étude étaient de :

- (i) démontrer que la 2'Ome varie chez l'Homme, et en particulier, dans des échantillons de patientes atteintes d'un cancer du sein

- (ii) déterminer l'association entre la variation de 2'Ome et des caractéristiques biologiques et cliniques particulières pour soutenir le fait que cette variabilité n'est pas aléatoire

C.2 Résultats principaux

Pour cette étude, nous avons utilisé la technologie de RiboMethSeq. Tout d'abord, nous avons dû adapter cette technique pour effectuer pour la première fois l'analyse de la 2'Ome à partir d'échantillons humains. Le RiboMethSeq a été effectué en collaboration avec l'équipe de Y. Motorin, qui avait adapté le RiboMethSeq couplé avec le protocole Illumina via la plateforme d'Epitranscriptomique Séquençage (EpiRNA-Seq) de l'université de Lorraine. Les adaptations résident dans une diminution de la quantité d'ARNr utilisée (100 ng au lieu de 250 ng) et dans le calcul du Cscore. En effet, nous avons évalué la performance du Cscore en faisant varier (1) la fenêtre glissante des nucléotides voisins ($k = 2, 4$ et 6) et (2) l'estimateur de la couverture locale de comptage des extrémités 5'/3' des lectures (moyenne ou médiane). Le Cscore calculé à l'aide d'une fenêtre de 6 nucléotides et d'une normalisation par rapport à la médiane des comptages des extrémités des lectures 5'/3' des nucléotides voisins, s'est avéré le plus performant (Figure S1).

Afin de démontrer la reproductibilité et la robustesse du protocole RiboMethSeq pour l'analyse des échantillons humains, nous avons effectué des analyses de contrôle qualité. Pour cela, nous avons utilisé un échantillon humain commercial d'ARN (séquencés en 5 réplicats techniques) et 20 tumeurs mammaires choisies au hasard (chaque échantillon ayant été séquencé 2 fois). Tous les réplicats ont montré une grande reproductibilité dans les comptages des extrémités 5'/3' des lectures et de très fortes corrélations de Cscore entre elles (Figure S1, S2), soulignant la reproductibilité et la robustesse du protocole RiboMethSeq, même lors du séquençage de matériel biologique délicat tel que des biopsies tumorales congelées.

Finalement, nous avons analysé la 2'Ome dans une série de 195 tumeurs mammaires primaires et démontré que :

- (i) Les molécules d'ARNr présentes dans un même échantillon tumoral ne sont pas également et complètement 2'Ome sur l'ensemble des 106 sites connus chez l'Homme (intra-variabilité de la 2'Ome), comme précédemment observé dans les lignées cellulaires ;

- (ii) Les molécules d'ARNr présentes dans des échantillons de cancer du sein ne sont pas également et complètement 2'Ome sur l'ensemble des 106 sites connus chez l'Homme (inter-variabilité de la 2'Ome);
- (iii) Le taux de 2'Ome est stable entre les échantillons pour environ 60 % des sites et variable pour environ 40 %, suggérant que la variabilité de 2'Ome des ARNr n'affecte que certains sites seulement, comme observé dans les lignées cellulaires cancéreuses mammaires;
- (iv) Le taux de 2'Ome de certains sites spécifiques diffère significativement entre les différents sous-types de cancer du sein et les grades de tumeur, suggérant que la variation du taux de 2'Ome des ARNr n'est pas aléatoire.

En conclusion, cette étude correspond à la première étude démontrant l'hétérogénéité de la 2'Ome des ARNr dans des échantillons humains. De plus, cette étude suggère que la variation de la 2'Ome des ARNr n'est pas aléatoire mais qu'elle pourrait refléter des caractéristiques cliniques et/ou biologiques de la tumeur. Ainsi, la 2'Ome des ARNr offre de nouvelles perspectives avec notamment l'identification de potentiels biomarqueurs originaux pour améliorer la prise en charge clinique des patientes atteintes d'un cancer du sein.

C.3 Ma participation

Pendant cette étude, j'ai participé à l'analyse des données de Cscore et au montage des figures de la première version du manuscrit, notamment pour mettre en évidence que le taux de 2'Ome de certains sites varie entre les sous-types de cancer du sein, les récepteurs hormonaux et HER2, et le grade de la tumeur. De plus, j'ai été responsable de l'analyse de données supplémentaires et de la préparation de la réponse aux reviewers lors de l'étape de révision de ce manuscrit.

C.4 Article

Ribosomal RNA 2'O-methylation as a novel layer of inter-tumour heterogeneity in breast cancer

Virginie Marcel^{1,*}, Janice Kielbassa², Virginie Marchand³, Kundhavai S. Natchiar^{4,5,6,7,8}, Hermes Paraqindes^{1,2}, Flora Nguyen Van Long¹, Lilia Ayadi^{3,9}, Valérie Bourguignon-Igel^{3,9}, Piero Lo Monaco¹, Déborah Monchiet¹, Véronique Scott¹⁰, Laurie Tonon², Susan E. Bray¹¹, Alexandra Diot¹², Lee B. Jordan¹³, Alastair M. Thompson^{12,14}, Jean-Christophe Bourdon¹², Thierry Dubois¹⁵, Fabrice André¹⁰, Frédéric Catez¹, Alain Puisieux¹, Yuri Motorin^{3,9}, Bruno P. Klaholz^{4,5,6,7,8}, Alain Viari^{2,16} and Jean-Jacques Diaz^{1,*}

¹Univ Lyon, Université Claude Bernard Lyon 1, Inserm 1052, CNRS 5286, Centre Léon Bérard, Centre de Recherche en Cancérologie de Lyon, 69008 Lyon, France, ²Synergie Lyon Cancer, Gilles Thomas Bioinformatics Platform, Centre Léon Bérard, 69008 Lyon, France, ³UMS2008 IBSLor CNRS-INSERM-Lorraine University, Biopôle, 9 avenue de la forêt de haye, 54505 Vandoeuvre-les-Nancy, France, ⁴Centre for Integrative Biology (CBI), Department of Integrated Structural Biology, IGBMC, CNRS, Inserm, Université de Strasbourg, 1 rue Laurent Fries, 67404 Illkirch, France, ⁵Institut of Genetics and of Molecular and Cellular Biology (IGBMC), 1 rue Laurent Fries, 67404 Illkirch, France, ⁶Centre National de la Recherche Scientifique (CNRS), UMR 7104, 67404 Illkirch, France, ⁷Institut National de la Santé et de la Recherche Médicale (Inserm), U964, 67404 Illkirch, France, ⁸Université de Strasbourg, 67404 Illkirch, France, ⁹IMoPA, UMR 7365 CNRS-UL, Biopole UL, 54505 Vandoeuvre-les-Nancy, France, ¹⁰Predictive biomarkers and novel therapeutic strategies Group, Institut Gustave Roussy, University of Paris Sud, INSERM 981, Université Paris Saclay, 114 rue Edouard Vaillant, 94800 Villejuif, France, ¹¹Tayside Tissue Bank, Ninewells Hospital and Medical School, NHS Tayside, Dundee DD1 9SY, Scotland, UK, ¹²Division of Cancer Research, University of Dundee, Ninewells Hospital and Medical School, Dundee DD1 9SY, Scotland, UK, ¹³Department of Pathology, University of Dundee, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK, ¹⁴Olga Keith Wiess Chair of Surgery, Dan L Duncan Breast Center, Division of Surgical Oncology, Baylor College of Medicine, Houston, TX 77030, USA, ¹⁵Breast Cancer Biology Group, Translational Research Department, Institut Curie-PSL Research University, 26 rue d'Ulm, 75005 Paris, France and ¹⁶INRIA Grenoble Rhône-Alpes, 38330 Montbonnot-Saint-Martin, France

Received July 08, 2020; Revised November 20, 2020; Editorial Decision November 20, 2020; Accepted November 23, 2020

ABSTRACT

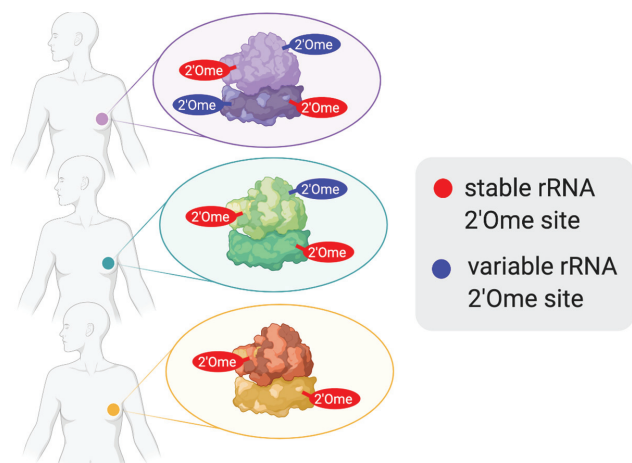
Recent epitranscriptomics studies unravelled that ribosomal RNA (rRNA) 2'O-methylation is an additional layer of gene expression regulation highlighting the ribosome as a novel actor of translation control. However, this major finding lies on evidences coming mainly, if not exclusively, from cellular models. Using the innovative next-generation RiboMeth-seq technology, we established the first rRNA 2'O-methylation landscape in 195 primary human breast tumours. We uncovered the existence of compulsory/stable sites, which show limited inter-patient variability in their 2'O-methylation level, which map on functionally important sites of the hu-

man ribosome structure and which are surrounded by variable sites found from the second nucleotide layers. Our data demonstrate that some positions within the rRNA molecules can tolerate absence of 2'O-methylation in tumoral and healthy tissues. We also reveal that rRNA 2'O-methylation exhibits intra- and inter-patient variability in breast tumours. Its level is indeed differentially associated with breast cancer subtype and tumour grade. Altogether, our rRNA 2'O-methylation profiling of a large-scale human sample collection provides the first compelling evidence that ribosome variability occurs in humans and suggests that rRNA 2'O-methylation might represent a relevant element of tumour biology useful in clinic. This novel variability at molecular level offers

*To whom correspondence should be addressed. Tel: +33 4 26 55 67 45; Email: virginie.marcel@lyon.unicancer.fr
Correspondence may also be addressed to Jean-Jacques Diaz. Tel: +33 4 78 78 28 19; Email: jean-jacques.diaz@lyon.unicancer.fr

an additional layer to capture the cancer heterogeneity and associates with specific features of tumour biology thus offering a novel targetable molecular signature in cancer.

GRAPHICAL ABSTRACT



INTRODUCTION

Emerging evidence suggests that changes in ribosomal RNA 2'O-ribose methylation (rRNA 2'Ome) in the human ribosome play a key role in regulating translation thereby contributing in setting particular phenotypes such as hallmarks of cancer cells (1–5). Using cancer cellular models, we and others indeed reported an association between variation in rRNA 2'O-methylation and modulation of translation efficiency of particular mRNA subsets in cancer (1,6–9). In particular, we demonstrated that inactivation of the tumour suppressor p53 protein in mammary epithelial cells is associated with an increased expression of the 2'O-methyltransferase *FBL*, an alteration of the 2'O-methylation level at some of the 106 rRNA-methylated sites and an increased translation of some cellular Internal Ribosome Entry Sites (IRES)-containing mRNAs such as those encoding the oncogenic IGF1R and cMyc proteins (1). Moreover, we have shown that *FBL* overexpression promotes IGF1R-dependent proliferation, colony formation and resistance to chemotherapy in breast cancer cell lines, suggesting a role of rRNA 2'O-methylation in mammary tumorigenesis and breast cancer progression (1). Importantly, using cell-free *in vitro* translation assays, we recently demonstrated that modulation of rRNA 2'O-methylation directly affects translation initiation of IRES-containing mRNAs, the IGF1R mRNA and the Cricket paralysis virus mRNA (2). This new layer of gene expression regulation recently uncovered, joins the numerous descriptions of chemical modifications of both coding and non-coding RNAs that regulates post-transcriptional processes, a field known as epitranscriptomics (3,4,10). However at present, it is unclear whether rRNA 2'O-methylation varies in humans.

Although genetics, genomics, epigenetics, transcriptomics and proteomics revealed the heterogeneity of breast

tumours at the molecular level, identification of novel molecular layers are required to fully capture all the inter-patient variability, essential for developing clinical applications (11). Amongst them, epigenetics appears as promising additional molecular fingerprints to improve patient classification, prognosis or treatment, including in breast cancer (11–13). In these molecular portraits of breast cancers, RNA epitranscriptomics and more specifically rRNA 2'O-methylation, has never been investigated. Thanks to the recent resolution of technological issues, including the development of RNA-seq-based approaches dedicated to the profiling of rRNA 2'O-methylation (4,5,14,15) and the recent developments of high-resolution cryo-EM that has provided insights into the detailed structure of the human ribosome including rRNA modifications (16–18), rRNA 2'O-methylation profiling can now be achieved in human samples to determine whether it can provide a molecular fingerprint in breast cancer.

Here, we present the first rRNA 2'O-methylation landscape of human breast tumours. We used the next-generation RiboMeth-seq technology that we remodelled to perform high-throughput analyses of human clinical samples. Our data reveal the unexpected existence of specific modulation of 2'O-methylation at only particular rRNA positions between human breast tumours that are associated with clinical outcome and biological features.

MATERIALS AND METHODS

Human samples

A series of 195 female primary breast tumours were collected from 1997 to 2011 and maintained by the Tayside Tissue Bank (TTB, Dundee, Scotland, UK) under ethical approval (REC Reference 07/S1402/90) (Supplementary Table ST1) (19). This retrospective series of primary breast tumours is composed of both non-invasive (45%) and invasive (55%) lymph node tumours (Supplementary Table ST1). Total RNA was extracted from frozen tissues by TTB services as already described (19,20). A second series of 7 mammaplasties (i.e. non-transformed, non-proliferative but mostly adipocyte cells) derived from healthy donors was sequenced independently (Institut Curie, France) (21). A third series composed of 10 mammary malignant tumours and 8 mammary benign tumours (i.e. non-transformed but hyper-proliferative epithelial cells) were collected between 2006 and 2007 (Institut Gustave Roussy, France) from consenting patients using fine-needle aspiration after suspicion of breast lesion (Supplementary Table ST2) (22). A human RNA reference sample (i.e. RNA reference) was used as a calibrated source of rRNA (Human XpressRef Universal Total RNA, Qiagen) prepared from 20 different human adult and foetal normal major organs.

RiboMeth-seq

Levels of rRNA 2'O-methylation at the 106 rRNA 2'O-methylated sites were determined by RiboMeth-seq (14,15,23–25). Presence of 2'O-methylation protects the phosphodiester bond located at the 3' of the 2'O-methylated nucleotide from alkaline hydrolysis. Thus, the presence of

2′O-methylation at the given nucleotide n induces under-representation of RNA fragments starting at the nucleotide $n + 1$ and ending at position n allowing to calculate a 2′O-methylation level at the corresponding nucleotide position (or C-score) varying from 0 to 1 (15). RiboMeth-seq was performed using the Illumina sequencing technology and raw data were processed as previously described (14,25). The median number of total reads reaches 7.2 millions after trimming, these reads being aligned on the 7.2 kb-long rRNA sequences, that corresponds to the optimal sequencing depth (14,26).

Analysis of the breast cancer series

Amongst an initial series of 214 primary breast tumours, RiboMeth-seq data of 195 primary breast tumour samples passed the QC criteria (representing 91% of the initial series) and were thus retained for the downstream analyses (Supplementary Table ST1). Unsupervised data analysis was performed (hierarchical clustering and principal component analysis (PCA)) using the C-scores at the 106 individual rRNA 2′O-methylated sites of the 195 samples. rRNA 2′O-methylation profiles were shown as either box-and-whiskers plots, line charts or barcharts. Classification of the rRNA 2′O-methylation sites into ‘stable’ and ‘variable’ classes was done empirically based on the variability of the sites (in terms of interquartile range, IQR), the cut-off site being the site from which the IQR values no longer lie on this straight line (see Supplementary Materials and Methods for details). Between-group comparisons were performed using Fisher’s exact test for categorical data or Mann–Whitney test for quantitative data. Bonferroni or false discovery rate correction methods were applied for multiple comparisons (P.adj). All P -values corresponded to two-tailed P -values. A P -value < 0.05 was considered to be statistically significant. Statistical analyses and graphical representations were performed using either R v3.6.3 or GraphPad Prism v7.0a software (GraphPad Software, Inc).

rRNA 2′O-methylation evolution and mapping

To further characterize the two classes of human rRNA 2′O-methylated sites, their conservation during evolution was compared using box C/D snoRNA guides as surrogates of rRNA 2′O-methylated sites, since to date availability of rRNA 2′O-methylation profiles is limited to only few organisms. Three evolution groups were established using the snoRNA-LBME-db database (www-snoRNA.biotoul.fr). Using Fisher’s exact test, enrichment analysis was performed to compare the proportion of stable versus variable sites amongst the three groups of rRNA 2′O-methylated sites. The rRNA 2′O-methylated sites were mapped on the structure of the HeLa cancer cell human ribosome determined by cryo-EM (16–18) and images were drawn using the PyMol software. Observations were based on previous reported three-dimensional molecular docking analysis of tRNAs or ribosome-associated factors, as already discussed in (ref 16–18).

RESULTS

Optimization of RiboMeth-seq technology for human samples

To profile rRNA 2′O-methylation in human primary breast tumours, we used the novel RiboMeth-seq technology. The measurement of 2′O-methylation frequency by RiboMeth-seq technology relies on a partial alkaline hydrolysis of the rRNA phosphodiester bonds, which become refractory to hydrolysis when adjacent riboses are methylated in position 2′ (14,15,25). RiboMeth-seq processing yields a score (i.e. C-score) at each of the 106 rRNA 2′O-methylated positions, reflecting the level of 2′O-methylation.

Since rRNA 2′O-methylation profiling has never been performed on large series of human samples, we developed RiboMeth-seq-dedicated extensive quality controls based on numerous metrics (see Supplementary Materials and Methods). Using human RNA reference samples composed of a mix of 20 different human adult and foetal normal major organs and a test set of 20 primary breast tumour samples, we show a strong correlation in the C-scores amongst the technical replicates (Supplementary Figures S1 and 2). Overall, our data demonstrate that the high-throughput RiboMeth-seq technology is a reproducible and robust technology even when sequencing delicate biological material such as frozen tumour biopsies.

Identification of two classes of rRNA 2′O-methylation sites

We then profiled rRNA 2′O-methylation in a series of 195 primary breast cancers composed of a mix of invasive and non-invasive tumours using RiboMeth-seq (Supplementary Table S1) (20). Unsupervised statistical analysis revealed variability in rRNA 2′O-methylation levels (Figure 1). It first showed that, for a given tumour, the 2′O-methylation level varies between 0 and 0.98 amongst the 106 rRNA 2′O-methylation sites (Figure 1 and Supplementary Figure S3a). This demonstrates that in a single human tumour sample, all rRNA molecules are not fully and equally 2′O-methylated at the 106 sites.

Hierarchical clustering also revealed that, for a given rRNA site, 2′O-methylation levels differ when comparing the 195 human tumours (i.e. inter-patient variability) (Figure 1; Supplementary Figures S3a and 4). We defined two classes of rRNA 2′O-methylated sites based on the variability of their 2′O-methylation level amongst patients (Supplementary Figure S3c). The classification into these two classes was done empirically based on the intra-variability of the sites (in terms of interquartile range, IQR, at a given site). The first class contains 60 rRNA 2′O-methylated sites (56.6% of all the rRNA 2′O-methylated positions), the levels of which exhibit low inter-patient variability between the 195 tumour samples, despite these series representing the full spectrum of breast cancer subtypes (Figure 2a and Supplementary Figure S4, red). Thereafter, they are termed ‘stable’ sites since they display the most stable C-scores between the 195 tumours. The second class corresponds to a limited number of rRNA 2′O-methylated sites (46 sites, around 43.4% of the rRNA 2′O-methylated positions) that exhibit high inter-patient vari-

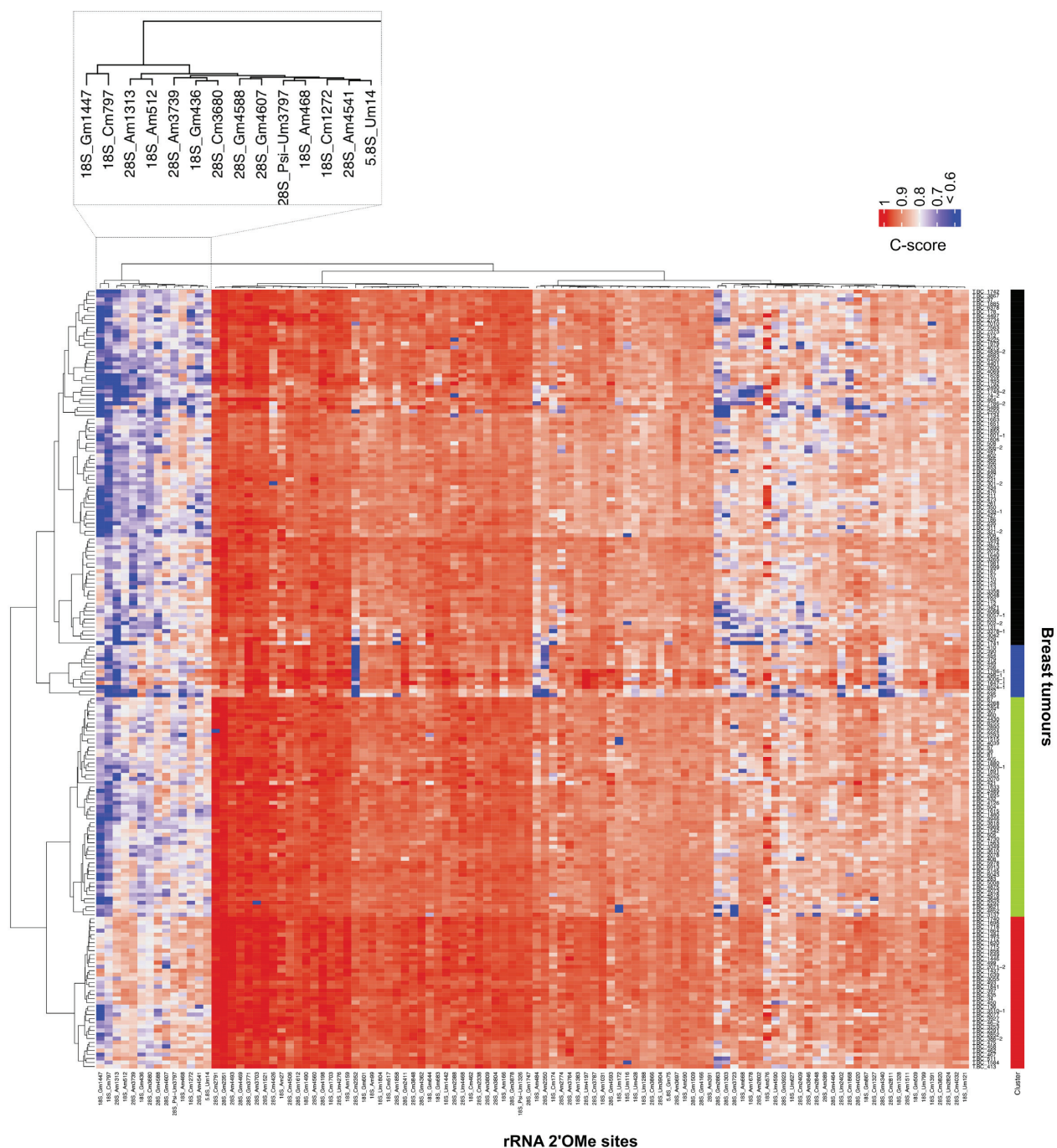


Figure 1. Stability and variability of rRNA 2'O-methylation levels between breast tumours. Levels of rRNA 2'O-methylation (i.e. C-score) were determined at the 106 rRNA 2'O-methylated sites using RiboMeth-seq in a series of 195 primary human breast tumours. Data are presented as a hierarchical clustering where the C-score of each position (column) for each tumour (row) is presented (colour scale). Dendrograms represent relationships of similarity between breast tumours on the basis of their rRNA 2'O-methylation profiles (left panel) that identify four groups of breast tumour samples (right colour panel), or in the rRNA 2'O-methylation level at a given site between tumours (top panel). An enlarged view of a group of clustered rRNA 2'O-methylated sites is presented.

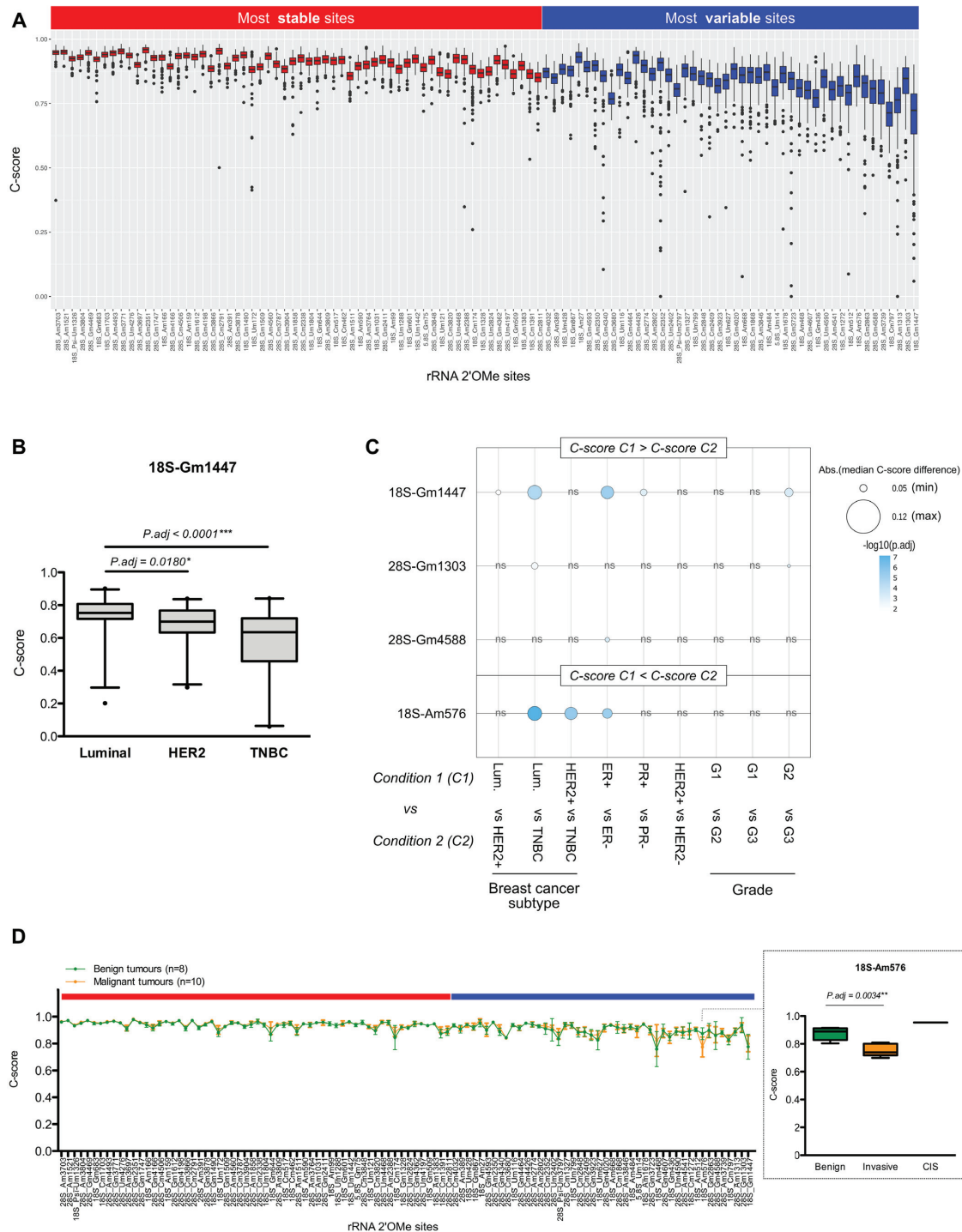


Figure 2. Biological relevance of variable rRNA 2'O-methylation levels in breast cancer. (A) C-score variation amongst the 195 human primary breast tumours at each of the 106 rRNA 2'O-methylated sites was ranked by increasing interquartile range (IQR). Based on IQR divergence, two classes of rRNA 2'O-methylated sites were defined: sites with the most stable C-scores (red); and sites with the most variable C-scores (blue). (B) Example of significant change in rRNA 2'O-methylation levels at the 18S-Gm1447 site between luminal, HER2+ and TNBC breast tumours. (C) Summary of the four variable rRNA sites, whose 2'O-methylation levels were significantly different between breast cancer subtypes, hormonal and HER2 receptors, and tumour grade. ER: oestrogen receptor; PR: progesterone receptor; Luminal: ER+ PR± HER2-; HER2+: ER+/- PR+/- HER2+; TNBC: ER- PR- HER2-; G: grade. (D) Comparison of rRNA 2'O-methylation profiles in a series of 8 benign (green) and 10 malignant (orange) tumours. An enlarged view presents the significant change in rRNA 2'O-methylation levels at the variable 18S-Am576 site in malignant compared to benign mammary tumours, with carcinoma in situ (CIS) presenting the highest rRNA 2'O-methylation level. *: adjusted P -value < 0.05; **: P_{adj} < 0.01; ***: P_{adj} < 0.0001.

ability in their 2′O-methylation level (‘variable’ sites, blue). Amongst them, the 28S-Am1313, 28S-Cm2352 and 28S-Gm3723 sites displayed important inter-patient variability, the C-scores ranging from 0 to >0.90 and a null C-score being observed in different tumours (Figure 2a and Supplementary Figure S3a). Position-by-position comparison of the δ C-score [(mean C-score + 2sd) – (mean C-score – 2sd)] between the 195 breast tumours and the statistical analysis based on five technical replicates of normal human RNA reference samples reveals that the high inter-patient variability in C-scores corresponds to a biological variability rather than to technical variability (Supplementary Figure S3b). Similarly, these two classes of most stable and most variable sites derived from the 195 breast tumours were observed in a series of 7 mammoplasties derived from healthy female donors, most of the sites being classified in the same classes with the exception of few ones (Supplementary Figure S5). The difference might come either from the small size of this series or from the fact that variability in rRNA 2′O-methylation might be tissue-specific, mammoplasties mostly corresponding to adipocyte cells. These data show that rRNA 2′O-methylation varies between human samples and demonstrate the co-existence of stable and variable classes of rRNA 2′O-methylated sites within human samples.

Association of rRNA 2′O-methylation level with biological and clinical characteristics

To assess the biological significance of the co-existence of two classes of rRNA 2′O-methylated positions displaying either the most ‘stable’ 2′O-methylation level between human tumour samples or the most ‘variable’ 2′O-methylation level, we evaluated the association between variations in C-score at the 106 rRNA 2′O-methylated positions and biological or clinical characteristics of primary breast tumours (Figure 2B and C). Remarkably, amongst the stable rRNA 2′O-methylated sites, no association between change in C-score and well-known characteristics of breast cancers was observed (i.e. tumour grade, tumour size, lymph node involvement, hormonal/HER2 receptor status, breast cancer subtype or *TP53* mutation). In contrast, we identified four variable rRNA 2′O-methylated sites (18S-Gm1447, 28S-Gm1303, 28S-Gm4588 and 18S-Am576), the levels of which are significantly different between breast cancer subtypes, oestrogen and progesterone statuses, as well as tumour grades (Figure 2B and C; Supplementary Table S3). In both human tumour samples and cell lines, the 2′O-methylation level of the 18S-Gm1447 site decreased in triple-negative breast cancers (TNBCs) compared to the luminal tumours, whilst that of the 18S-Am576 site increased in TNBCs (Figure 2B and C; Supplementary Figure S6). Furthermore, we compared rRNA 2′O-methylation profile in a small series of eight benign (i.e. hyperproliferative, but non-transformed epithelial cells) and 10 malignant tumours, the former do not represent the full breast cancer inter-patient heterogeneity. In this series, 2′O-methylation level at the 18S-Am576 site distinguished benign from malignant mammary tumours (Figure 2D).

Secondly, we evaluated the association between breast cancer patients harbouring tumours exhibiting similar

rRNA 2′O-methylation profiles and biological/clinical characteristics of breast tumours. Hierarchical clustering and PCA methods identified 4 clusters of breast cancer tumours based on their rRNA 2′O-methylation profiles (Figures 1 and 3A). Analysis of survival indicates that, although no significant association was observed, the patients carrying breast tumours characterized by distinct rRNA 2′O-methylation profile display distinguishable overall survival (Figure 3B). In particular the ‘blue’ group displays the poorest overall survival. This blue group was enriched in TNBC and in tumours of high grade compared with the other groups (Figure 3C and D; Supplementary Figure S7). Regarding the fact that tumours in the blue groups exhibited the lowest C-score at most of the rRNA 2′O-methylated sites (Figure 1), these data suggest that global decrease in rRNA 2′O-methylation might be associated with breast tumour aggressiveness. Finally, clustering breast cancer tumours on the basis of their 2′O-methylation level at the 106 rRNA sites highlights at diagnosis, patients carrying small tumours size who display prognosis as poor as patients carrying large tumours size (Figure 3E, blue group). Indeed, patients carrying small tumour size exhibit different survival rates depending on the rRNA 2′O-methylation profile. Overall, these data indicate that 2′O-methylation levels at variable rRNA sites are associated with breast cancer subtypes and tumour grade.

Evaluation of the structural impact of the two classes of rRNA 2′O-methylation sites

To further characterize the two classes of human rRNA 2′O-methylated positions identified in the 195 breast tumours, we compared their conservation during evolution. Amongst the 106 rRNA 2′O-methylation sites described in humans, some sites have also been identified in yeast and bacteria (snoRNA-LBME-db database, www.snorna.biotoul.fr). We thus compared the proportion of either ‘stable’ or ‘variable’ sites classified on the basis of our data, amongst three groups: the group of rRNA 2′O-methylated sites only detected in humans (‘humans’, $n = 68$); the group of rRNA 2′O-methylated sites detected both in humans and yeast (‘eukaryotes’, $n = 35$); and the group of rRNA 2′O-methylated sites conserved in humans, yeast and bacteria (‘universal’, $n = 3$; see also annotation on the human ribosome structure in ref 17). Enrichment analysis showed an increased proportion of the variable rRNA 2′O-methylated sites in humans compared to other organisms (Figure 4A), suggesting a link between the appearance of variable rRNA 2′O-methylated sites and evolution, characterized by an increasing number of fine-tuning gene expression mechanisms.

We then mapped the stable and variable rRNA 2′O-methylated positions on the human ribosome structure (16–18). It revealed an interesting distribution in space, where the decoding centre (DC), the peptidyl-transferase centre (PTC) and the polypeptide exit tunnel are mostly decorated with stable 2′O-methylated positions, whilst some variable positions are found in 1 or 2 nt layers away from the functional sites (i.e. maximum 4.5 Å away from the functional sites) (Figure 4B and C, annotated sites). However, 2′O-methylated sites are also found in functional sites, methyl

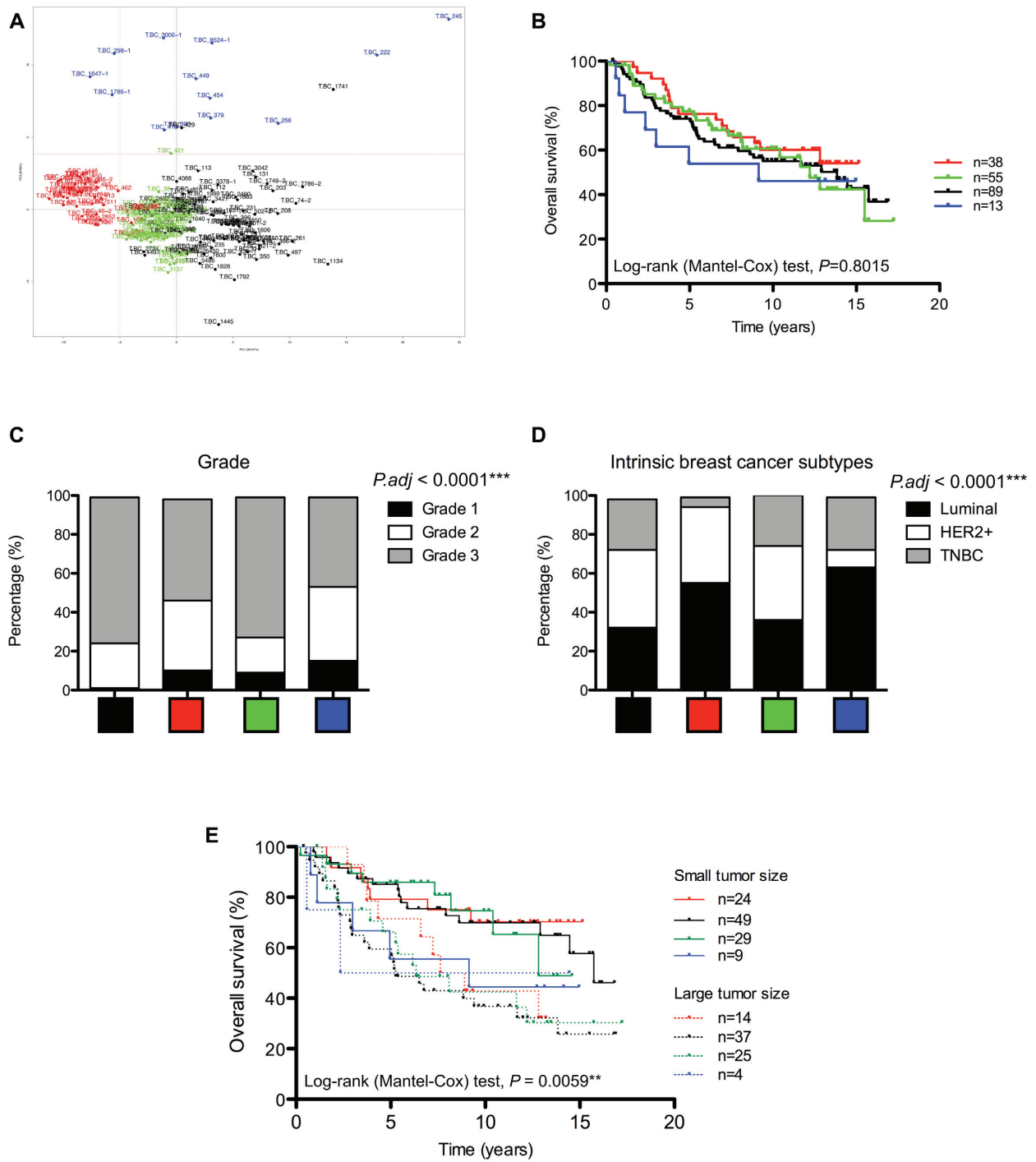


Figure 3. Association of rRNA 2'O-methylation profiles with biological and clinical breast cancer characteristics. (A) PCA showed four clusters of breast tumours based on their rRNA 2'O-methylation profiles that correspond to the ones identified using the hierarchical clustering approach (Figure 1). (B) Kaplan-Meier curve suggested the patients carrying breast tumours characterized by a 'blue' rRNA 2'O-methylation profile display the poorest overall survival. (C and D) Significant differences in repartition amongst the four clusters were observed regarding tumour grade (C) and breast cancer subtype (D). Luminal: ER+ PR± HER2-; HER2+: ER+/- PR+/- HER2+; TNBC: ER- PR- HER2-. (E) Kaplan-Meier curve depicts the overall survival of patients grouped on the basis of both their tumour size at diagnosis and their rRNA 2'O-methylation profiles. **: $P < 0.01$; ***: $P < 0.0001$.

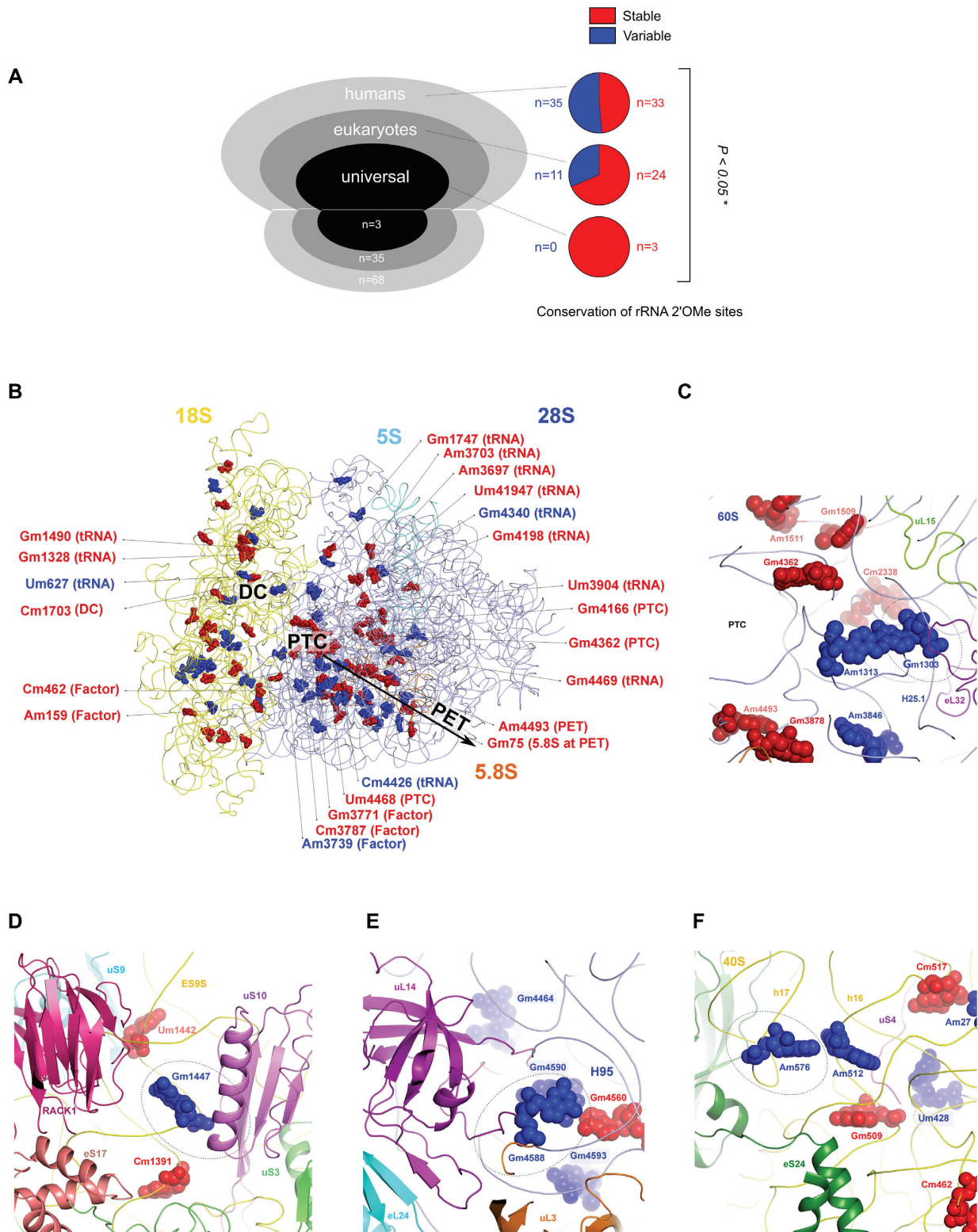


Figure 4. Structural properties of stable and variable rRNA 2'O-methylated sites. (A) Amongst the 106 rRNA 2'O-methylated sites in humans, 68 are specific to humans (humans), 35 are found across eukaryotes (eukaryotes) and 3 are conserved from bacteria to humans (universal) (left panel). The proportion of stable and variable sites amongst these three groups is represented using pie charts (right panel). (B and C), Mapping of the two classes of rRNA 2'O-methylated sites on the human ribosome (B) and at the PTC (C). Only rRNA 2'O-methylated positions directly involved in function domains or in interaction with components of the translational machinery have been annotated. (D–F) Structural environment of the variable rRNA 2'O-methylated sites significantly associated with breast cancer features. Red: stable sites; blue: variable sites; dotted circles: rRNA 2'O-methylation sites of interest; DC: decoding centre; PTC: peptidyl-transferase centre; PET: peptide exit tunnel. *: $P < 0.05$.

carbon in very few 2′O-methylated sites are directly involved in van der Waals contacts with tRNAs, mRNA and other ribosome functional large ligands, as can be predicted from modelling these into the human ribosome structure. Most tRNAs binding regions in the large and small subunits are decorated with stable 2′O-methylated positions with the exception of 4 variable 2′O-methylated sites (18S-Um627, 28S-Am3739, 28S-Gm4340 and 28S-Cm4426) (Figure 4B). In particular, the 28S-Am3739 nucleotide is located at the hairpin loop of helix H69, which is located right next to the inter-subunit rotation centre describing pre- and post-translocation rotational states of the 40S with respect to the 60S ribosomal subunit and whose 2′O-methylation may influence P-site tRNA binding. The variable 28S-Gm4340 site is located at the CCA end region of the E-site tRNA that provides favourable van der Waals contacts to the E-site tRNA as observed experimentally in the human 80S structure. Finally, the four rRNA sites displaying changes in 2′O-methylation associated with biological and clinical features of breast cancer (28S-Gm1303, 18S-Gm1447, 28S-Gm4588 and 18S-Am576), are located at the periphery of the PTC of the ribosomes and are accessible for tRNAs, mRNA and other ribosome functional large ligands such as GTPases that associate transiently during the translation process (Figure 4C–F). Various factor binding pockets and their vicinity are mostly decorated with variable sites. Molecular docking tRNA and ribosomal factors suggest that stable rRNA 2′O-methylated sites are mostly located at the catalytic sites of the ribosomes, whilst the variable sites are found at the regulatory sites of the ribosomes.

DISCUSSION

Taken together, we show here for the first time that rRNA molecules exist naturally in humans as molecules not fully and equally 2′O-methylated at all the 106 positions. Indeed, our structure/function and evolution enrichment analyses as well as the differential association with biological characteristics, demonstrates that at least two classes of rRNA 2′O-methylated sites concur in human breast cancers, depending on their ability to tolerate (i.e. variable sites) or not (i.e. stable sites) the absence of chemical modifications. This unique discovery made by comparing human samples demonstrates the co-existence of stable and variable 2′O-methylation at specific rRNA positions.

Historically, plethora of studies mainly performed in both yeast and *Xenopus laevis* models, which have been the historical prototypic, eukaryotic models used during many years finely decipher the highly complex molecular mechanisms of rRNA transcription, chemical modifications and processing (27–31). These studies allowed the mapping of rRNA 2′O-methylation sites but also the demonstration of the importance of 2′O-methylation in the conserved intrinsic activity of the ribosome, on which life relies. Indeed, it appeared over the past years that loss of 2′O-methylation at a few number of rRNA sites in yeast is critical for ribosome functions and induces cell death (27,32). However, more recent findings indicate that in yeast, loss of rRNA 2′O-methylation at other sites have only little or no effect.

Our finding sustain the emerging notion that rRNA chemical modifications provide plasticity to the ribosome

structure on which its regulatory effects rely, thus extending the modulation and activity of the ribosome to translational regulation in humans, including in cancer (4,33). In a single human sample, it appears that rRNA 2′O-methylation varies between the 106 rRNA 2′O-methylated positions ($0 < C\text{-score} < 0.98$). It indicates that even within a single human sample, rRNA molecules exist naturally as molecules not fully and equally 2′O-methylated at all the 106 positions. This observation is consistent with previous studies that were performed, albeit in limited number of models, either in cell lines (i.e. HeLa, PA1, HCT-116 and 293) (2,8,18,34–37), in cellular models describing tumour initiation and progression of solid and haematological cancers (1,6,38) or in few number of adult/embryonic organs (39). In addition to these previous studies, our findings also demonstrate that difference in rRNA 2′O-methylation occur between different human samples and reflects inter-patient variability. Variation in rRNA 2′O-methylation level between tumours might result from different mechanisms. Change in 2′O-methylation levels can result from alteration in expression of each individual component of the rRNA 2′O-methylation complex (i.e. the rRNA methyltransferase FBL, the structural proteins Nop56, Nop58 or NHP2L1, or the small nucleolar snoRNA, which hybridizes in a sequence-specific manner the rRNA and thus specifically guides the enzyme) (1,7,9,28). Expression of these different components has been correlated with genomic alteration, loss of tumour suppressor or oncogene activation (1,7,28,40), suggesting that the 106 rRNA 2′O-methylation sites can reflect numerous alterations of the tumour. Moreover, changes in the ratio substrate:enzyme (rRNA:rRNA 2′O-methylation complex) might be sufficient to alter rRNA 2′O-methylation level, the rRNA level being also dependent of particular tumour suppressors and oncogenes (41). Interestingly, the rRNA 2′O-methylated positions that are not fully 2′O-methylated in a single human sample overlap with those displaying strong inter-variability between breast cancer samples. Development of statistical and methodological tools for future meta-analyses will be required to identify the complete list of rRNA sites accepting lack of 2′O-methylation, either commonly or specifically to a particular physio/pathological context.

The diversity of rRNA 2′O-methylation profiles naturally occurring in humans reinforces the altered translational control induced by experimental modulation of rRNA 2′O-methylation (1,2,28). Amongst the variable rRNA 2′O-methylation positions, the 18S-Gm1447 displays the highest variability in rRNA 2′O-methylation amongst tumours with the identification of human tumours and breast cancer cell lines that lack 2′O-methylation at this particular nucleotide. 18S-Gm1447 is in close vicinity of the uS10 ribosomal protein, which has been shown to bind to HCV IRES, regulates translation initiation of short 5′UTR containing mRNAs and contributes in resolving stalled ribosomes (42–46). Thus, absence or presence of rRNA 2′O-methylation at 18S-Gm1447 might finely modulate uS10 activity, which applies to the intrinsic ribosome function in general depending on its plasticity. Overall, our data reinforce, in humans, both the existence of rRNA 2′O-methylated compulsory/stable sites, which display no or low inter-patient variability and on which ribosome activities might rely as suggested by

structural observations, and of rRNA 2'O-methylated plastic sites, which provides novel properties to the ribosome that can now be seen as a direct actor of translational regulation.

In addition to 2'O-methylation, rRNA exhibits other chemical modifications that affect translational regulation and contribute in the acquisition of cancer hallmarks. rRNA pseudouridylation (Ψ) remains the prototypic example. Alteration in Ψ rRNA profile induced by mutant DKC1, the enzyme-directed rRNA pseudouridine synthesis, has been shown to inhibit translation of the tumour suppressor p53 protein (47,48). More recently, the 18S-U609/U863 Ψ has been shown to display tumour suppressive activity in response to KRAS-induced oncogenic stress in liver cancer (49). Similarly, loss of 18S-m¹acp³U1248 Ψ and 28S-m⁶A4220 rRNA are associated with colorectal and hepatocellular carcinoma, respectively (50,51). Thus, like other rRNA chemical modifications, rRNA 2'O-methylation can regulate translation and might provide novel molecular understanding of cancer biology, as it has been previously suggested, alteration of rRNA 2'O-methylation profile impairing cell proliferation and resistance to chemotherapy (1). Future studies will thus be required to finely identify rRNA 2'O-methylation sites involved in tumorigenesis by comparing series of normal and tumour breast tissues. Here, using benign tissue corresponding to non-transformed epithelial, we identify that in hyper-proliferative cells 2'O-methylation level of the 18S-Am576 site differs between benign and malignant tumours, suggesting that the rRNA 2'O-methylation profile is different between transformed and non-transformed tissue. It has to be noted that due to the cycling evolution of breast cancer through life (i.e. menstruation, pregnancy, menopause...) and the discrepancies between benign tumours (i.e. hyper-proliferative), mastoplasia (i.e. astrocyte enrichment) or stromal tissue (i.e. molecular traits) and breast cancer tissues (52), it remains difficult to build matched collection of normal mammary tissues.

Our data also demonstrate that rRNA 2'O-methylation corresponds to an additional layer of molecular characterization of breast cancer patient heterogeneity. Indeed, we show for the first time that rRNA 2'O-methylation levels differ between breast cancer subtypes and tumour grades. rRNA 2'O-methylation corresponds to an additional layer of molecular characterization that may help in the near future to improve breast cancer patient classification but also patients management. First, rRNA 2'O-methylation might be useful to improve the refinement of breast cancer patients stratification required in precision medicine. For example, the rRNA 2'O-methylation profile could help in identifying patients with the poorest outcome although they carry small tumours that are generally associated with a low risk factor in breast cancer. Such patients may thus benefit from a more aggressive treatment protocol usually reserved to treat largest breast tumours. Second, cancer-specific alterations of ribosome structure induced by naturally occurring variation in rRNA 2'O-methylation might be the source of powerful and specific anti-cancer molecules (18,53). Indeed, presence or absence of 2'O-methylation at a particular position might provide formation of specific ligand

pockets to specifically target cancer-related ribosomes. In cancer, ribosome indeed appears as an emerging therapeutic target, inhibitors of RNA polymerase I activity specifically killing cancer cells being currently evaluated in clinical trials (54) and small molecules such as eukaryote-specific antibiotics directly targeting the human ribosome reducing proliferation of cancer cells (18,53,55–57). However, future studies will be required to determine whether diversity in rRNA 2'O-methylation profiles result from the intratumoral heterogeneity and/or heterogeneity in chemically modified rRNA molecules between and/or within a single cell.

The discovery of rRNA chemical modifications promoting ribosome plasticity will undoubtedly lead to a deeper and more complete understanding of rRNA chemical modifications in ribosome function and of cell fate determination providing a novel layer of cancer-related molecular fingerprinting.

DATA AVAILABILITY

Deposit FastQ and/or raw counts as dataset in GEO Profile (GSE143415).

To review GEO accession GSE143415:

Go to <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE143415>.

Enter token wpknuomppoxlmh into the box.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Cancer Online.

ACKNOWLEDGEMENTS

We thank Brigitte Manship for English editing.

Author contributions: V.Marcel, J.K., V.Marchand, H.P., F.N.V.L., L.A., V.B.I., P.L.M., D.M. and L.T. performed and analysed experiments. V.Marcel, J.K., H.P., L.T., Y.M. and A.V. performed statistical analyses. F.N.V.L., V.S., S.E.B., A.D., L.B.J., A.M.T., J.C.B., T.D. and F.A. provide human samples, clinical data and cellular models. V.Marcel, J.K., S.K.N., F.C., A.P., B.O.K. and J.J.D. analysed and interpreted the data. V.Marcel and J.J.D. shaped the research question and coordinated the project. V.Marcel, J.K., F.C. and A.V. supervised the daily experiments. V.Marcel, A.P. and J.J.D. raised funding. V.Marcel, J.K., S.K.N., B.P.K. and J.J.D. wrote the first draft of the manuscript.

FUNDING

Institut National du Cancer (INCa) [PLBIO program n°2019-138 MARACAS, PRT-K program n°2018-024 EMT-CoNCEPT, PAIR Sein program ARC.INCa.LNCC.7625 RiboTEM, and to B.P.K.]; Fondation ARC pour la recherche sur le cancer [20161204686 MARACAS.v0, and to B.P.K.]; ITMO Cancer [18CN044-00 RESIST]; SIRIC program [INCa-DGOS-Inserm.12563 LyRICAN]; Labex program [DEVCAN2UMAN]; Synergie Lyon Cancer Foundation; Fondation pour la Recherche Médicale (FRM) (to

B.P.K.); Ligue Nationale Contre le Cancer (LNCC) (to B.P.K.); Agence Nationale pour la Recherche (ANR) (to B.P.K.); USIAS of the University of Strasbourg [USIAS-2018-012 to B.P.K.]; LNCC Fellowship (to H.P., F.N.V.L.); FRM Fellowship (to F.N.V.L.). French Research Minister Fellowship (to P.L.M.); Centre National de la Recherche Scientifique (CNRS) (to S.K.N., F.C., B.K.); Institut National de la Santé et de la Recherche Médicale (Inserm) (to J.-J.D., V.Marcel); Breast Cancer Now Fellowship [2012MaySF127 to A.D., J.-C.B.]; Chief Scientist Office; NHS Scotland; University of Dundee.

Conflict of interest statement. None declared.

This paper is linked to: [doi:10.1093/narcan/zcaa035](https://doi.org/10.1093/narcan/zcaa035).

REFERENCES

1. Marcel, V., Ghayad, S.E., Belin, S., Therizols, G., Morel, A.P., Solano-González, E., Vendrell, J.A., Hacot, S., Mertani, H.C., Albaret, M.A. *et al.* (2013) P53 acts as a safeguard of translational control by regulating fibrillarin and rRNA methylation in cancer. *Cancer Cell*, **24**, 318–330.
2. Erales, J., Marchand, V., Panthu, B., Gillot, S., Belin, S., Ghayad, S.E., Garcia, M., Laforets, F., Marcel, V. *et al.* (2017) Evidence for rRNA 2'-O-methylation plasticity: Control of intrinsic translational capabilities of human ribosomes. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 12934–12939.
3. Shi, Z., Fujii, K., Kovary, K.M., Genuth, N.R., Röst, H.L., Teruel, M.N. and Barna, M. (2017) Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. *Mol. Cell*, **67**, 71–83.
4. Genuth, N.R. and Barna, M. (2018) Heterogeneity and specialized functions of translation machinery: from genes to organisms. *Nat. Rev. Genet.*, **19**, 431–452.
5. Helm, M. and Motorin, Y. (2017) Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.*, **18**, 275–291.
6. Belin, S., Beghin, A., Solano-González, E., Bezin, L., Brunet-Manquat, S., Textoris, J., Prats, A.-C., Mertani, H.C., Dumontet, C. and Diaz, J.-J. (2009) Dysregulation of ribosome biogenesis and translational capacity is associated with tumor progression of human breast cancer cells. *PLoS One*, **4**, e7147.
7. Gachet, S., El-Chaar, T., Avran, D., Genesca, E., Catez, F., Quentin, S., Delord, M., Therizols, G., Briot, D. *et al.* (2018) Deletion 6q drives T-cell leukemia progression by ribosome modulation. *Cancer Discov.*, **8**, 1614–1631.
8. D'Souza, M.N., Gowda, N.K.C., Tiwari, V., Babu, R.O., Anand, P., Dastidar, S.G., Singh, R., Selvaraja, B., Pal, R., Ramesh, A. *et al.* (2018) FMRP interacts with C/D box snoRNA in the nucleus and regulates ribosomal RNA methylation. *iScience*, **9**, 399–411.
9. Su, H., Xu, T., Ganapathy, S., Shadfan, M., Long, M., Huang, T.H.-M., Thompson, I. and Yuan, Z.-M. (2014) Elevated snoRNA biogenesis is essential in breast cancer. *Oncogene*, **33**, 1348–1358.
10. Esteller, M. (2011) Non-coding RNAs in human disease. *Nat. Rev. Genet.*, **12**, 861–874.
11. Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., Ruddy, K., Tsang, J. and Cardoso, F. (2019) Breast cancer. *Nat. Rev. Dis. Prim.*, **5**, 66–97.
12. Beekman, R., Chapaprieta, V., Russiñol, N., Vilarrasa-Blasi, R., Verdagué-Dot, N., Martens, J.H.A., Duran-Ferrer, M., Kulis, M., Serra, F., Javierre, B.M. *et al.* (2018) The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. *Nat. Med.*, **24**, 868–880.
13. Capper, D., Jones, D.T.W., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., Koelsche, C., Sahm, F., Chavez, L., Reuss, D.E. *et al.* (2018) DNA methylation-based classification of central nervous system tumours. *Nature*, **555**, 469–474.
14. Marchand, V., Blanloeil-Oillo, F., Helm, M. and Motorin, Y. (2016) Illumina-based RiboMethSeq approach for mapping of 2'-O-Me residues in RNA. *Nucleic Acids Res.*, **44**, 135–147.
15. Birkedal, U., Christensen-Dalsgaard, M., Krogh, N., Sabarinathan, R., Gorodkin, J. and Nielsen, H. (2015) Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew. Chem. Int. Ed.*, **54**, 451–455.
16. Khatter, H., Myasnikov, A.G., Natchiar, S.K. and Klaholz, B.P. (2015) Structure of the human 80S ribosome. *Nature*, **520**, 640–645.
17. Natchiar, S.K., Myasnikov, A.G., Kratzat, H., Hazemann, I. and Klaholz, B.P. (2017) Visualization of chemical modifications in the human 80S ribosome structure. *Nature*, **551**, 472–477.
18. Natchiar, S., Myasnikov, A., Hazemann, I. and Klaholz, B. (2018) Visualizing the role of 2'-OH rRNA methylations in the human ribosome structure. *Biomolecules*, **8**, 125–136.
19. Bourdon, J.-C., Khoury, M.P., Diot, A., Baker, L., Fernandes, K., Aoubala, M., Quinlan, P., Purdie, C.A., Jordan, L.B., Prats, A.-C. *et al.* (2011) p53 mutant breast cancer patients expressing p53y have as good a prognosis as wild-type p53 breast cancer patients. *Breast Cancer Res.*, **13**, R7.
20. Nguyen Van Long, F., Lardy-Cleaud, A., Bray, S., Chabaud, S., Dubois, T., Diot, A., Jordan, L.B., Thompson, A.M., Bourdon, J.-C. *et al.* (2018) Druggable *Nucleolin* identifies breast tumours associated with poor prognosis that exhibit different biological processes. *Cancers (Basel)*, **10**, 390–401.
21. Maire, V., Baldeyron, C., Richardson, M., Tesson, B., Vincent-Salomon, A., Gravier, E., Marty-Prouvost, B., De Koning, L., Rigault, G., Dumont, A. *et al.* (2013) TTK/hMPS1 is an attractive therapeutic target for triple-negative breast cancer. *PLoS One*, **8**, e63712.
22. André, F., Michiels, S., Dessen, P., Scott, V., Suciu, V., Uzan, C., Lazar, V., Lacroix, L., Vassal, G., Spielmann, M. *et al.* (2009) Exonic expression profiling of breast cancer and benign lesions: a retrospective analysis. *Lancet Oncol.*, **10**, 381–390.
23. Krogh, N. and Nielsen, H. (2018) Sequencing-based methods for detection and quantitation of ribose methylations in RNA. *Methods*, **156**, 5–15.
24. Birkedal, U., Christensen-Dalsgaard, M., Krogh, N., Sabarinathan, R., Gorodkin, J. and Nielsen, H. (2015) Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew. Chem. Int. Ed. Engl.*, **54**, 451–455.
25. Marchand, V., Ayadi, L., El-Hajj, A., Blanloeil-Oillo, F., Helm, M. and Motorin, Y. (2017) High-throughput mapping of 2'-O-Me residues in RNA using next-generation sequencing (Illumina RiboMethSeq Protocol). In: Lusser, A. (ed). *RNA Methylation: Methods and Protocols, Methods in Molecular Biology*, Springer, NY, pp. 171–187.
26. Pichot, F., Marchand, V., Ayadi, L., Bourguignon-Igel, V., Helm, M. and Motorin, Y. (2020) Holistic optimization of bioinformatic analysis pipeline for detection and quantification of 2'-O-Methylations in RNA by RiboMethSeq. *Front. Genet.*, **11**, 38.
27. Decatur, W.A. and Fournier, M.J. (2002) rRNA modifications and ribosome function. *Trends Biochem. Sci.*, **27**, 344–351.
28. Lo Monaco, P., Marcel, V., Diaz, J.J. and Catez, F. (2018) 2'-O-methylation of ribosomal RNA: towards an epitranscriptomic control of translation? *Biomolecules*, **8**, 106–119.
29. Maden, B.E.H., Corbett, M.E., Heeney, P.A., Pugh, K. and Ajuh, P.M. (1995) Classical and novel approaches to the detection and localization of the numerous modified nucleotides in eukaryotic ribosomal RNA. *Biochimie*, **77**, 22–29.
30. Maden, B.E.H. (1980) Methylation map of xenopus laevis ribosomal RNA. *Nature*, **288**, 293–296.
31. Sirum-Connolly, K. and Mason, T.L. (1993) Functional requirement of a site-specific ribose methylation in ribosomal RNA. *Science*, **262**, 1886–1889.
32. Baudin-Baillieu, A., Fabret, C., Liang, X.H., Piekna-Przybylska, D., Fournier, M.J. and Rousset, J.P. (2009) Nucleotide modifications in three functionally important regions of the *Saccharomyces cerevisiae* ribosome affect translation accuracy. *Nucleic Acids Res.*, **37**, 7665–7677.
33. Truitt, M.L. and Ruggero, D. (2016) New frontiers in translational control of the cancer genome. *Nat. Rev. Cancer*, **16**, 288–304.
34. Sharma, S., Marchand, V., Motorin, Y. and Lafontaine, D.L.J. (2017) Identification of sites of 2'-O-methylation vulnerability in human ribosomal RNAs by systematic mapping. *Sci. Rep.*, **7**, 11490.
35. Krogh, N., Jansson, M.D., Häfner, S.J., Tehler, D., Birkedal, U., Christensen-Dalsgaard, M., Lund, A.H. and Nielsen, H. (2016) Profiling of 2'-O-Me in human rRNA reveals a subset of fractionally

- modified positions and provides evidence for ribosome heterogeneity. *Nucleic Acids Res.*, **44**, 7884–7895.
36. Incarnato, D., Anselmi, F., Morandi, E., Neri, F., Maldotti, M., Rapelli, S., Parlato, C., Basile, G. and Oliviero, S. (2016) High-throughput single-base resolution mapping of RNA 2'-O-methylated residues. *Nucleic Acids Res.*, **45**, 1433–1441.
 37. Zhu, Y., Pirnie, S.P. and Carmichael, G.G. (2017) High-throughput and site-specific identification of 2'-O-methylation sites using ribose oxidation sequencing (RibOxi-seq). *RNA*, **23**, 1303–1314.
 38. Zhou, F., Liu, Y., Rohde, C., Pauli, C., Gerloff, D., Kohn, M., Misiak, D., Baumer, N., Cui, C., Gollner, S. *et al.* (2017) AML1-ETO requires enhanced C/D box snoRNA/RNP formation to induce self-renewal and leukaemia. *Nat. Cell Biol.*, **19**, 844–855.
 39. Hebras, J., Krogh, N., Marty, V., Nielsen, H. and Cavallé, J. (2020) Developmental changes of rRNA ribose methylations in the mouse. *RNA Biol.*, **17**, 150–164.
 40. Mourksi, N.-E.-H., Morin, C., Fenouil, T., Diaz, J.-J. and Marcel, V. (2020) snoRNAs offer novel insight and promising perspectives for lung cancer understanding and management. *Cells*, **9**, 541–565.
 41. Bywater, M.J., Pearson, R.B., McArthur, G.A. and Hannan, R.D. (2013) Dysregulation of the basal RNA polymerase transcription apparatus in cancer. *Nat. Rev. Cancer*, **13**, 299–314.
 42. Haimov, O., Sinvani, H., Martin, F., Ulitsky, I., Emmanuel, R., Tamarkin-Ben-Harush, A., Vardy, A. and Dikstein, R. (2017) Efficient and accurate translation initiation directed by TISU involves RPS3 and RPS10e binding and differential eukaryotic initiation factor 1A regulation. *Mol. Cell Biol.*, **37**, e00150-17.
 43. Malygin, A.A., Kossinova, O.A., Shatsky, I.N. and Karpova, G.G. (2013) HCV IRES interacts with the 18S rRNA to activate the 40S ribosome for subsequent steps of translation initiation. *Nucleic Acids Res.*, **41**, 8706–8714.
 44. Sundaramoorthy, E., Leonard, M., Mak, R., Liao, J., Fulzele, A. and Bennett, E.J. (2017) ZNF598 and RACK1 regulate mammalian ribosome-associated quality control function by mediating regulatory 40S ribosomal ubiquitylation. *Mol. Cell*, **65**, 751–760.
 45. Matsuo, Y., Ikeuchi, K., Saeki, Y., Iwasaki, S., Schmidt, C., Udagawa, T., Sato, F., Tsuchiya, H., Becker, T., Tanaka, K. *et al.* (2017) Ubiquitination of stalled ribosome triggers ribosome-associated quality control. *Nat. Commun.*, **8**, 159–173.
 46. DiGiuseppe, S., Rollins, M.G., Bartom, E.T. and Walsh, D. (2018) ZNF598 plays distinct roles in interferon-stimulated gene expression and poxvirus protein synthesis. *Cell Rep.*, **23**, 1249–1258.
 47. Bellodi, C., Krasnykh, O., Haynes, N., Theodoropoulou, M., Peng, G., Montanaro, L. and Ruggero, D. (2010) Loss of function of the tumor suppressor DKC1 perturbs p27 translation control and contributes to pituitary tumorigenesis. *Cancer Res.*, **70**, 6026–6035.
 48. Ruggero, D. (2003) Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification. *Science (80-)*, **299**, 259–262.
 49. McMahon, M., Contreras, A., Holm, M., Uechi, T., Forester, C.M., Pang, X., Jackson, C., Calvert, M.E., Chen, B., Quigley, D.A. *et al.* (2019) A single H/ACA small nucleolar RNA mediates tumor suppression downstream of oncogenic RAS. *Elife*, **8**, e48847.
 50. Babaian, A., Rothe, K., Girodat, D., Minia, I., Djondovic, S., Milek, M., Spencer Miko, S., Wieden, H.-J., Landthaler, M., Morin, G. *et al.* (2019) Loss of m¹acp³Ψ ribosomal RNA modification is a major feature of cancer. *Cell Rep.*, **31**, 107611.
 51. Ma, H., Wang, X., Cai, J., Dai, Q., Natchiar, S.K., Lv, R., Chen, K., Lu, Z., Chen, H., Shi, Y.G. *et al.* (2019) Methyladenosine methyltransferase ZCCHC4 mediates ribosomal RNA methylation. *Nat. Chem. Biol.*, **15**, 88–94.
 52. Aran, D., Camarda, R., Odegaard, J., Paik, H., Oskotsky, B., Krings, G., Goga, A., Sirota, M. and Butte, A.J. (2017) Comprehensive analysis of normal adjacent to tumor transcriptomes. *Nat. Commun.*, **8**, 1077–1091.
 53. Myasnikov, A.G., Kundhavi, Natchiar, S., Nebout, M., Hazemann, I., Imbert, V., Khatter, H., Peyron, J.-F.F. and Klaholz, B.P. (2016) Structure-function insights reveal the human ribosome as a cancer target for antibiotics. *Nat. Commun.*, **7**, 12856.
 54. Catez, F., Dalla Venezia, N., Marcel, V., Zorbas, C., Lafontaine, D.L.J. and Diaz, J.-J. (2019) Ribosome biogenesis: an emerging druggable pathway for cancer therapeutics. *Biochem. Pharmacol.*, **159**, 74–81.
 55. Boer, R.E. and Schneekloth, J.S. (2018) Targeting mammalian translational inhibition with tetracyclines. *Cell Chem. Biol.*, **25**, 1437–1438.
 56. Gilles, A., Frechin, L., Natchiar, K., Biondani, G., Loeffelholz, O. von, Holvec, S., Malaval, J.-L., Winum, J.-Y., Klaholz, B.P. and Peyron, J.-F. (2020) Targeting the human 80S ribosome in cancer: from structure to function and drug design for innovative adjuvant therapeutic strategies. *Cells*, **9**, 629–651.
 57. Pellegrino, S., Meyer, M., Zorbas, C., Bouchta, S.A., Saraf, K., Pelly, S.C., Yusupova, G., Evidente, A., Mathieu, V., Kornienko, A. *et al.* (2018) The amaryllidaceae alkaloid haemanthamine binds the eukaryotic ribosome to repress cancer cell growth. *Structure*, **26**, 416–425.

Supplemental informations

Ribosomal RNA 2'O-methylation as a novel layer of inter-tumour heterogeneity in breast cancer

Virginie Marcel^{1,#}, Janice Kielbassa², Virginie Marchand³, S. Kundhavai Natchiar^{4,5,6,7,8}, Hermes Paragindes^{1,2}, Flora Nguyen Van Long¹, Lilia Ayadi^{3,9}, Valérie Bourguignon-Igel^{3,9}, Piero Lo Monaco¹, Déborah Monchiet¹, Véronique Scott¹⁰, Laurie Tonon², Susan E. Bray¹¹, Alexandra Diot¹², Lee B. Jordan¹³, Alastair M. Thompson^{12,14}, Jean-Christophe Bourdon¹², Thierry Dubois¹⁵, Fabrice André¹⁰, Frédéric Catez¹, Alain Puisieux¹, Yuri Motorin^{3,9}, Bruno P. Klaholz^{4,5,6,7,8}, Alain Viari^{2,16}, Jean-Jacques Diaz^{1,#}

Supplementary material

Supplementary Methods

Human breast tumour samples. The series of 195 primary breast tumours include biological materials and clinical data (last update 2016, TTB, Dundee, Scotland, UK) maintained under ethical approval (REC Reference 07/S1402/90). Tumours were from women with previously untreated primary breast cancer gathered after macrodissection at surgical resection by a specialist breast cancer pathologist and snap-frozen (19). Hormone receptors (oestrogen ER and progesterone PR) and HER2/ErbB2 receptor status were determined by immunohistochemistry on FFPE sections (19). These three markers allowed identification of the main, clinically relevant, breast cancer subtypes: ER+ PR+/- HER2-, considered to be equivalent to the luminal subtypes; ER+/- PR+/- HER2+, to the HER2 amplified subtype (HER2+); and ER- PR- HER2-, to the triple-negative subtype (TNBC). The *TP53* mutation status was determined as previously described using the AmpliChip p53 Test (Roche Diagnostics) (19). The median follow-up of the 195 primary breast cancers was of 12.3 years. This series exhibited expected clinical characteristics (Supplementary Table ST1), in particular regarding association of overall survival with tumour size ($P < 0.0001^{***}$), lymph node invasion ($P = 0.0079^{**}$), grade ($P = 0.0151^*$) and breast cancer subtypes ($P = 0.0569$).

The series of mammoplasties issued from 7 female healthy donors (Institut Curie, France) corresponds mostly to non-transformed, non-proliferative but adipocyte cells (21). These female healthy donors are younger than the 195 breast cancer patients (mean age healthy donors = 42 years, about 17 years younger than breast cancer patients). Due to the small number of samples and the difference in age and cell origin, these samples were not compared to the 195 breast cancer patient series.

The series of 18 mammary samples (Institut Gustave Roussy, France) was designated either as malignant or benign tumours by a specialist breast cancer pathologist (Supplementary Table ST2) (22). Diagnosis of malignant tumours (n = 10) was based on the identification of breast cancer cells either on the core-needle biopsy or surgical specimen, while diagnosis of benign tumours (n = 8) was based on the lack of cancer cells together with the lack of radiological change after at least 3 months of follow-up. As for the Dundee samples, hormone and HER2 receptor statuses were determined by immunohistochemistry.

Breast cancer cell lines. Four breast cancer cell lines were maintained in 2D culture under 5% CO₂ atmosphere at 37°C. The MCF7 and BT474 cell lines considered to be representative of the luminal breast cancer subtype were maintained respectively, in MEM(Eagle) L-glutamine (-) (Thermofisher) supplemented with 10% foetal bovine serum (Sigma), 1% penicillin/streptomycin (Gibco), 1% glutamax (Gibco), 1.5 g/L sodium bicarbonate, 1 mM sodium pyruvate and 0.1 mM non-essential amino-acids or in RPMI1640 Glutamax supplemented with 10% foetal bovin serum and 1% glutamax. The MDA-MB-468 and MDA-MB-231 cell lines considered to be representative of the TNBC subtype were maintained respectively, in RPMI1640 Glutamax or DMEM/F12 Glutamax, the two media being supplemented with 10% foetal bovine serum and 1% glutamax.

RiboMeth-seq. Levels of rRNA 2'O-methylation at the 106 rRNA 2'O-methylated sites were determined in biological samples by RiboMeth-seq, the only approach allowing quantification of rRNA 2'O-methylation level to date (14, 15, 23, 25). The C-score, which reflects the 2'O-methylation level, varies from 0 to 1; a C-score of 0 meaning that all the rRNA molecules are not 2'O-methylated at the given site, a C-score of 1 indicating that all the rRNA are fully 2'O-methylated at the given site, and a $0 < \text{C-score} < 1$ meaning that the sample displays a mix of 2'O-methylated and un-2'O-methylated rRNA molecules at the given site.

RiboMeth-seq was performed as previously described using the Illumina sequencing technology (14, 25). Briefly, total RNA was first subjected to partial alkaline hydrolysis to generate RNA fragments ranging from 20 to 50 bp. Libraries were then prepared using NEBNext Multiplex Small RNA Library Prep Set (Illumina) and multiplexed. Sequencing was performed using HiSeq1000 sequencer (Illumina) in SR50 mode.

Bioinformatics pre-processing. After sequencing, FastQ files were pre-processed, aligned and C-scores were computed from raw counts as previously described (25). Briefly, Trimmomatic v0.36 was run to remove adapter sequences from the reads and to discard reads shorter than 17 nucleotides. The remaining reads were aligned to a human rRNA sequence (NR_046235) using Bowtie 2.0. 5'/3' ends of reads were counted at each rRNA nucleotide position and used to compute a 2'O-methylation score, the C-score, as described (15). The C-score is defined as the level of 2'O-methylation at a given site and is computed from the read coverage. More precisely, the C-score at a position is calculated from the ratio of the 5'/3' end reads raw count at the position i to the local read coverage, where the latter is inferred from the 5'/3' end read raw counts of the flanking positions (i.e., $k=6$ nucleotides located upstream and downstream from i).

Numbering of rRNA 2'O-methylated sites was assigned according to the numbering in the snoRNA-LBME-db to facilitate comparison between studies. A table is given with the correspondence of the nomenclature used in the present study and in that of Natchiar and colleagues (Supplementary Table ST4) (17).

As this study corresponds to the first rRNA 2'O-methylation profiling of human samples, we assessed the performance of the C-Score by computing six different versions of the C-score using 5'/3' end read counts at the 106 rRNA 2'O-methylated sites derived from the human RNA reference sample (Supplementary Figure S1a). We varied the window size of the neighbouring nucleotides ($k=6, 4, \text{ or } 2$) and used either the arithmetic mean or the median of raw read counts

as an estimator of the coverage (i.e., total rRNA molecules). The gold-standard corresponds to the C-score of the 106 rRNA 2'O-methylation sites calculated using arithmetic mean of a 6 window size previously described (14, 15, 25). By considering each score value as a threshold, we computed the sensitivity and specificity of the different C-score versions at the 106 rRNA methylated sites and plotted the ROC (*Receiver Operating Characteristics*) curve (Supplementary Figure S1a). In the present study, a C'-score that was computed using a 6 nucleotide window and a normalization against the median raw counts of neighbouring nucleotides was found as the best performer. This C'-score was termed C-score throughout the manuscript.

Quality control. We performed exhaustive quality controls (QC) to assess the technical reproducibility and robustness of the RiboMeth-seq protocol and to ensure the high quality of the data for further downstream analyses of the human tumour series. To this end, we first used the human RNA reference sample as a calibrated source of rRNA. Five of the human RNA reference samples were processed in parallel with different batches of human tumour samples. Second, 20 primary breast tumours corresponding to randomly chosen samples were processed independently twice to obtain a set of technical duplicates from relevant biological tissue.

We computed Spearman correlations between replicates both in terms of raw counts and C-scores at all genomic positions of the rRNA sequences, as well as at the 106 2'O-methylated sites exclusively. To assess the reproducibility of the data, we calculated the QC metrics for the five replicates of the human RNA reference, as well as for the 20 tumour duplicates (Supplementary Fig. 1b-c and 2). All replicates showed a high technical reproducibility and very strong correlations between each other, emphasizing the reproducibility and robustness of the RiboMeth-seq protocol even when sequencing delicate biological material such as frozen tumour biopsies.

Analysis of the breast cancer series data. In our first exploratory, unsupervised data analysis, we performed hierarchical clustering (using Manhattan as distance and Ward's linkage method), and principal component analysis (PCA) using the C-scores at the 106 individual rRNA 2'O-methylated sites of the 195 samples. Both methods lead to the detection of similarities between samples and potentially to define groups among them. Hierarchical clustering yields a heatmap (i.e., a coloured matrix) that represents a grid of values where the rows correspond to the samples and the columns to the rRNA 2'O-methylated sites. Thus, each rectangle in a heatmap represents one C-score at a given rRNA 2'O-methylated site of a sample and is coloured according to the numerical value in the matrix, with a range of colours proportionate to the C-score. Samples and sites are ranked according to the hierarchical clustering dendrogram. Dendrograms thus identify subgroups (clusters of both breast tumours exhibiting the most similar rRNA 2'O-methylation profiles and rRNA 2'O-methylated sites exhibiting the most similar level) across the 195 primary breast tumours.

PCA yields a lower dimensional representation of the samples, preserving much of the variance of the original data set. The PCA plot shows the first two principal components and the amount of variance explained by each component.

rRNA 2'O-methylation profiles. To summarize the data, we constructed box-and-whiskers plots (boxplot) for each of the 106 rRNA 2'O-methylated sites. The ends of the box are the lower (Q1), and upper (Q3) quartiles (i.e., C-score delimiting 25% and 75% of the population), so the box spans the interquartile range (IQR). The horizontal line inside the box represents the median C-score. The lines outside the box (top and bottom whiskers) extend to the highest and lowest observations the 97.5% and 2.5th. Line charts and barcharts represent the mean and standard deviation of C-scores across the samples at each rRNA 2'O-methylated sites.

Classification of the rRNA 2'O-methylated sites into variable and stable ones. We observed that the variability of the C-scores across samples was depended on the rRNA 2'O-methylated site. The total variability observed at a given 2'O-methylated site is a composition of measurement error, biological noise and intrinsic biological variation. Although the first two terms should be independent of the 2'O-methylated site, this may not be the case for the intrinsic variation. We thus hypothesized that the total variability at a specific 2'O-methylated site beyond a certain threshold is mainly driven by its intrinsic biological variation that will allow the identification of two distinct classes of rRNA 2'O-methylated sites. To define this threshold, rRNA 2'O-methylated sites from the 195 breast tumor samples and IQR values were sorted in ascending order. We define empirically a threshold at the rRNA 2'O-methylation site. The cut-off site was defined as the site where the increase of the IQR between two successive sites is not constant anymore. 2'O-methylated sites having IQR values below the threshold were termed "stable", while sites having IQR values above were designated as "variable" (Supplementary Figure S3c).

Association of rRNA 2'O-methylation with biological and clinical characteristics.

Between-group comparisons were performed using Fisher's exact test for categorical data. Comparison of rRNA 2'O-methylation levels between two groups of breast tumours was performed using non-parametric Mann-Whitney tests. Bonferroni or FDR (False Discovery Rate) correction methods were applied for multiple comparisons (P_{adj}). Significant association was considered when the adjusted P -value < 0.05 and ΔC -score (C-score condition 1 – C-score condition 2) > 0.05 . Results were summarised using bubble plots where $-\log_{10}(P_{adj})$ are represented by blue colour scale, absolute difference in C-score between the two conditions of interest by circle size. Survival curves for overall survival (OS) with associated log-rank tests were generated using the Kaplan Meier method. OS corresponded to the timing from the date of diagnosis to either the date of death or last follow-up for censored patients. Survival median was estimated using the inverted Kaplan-Meier method.

Supplementary Tables

Supplementary Table ST1. Characteristics of TTB breast tumours.

		Human breast tumour series (n = 195)
Age (years)	Median (min – max)	59.0 (28.0 – 90.0)
Invasive grade	No data	3
	Grade 1	12 (6.3 %)
	Grade 2	49 (25.5 %)
	Grade 3	131 (68.2 %)
Tumour size (mm)	n Median (min – max)	192 28.0 (2.0 – 110.0)
Number of invaded lymph node	n Median (min – max)	195 1.00 (0.0 – 41.0)
Invaded lymph node	No (n = 0) Yes (n ≥ 1)	88 (45.1 %) 107 (54.9 %)
ER status	No data Negative Positive	1 69 (35.6 %) 125 (64.4 %)
PR status	No data Negative Positive	1 110 (56.7 %) 84 (43.3 %)
HER status	No data Negative Positive	13 113 (62.1 %) 69 (37.9 %)
Breast cancer subtype	No data ER+ PR+/- HER2- ER+/- PR+/- HER2+ ER- PR- HER2-	13 73 (40.1 %) 69 (37.9 %) 40 (22.0 %)
TP53 mutation status	No data Wild-type Mutant	13 127 (69.8 %) 55 (30.2 %)

Supplementary Table ST2. Characteristics of malignant and benign breast lesions (IGR).

Malignant tumours	n = 10
Luminal	8 (80 %)
Her2+	0 (0 %)
TNBC	2 (20 %)
Carcinoma in situ	1 (10 %)
Benign tumours	n = 8
Fibroadenoma	5 (62.5 %)
Cyst	2 (25 %)
Phyllodes	1 (12.5 %)

Supplementary Table ST3. Comparison of rRNA 2'O-methylation level between clinical breast cancer characteristics at four significant rRNA sites.

	18S-Gm1447	28S-Gm1303	28S-Gm4588	18S-Am576
Breast cancer subtype				
<i>Luminal (n=73) vs HER2+ (n=69)</i>				
Median [£] (Lum.)	0.75	0.87	0.83	0.82
Median [£] (HER2+)	0.70	0.83	0.79	0.84
∂ median [#]	-0.052	-0.039	-0.044	0.018
P.adj [§]	0.0180	0.0180	0.0180	0.8260
<i>Luminal (n=73) vs TNBC (n=40)</i>				
Median [£] (Lum.)	0.75	0.87	0.83	0.82
Median [£] (TNBC)	0.63	0.81	0.79	0.94
∂ median [#]	-0.118	-0.058	-0.043	+0.118
P.adj [§]	2.57e-05	0.0070	0.0151	5.57e-08
<i>HER2+ (n=69) vs TNBC (n=40)</i>				
Median [£] (HER2+)	0.70	0.83	0.79	0.84
Median [£] (TNBC)	0.63	0.81	0.79	0.94
∂ median [#]	-0.065	-0.019	0.0004	0.100
P.adj [§]	0.0933	0.7142	0.9860	2.85e-06
Receptor status				
<i>ER+ (n=125) vs ER- (n=69)</i>				
Median [£] (ER+)	0.75	0.86	0.82	0.84
Median [£] (ER-)	0.65	0.81	0.77	0.91
∂ median [#]	-0.102	-0.048	-0.05	0.075
P.adj [§]	6.80e-06	0.0010	0.0005	5.82e-06
<i>PR+ (n=84) vs PR- (n=110)</i>				
Median [£] (PR+)	0.75	0.86	0.82	0.83
Median [£] (PR-)	0.69	0.82	0.78	0.87
∂ median [#]	-0.058	-0.038	-0.044	0.034
P.adj [§]	0.0011	0.0014	0.0007	0.0007
<i>HER2+ (n=69) vs HER2- (n=113)</i>				
Median [£] (HER2+)	0.70	0.82	0.79	0.84
Median [£] (HER2-)	0.73	0.86	0.82	0.86
∂ median [#]	0.031	0.027	0.031	0.016
P.adj [§]	0.8095	0.6630	0.6386	0.6630
Tumor grade				
<i>Grade 1 (n=12) vs Grade 2 (n=49)</i>				
Median [£] (G1)	0.78	0.89	0.83	0.86
Median [£] (G2)	0.76	0.88	0.83	0.83
∂ median [#]	-0.025	-0.017	-0.008	-0.025
P.adj [§]	0.9911	0.9911	1.0000	0.9911

	18S-Gm1447	28S-Gm1303	28S-Gm4588	18S-Am576
<i>Grade 1 (n=19) vs Grade 3 (n=131)</i>				
Median [£] (G1)	0.78	0.89	0.83	0.86
Median [£] (G3)	0.69	0.83	0.80	0.86
∂ median [#]	-0.091	-0.068	-0.039	-4.04e-05
P.adj [§]	0.1763	0.1763	0.5386	0.8548
<i>Grade 2 (n=49) vs Grade 3 (n=131)</i>				
Median [£] (G2)	0.76	0.88	0.83	0.83
Median [£] (G3)	0.69	0.83	0.80	0.86
∂ median [#]	-0.067	-0.050	-0.031	0.025
P.adj [§]	0.0008	0.0008	0.0227	0.0635

[£]Median C-score; [#]Median C-score difference (condition 2 – condition 1); [§]P-value adjusted for multiple comparison; Significant when both P.adj < 0.05 and absolute (∂ median) > 0.05

Supplementary Table ST4. Correspondence of rRNA 2'O-methylation position name between the present study and Natchiar et al Nature 2017. (see Excel supplementary file)

Supplementary Figure Legends

Supplementary Figure S1. Technical variation of rRNA 2'O-methylation profiling in human RNA reference sample. **a**, As this study corresponds to the first rRNA 2'O-methylation profiling of human samples, we first assessed the performance of rRNA 2'O-methylation levels (or C-score). C-scores were determined at the 106 rRNA 2'O-methylated sites using the innovative RiboMeth-seq method, in five independent technical replicates of the human RNA reference sample. Performance of the C-score was assessed by computing ROC curve using six different versions of the C-score calculated with both variable numbers of neighbouring nucleotides ($k=6, 4$ or 2) and either the arithmetic mean or the median raw counts (table, right side). Compared to the original description of C-score (gold-standard: violet C-score M6), we determined a C'-score (termed C-score throughout the manuscript) based on the median raw counts and the 6 neighbouring positions (C-score- Med6) as the most robust measure of rRNA 2'O-methylation levels using human RNA reference. **b**, To verify the reproducibility and robustness of the RiboMeth-seq method using human tumour samples, we developed quality controls based on different metrics. Quality control was performed by comparing correlation (top) and profiles (bottom) of both 5'/3' end reads raw counts (left) and C-scores (right) of the longest 28S rRNA, between the five technical replicates of the human RNA reference sample. Strong reproducibility and robustness was observed using these two approaches (5'/3' end reads raw counts: $r^2 > 0.96$; C-scores: $r^2 > 0.92$ at the non 2'O-methylated sites and $r^2 > 0.88$ at the 106 2'O-methylated sites). **c**, Boxplot presents C-score variation among the five technical replicates of human RNA reference at each of the 106 rRNA 2'O-methylated sites ranked by increasing rRNA molecules and nucleotide positions. About 75% of rRNA 2'O-methylated sites exhibited only minor variation in their C-score across the five technical replicates ($\partial C\text{-score} = C\text{-score max} - C\text{-score min}$, $\partial C\text{-score} < 0.069$), the other sites having a $\partial C\text{-score}$ ranging from 0.070 to 0.2047. These data show the strong technical reproducibility and robustness of the RiboMeth-seq approach.

Supplementary Figure S2. Technical variation of rRNA 2'O-methylation profiling in 20 human primary breast tumours. **a**, In addition to assessing the reproducibility and robustness of the RiboMeth-seq approach using the human RNA reference sample, we also analysed in duplicate a test series composed of randomly chosen 20 human breast tumour samples. Example of quality control of RiboMeth-seq performed in two technical duplicates of human primary breast tumours that show the highest (top) and the lowest (bottom) duplicate quality. Correlation (top) and profiles (bottom) of 5'/3' end reads raw counts of 5.8S, 18S and 28S rRNAs were compared between the technical duplicates. While a r^2 correlation coefficient

ranging from 0.64 to 0.85 and depending on the considered rRNA was determined in the “lowest duplicate quality” sample, a strong correlation with $r^2 > 0.99$ was observed for the three rRNAs in the “highest duplicate quality” human breast sample. **b**, The r^2 correlation coefficient for the three rRNAs between technical duplicate of the 20 human breast tumours was shown as a cumulative level. The lowest, but nevertheless showing strong r^2 correlation coefficient, was observed for the 5.8S possibly due to the fact it corresponds to the smallest rRNA molecule (i.e., 157 bp compared to 1875 or 5035 bp for 18S and 28S, respectively) and is thus associated with only a small number of raw data. More than 50%, 85% and 65% of technical duplicates showed a $r^2 > 0.8$ for 5.8S, 18S and 28S rRNA respectively, further demonstrating the strong reproducibility and robustness of the RiboMeth-seq approach even in delicate biological material such as human frozen biopsies.

Supplementary Figure S3. Variability of rRNA 2’O-methylation level between the 195 human primary breast tumours. **a**, Boxplot presents C-score distribution among the 195 human primary breast tumours at each of the 106 rRNA 2’O-methylated sites ranked by increasing rRNA molecule and nucleotide positions. Median C-scores vary depending on the rRNA 2’O-methylated sites, the rRNA 2’O-methylation level showing strong variation across human primary breast tumours with C-score ranging from 0 to about 0.95. It has to be noted that a null C-score was observed at three rRNA 2’O-methylation sites (28S-Am1313, 28S-Cm2352 and 28S-Gm3723; stars) in three different breast tumours. These data show the high variability of rRNA 2’O-methylation level between human breast tumour samples. **b**, To determine whether variation in rRNA 2’O-methylation levels among the human breast tumour samples corresponds to technical or biological variation, the ∂ C-score [(mean C-score + 2 standard deviation) – (mean C-score – 2 standard deviation)] between the 195 primary breast tumours (black) and the five technical replicates of the human RNA reference sample (white) was compared. ∂ C-score of human breast samples were higher than the ∂ C-score of the human RNA reference sample, indicating that variation in rRNA 2’O-methylation levels observed between human breast tumours corresponds to biological variation rather than technical issues. **c**, We defined two classes of rRNA 2’O-methylated sites based on the variability of their level among the patients based on the C-score inter-patient variability measured by interquartile range (IQR). The plot represents the IQR of C-scores calculated using the 195 human primary breast tumours at each of the 106 rRNA 2’O-methylated sites ranked by increasing IQR. The cut-off site is the site from which the IQR values no longer lie on this straight line, i.e. beyond which a straight line should change its slope to pass through them. This allowed the determination of two classes of rRNA 2’O-methylated sites: the ones with the most stable C-score; and the ones with the most variable C-score. Overall, these data

indicate that 60% of the rRNA 2'O-methylated sites correspond to chemically modified nucleotides with conserved rRNA 2'O-methylation levels in humans, while 40% of these sites can be either 2'O-methylated or not.

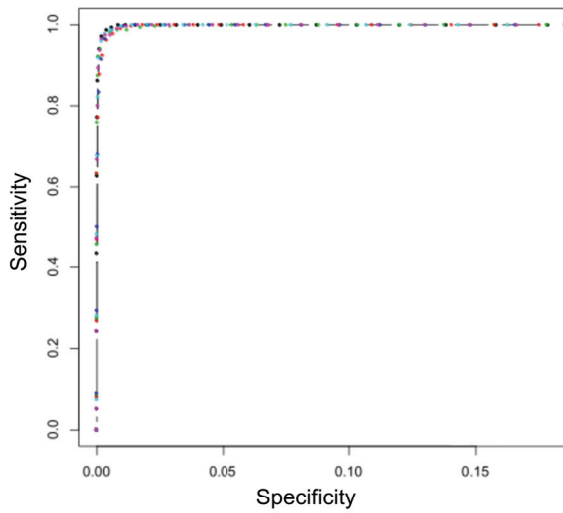
Supplementary Figure S4. Examples of variability in rRNA 2'O-methylation levels at three rRNA sites between breast tumours. The plot presents the C-score for each of the 195 human primary breast tumours at three rRNA 2'O-methylated sites that show either the smallest variability (**a**, 28S-Am3703), an intermediate one (**b**, 28S-Cm2811) or the highest variability (**c**, 18S-Gm1447). The red dotted lines indicate the mean \pm 2 standard deviations encompassing 95% of the primary breast tumours. These data illustrate the two classes of stable and variable rRNA 2'O-methylated sites with distinct degrees of variability between human breast tumour samples.

Supplementary Figure S5. rRNA 2'O-methylation profiling of 7 mammoplasties derived from healthy female donors. rRNA 2'O-methylation has been profiled using RiboMeth-seq approach and variability in C-score among 7 healthy donors is represented using box-plots. Compared to the classification derived from the 195 breast tumours (stable sites: red bar; variable sites: blue bar), it appears that most of the sites classified as either "stable" or "variable" in breast tumours are conserved in healthy donors. Only few sites didn't exactly fit this classification. This suggests either a lack of statistical power regarding the small series of 7 mammoplasties or that variability in rRNA 2'O-methylation is tissue-specific, the mammoplasties mostly corresponding to adipocytes.

Supplementary Figure S6. Variability of rRNA 2'O-methylation level between breast cancer cell lines. **a**, rRNA 2'O-methylation was profiled in luminal breast cancer cell lines (MCF-7 and BT474) and triple-negative breast cancer cell lines (MDA-MB-468 and MDA-MB-231). Mean C-score of two biological replicates issued from the different cell lines of each subtypes is presented at the 106 rRNA 2'O-methylated sites. The most stable (red) and the most variable (blue) rRNA 2'O-methylated sites are shown as identify in the series of 195 breast tumours (Fig. 2a). The most variable rRNA 2'O-methylated sites identified in the series of 195 primary breast tumours displayed also a strong variability in these breast cell lines. The rRNA 2'O-methylated sites, whose level was significantly associated with malignant mammary tumours, breast cancer subtypes and tumour grades in the 195 tumours samples, are indicated by a black arrow. *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$; ****: $P < 0.0001$. **b**, Variation of rRNA 2'O-methylation level at the sites associated with clinical and biological characteristics of breast tumours are shown in the different cell lines. While C-score is not associated with breast

cancer cell line subtypes at both 28S-Gm1303 and 28S-Gm4588 rRNA 2'O-methylated sites like in human breast tumour samples, C-score is strongly coherent among the two different subtypes of breast cancer cell lines at 18S-Gm1447 and 18S-Am576 sites. rRNA 2'O-methylation is almost absent in the TNBC (MDA-MB-468 and MDA-MB-231) cell lines at 18S-AmGm1447, while it is 2'O-methylated, even if at low level, in luminal breast cancer (MCF-7 and BT474) cell lines. In contrast, a decrease in rRNA 2'O-methylation level at 18S-Am576 site is observed in the luminal breast cancer cell lines compared to the TNBC cell lines.

Supplementary Figure S7. Association of rRNA 2'O-methylation profiles with biological and clinical breast cancer characteristics. Differences in the repartition of lymph node invasion status (a), tumour size (b), *TP53* mutation (c), oestrogen receptor (d), progesterone receptor (e) and HER2 (f) statuses among the 4 clusters of breast tumours defined on the basis of their rRNA 2'O-methylation profiles were assessed by statistical enrichment analyses. No significant difference was observed for tumour size, lymph node invasion and *TP53* mutation statuses while a significant difference in repartition between the 4 clusters was observed for ER, PR and HER2 statuses, supporting the association observed between these 4 clusters and breast cancer subtypes (Figure 3d).

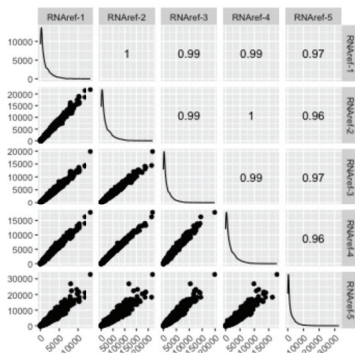
a

Method for average counts	Number of neighbouring positions	C-score
mean	6	C-score M6
	4	C-score M4
	2	C-score M2
median	6	C-score Med6
	4	C-score Med4
	2	C-score Med2

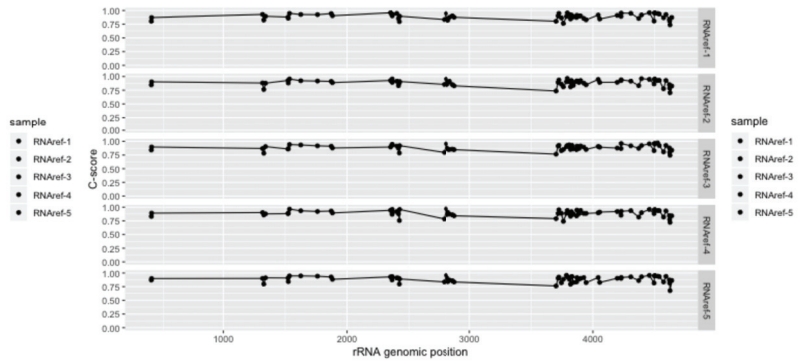
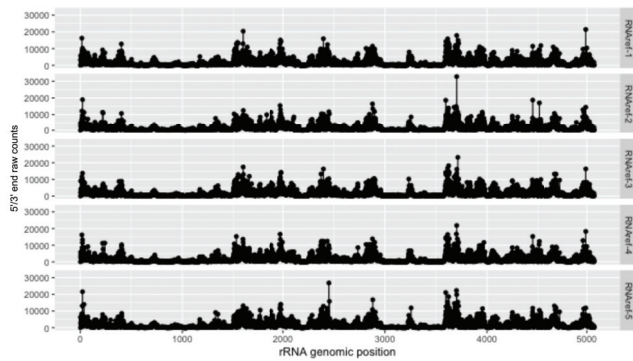
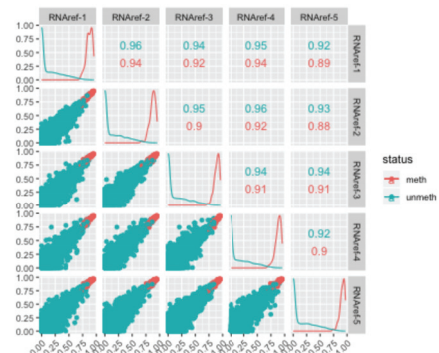
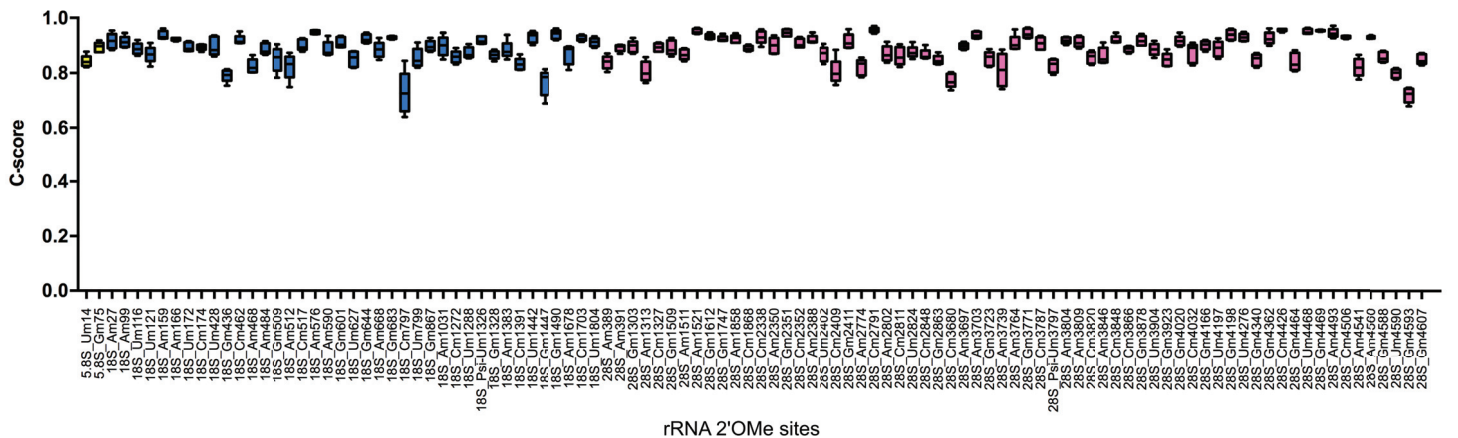
(Birkedal et al, 2015; Marchand et al, 2016)

b

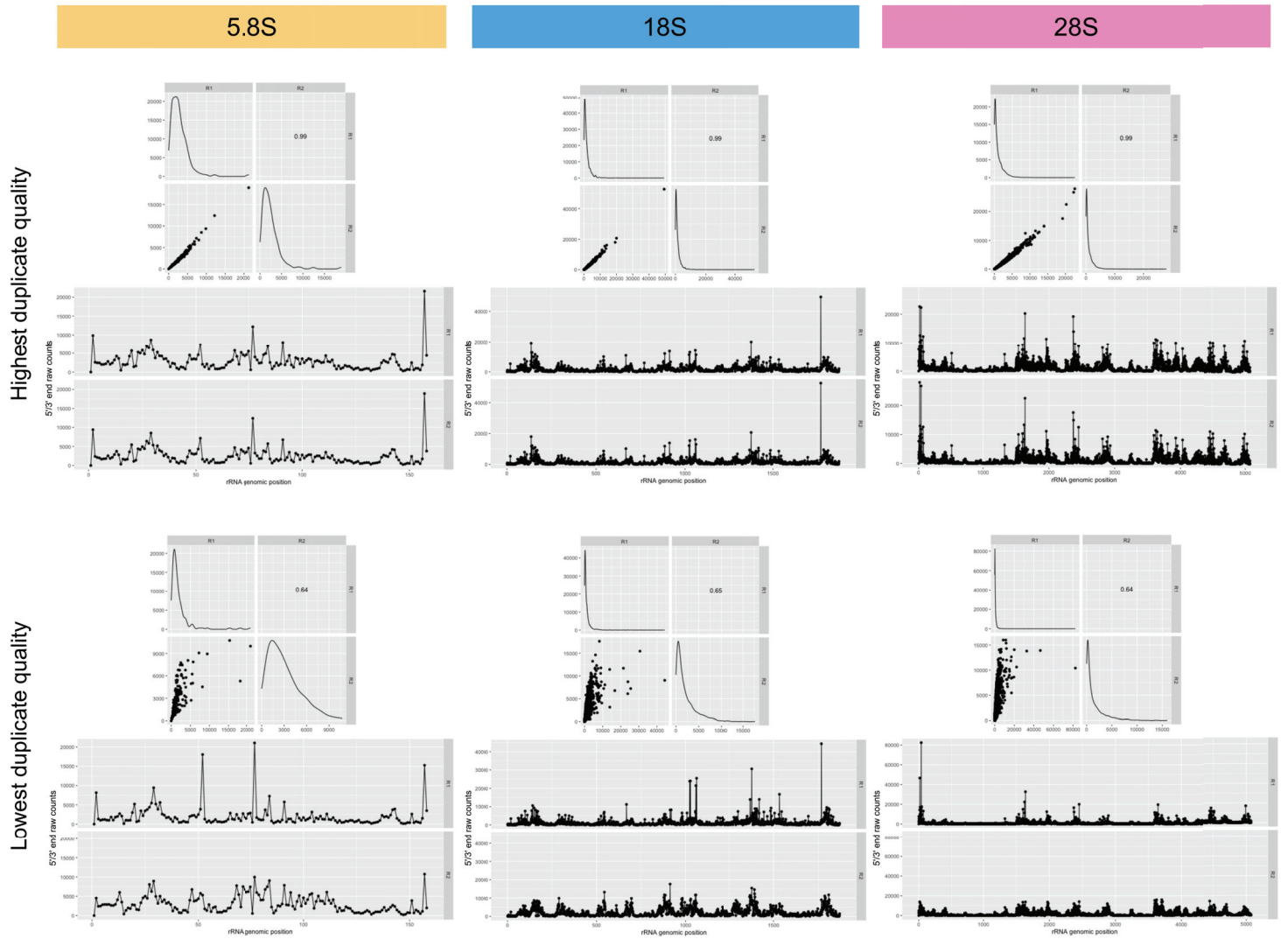
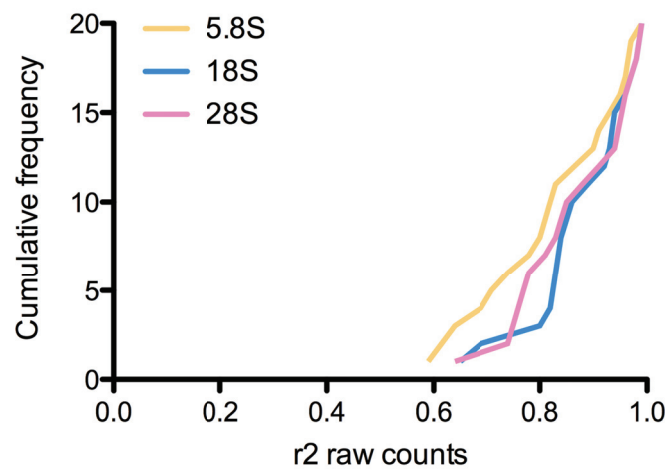
28S raw counts

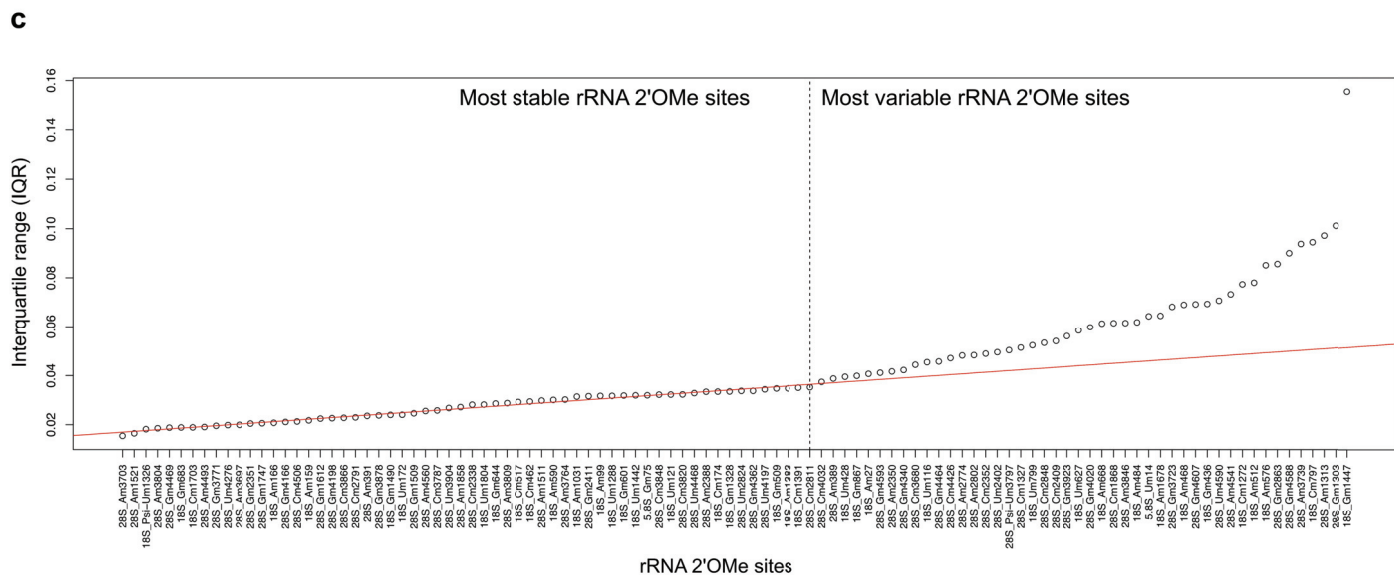
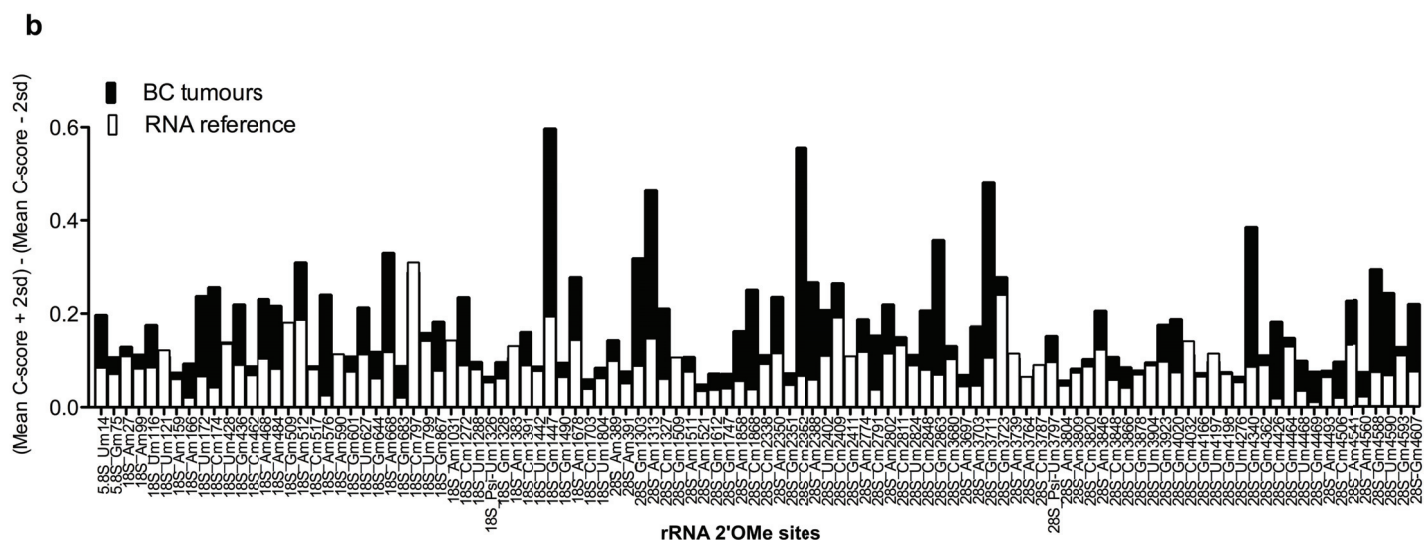
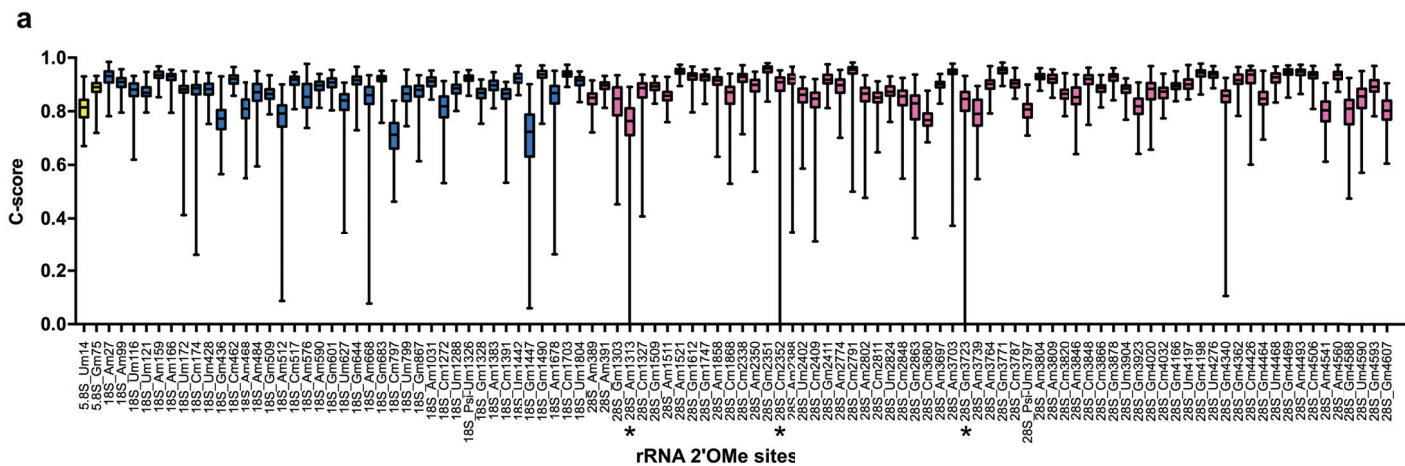


28S C-scores

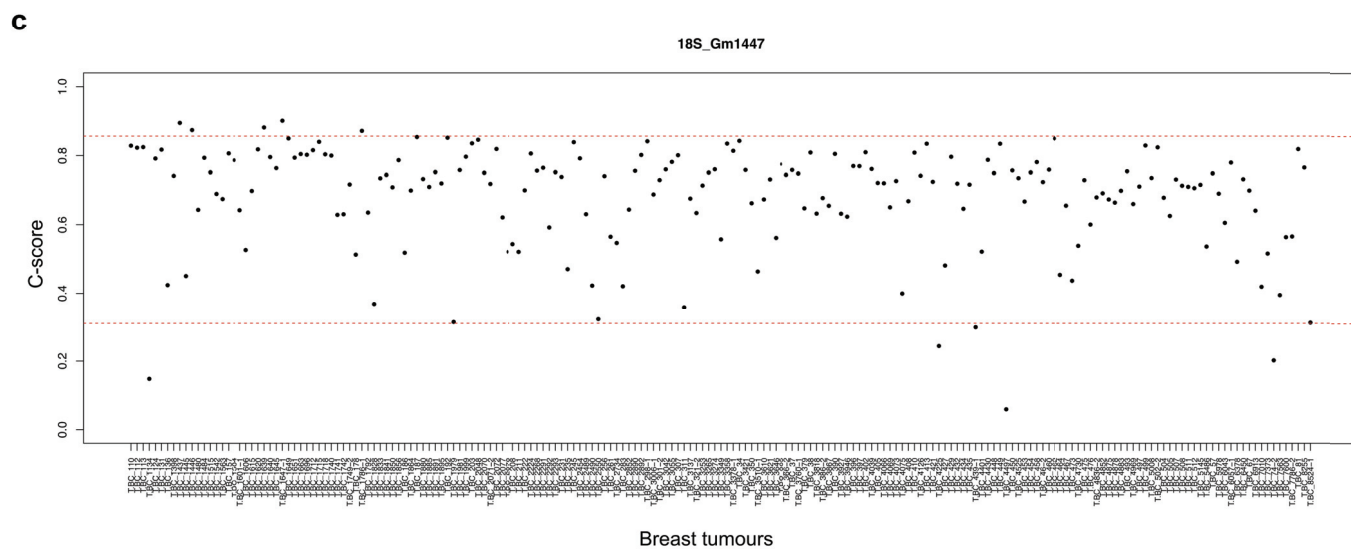
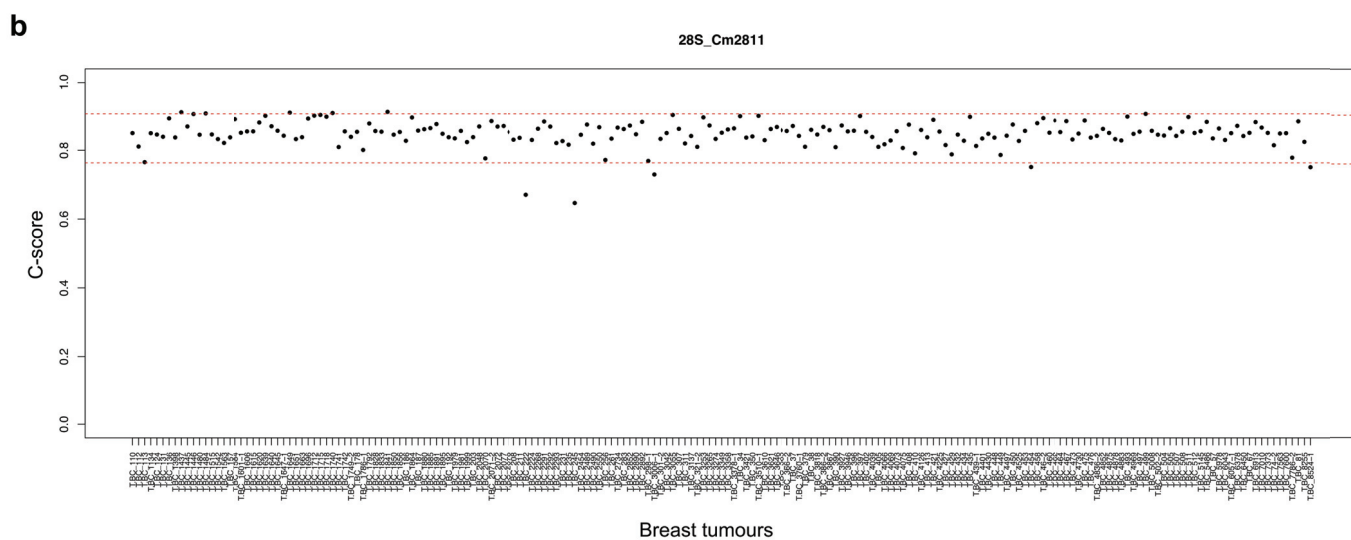
**c**

Supplementary Figure 1

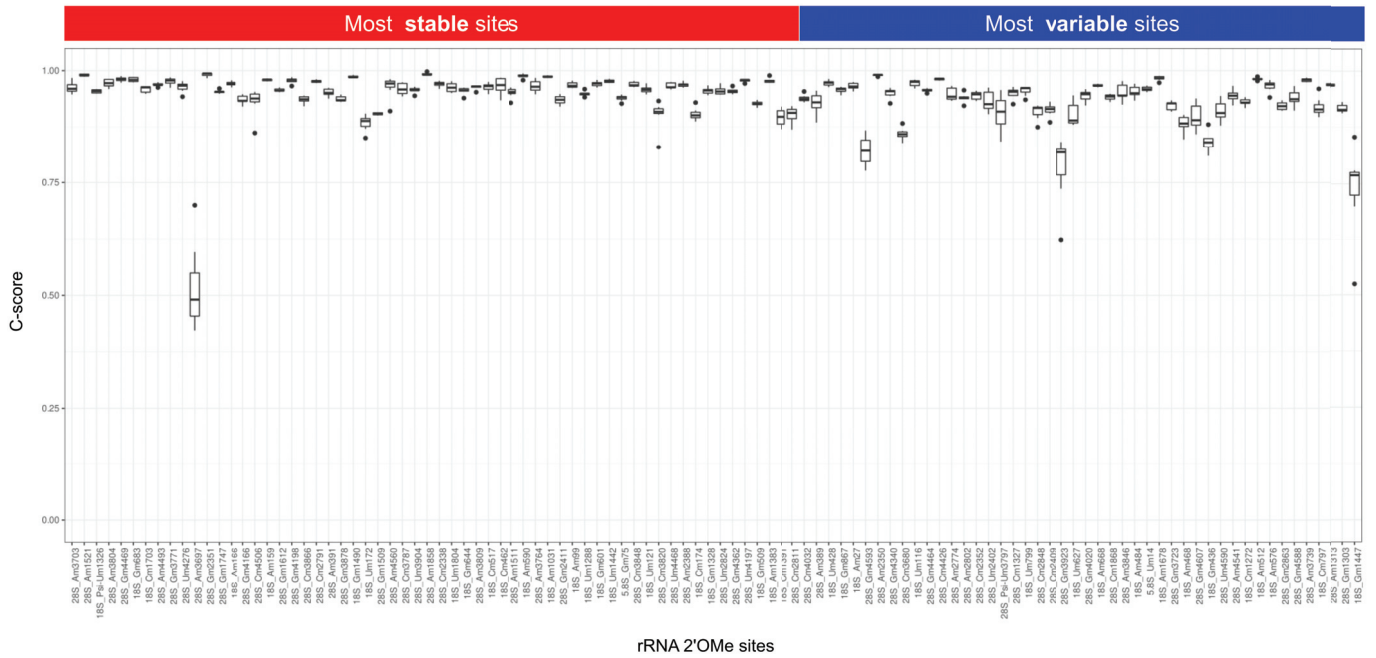
a**b**



Supplementary Figure 3

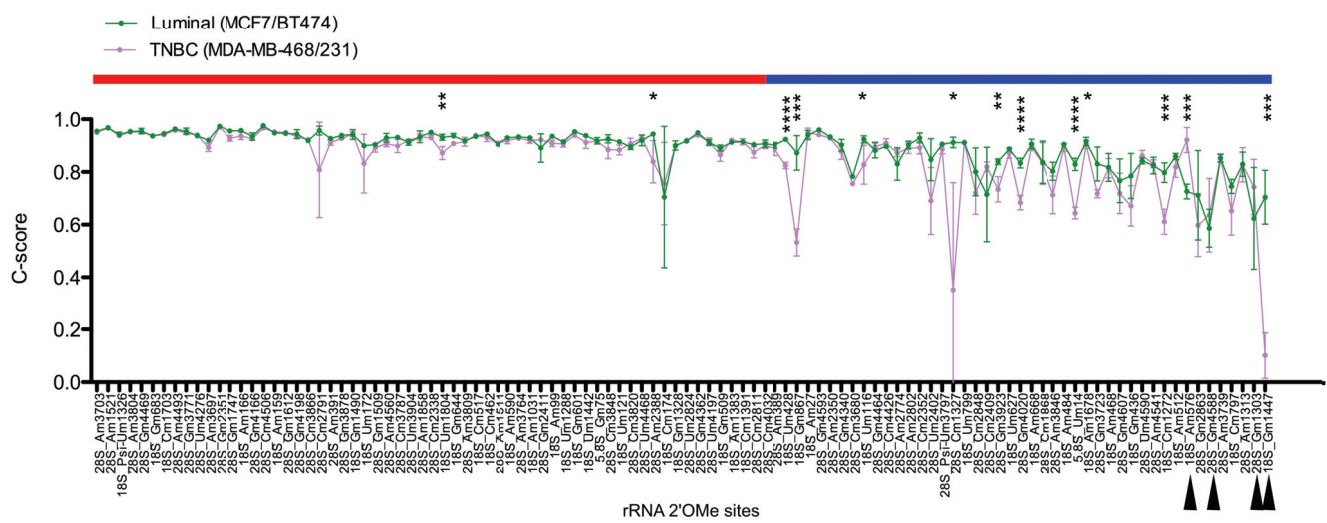


Supplementary Figure 4

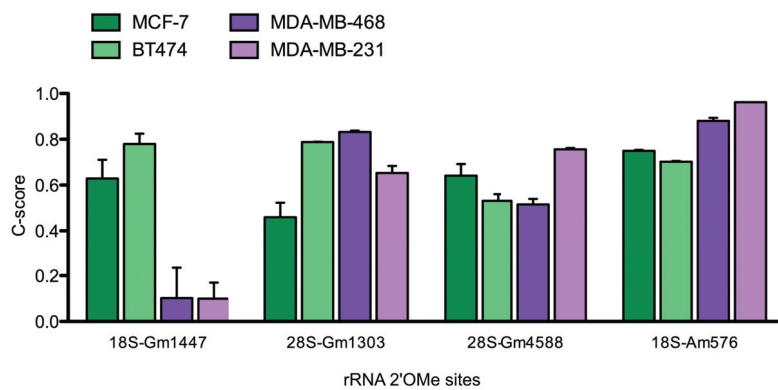


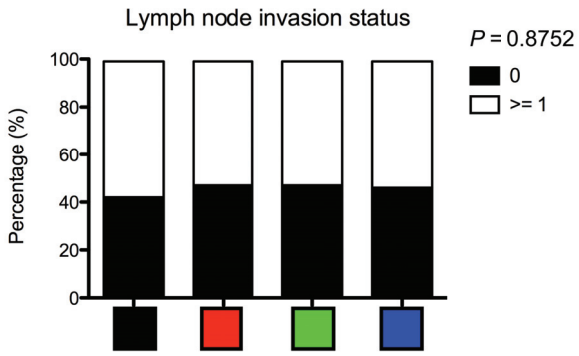
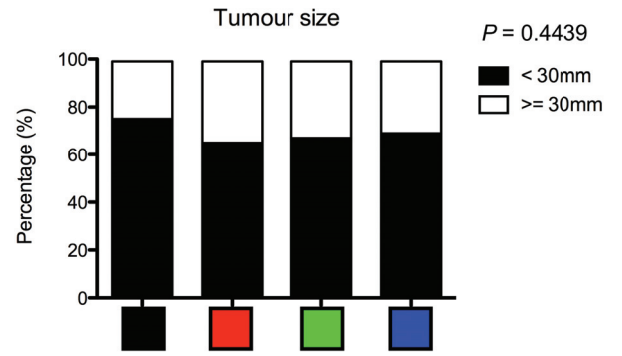
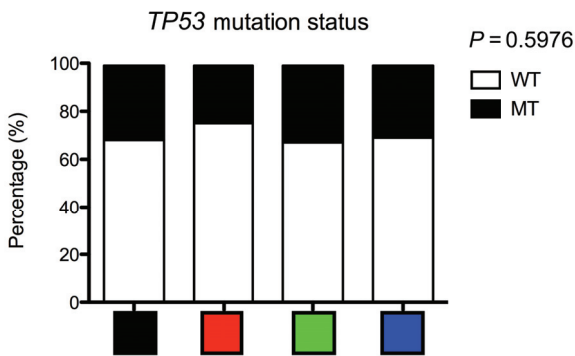
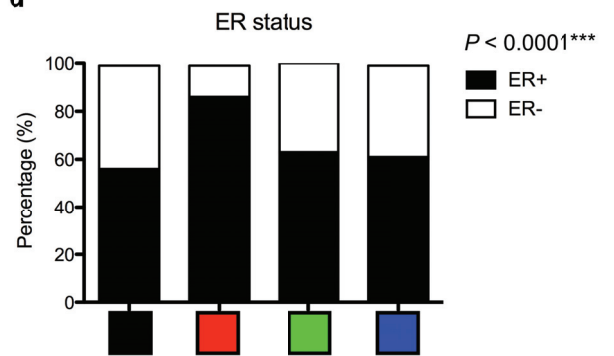
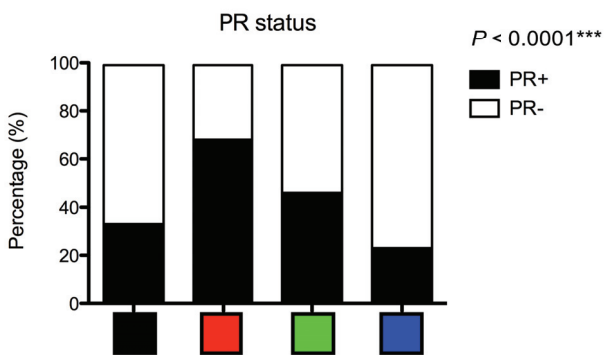
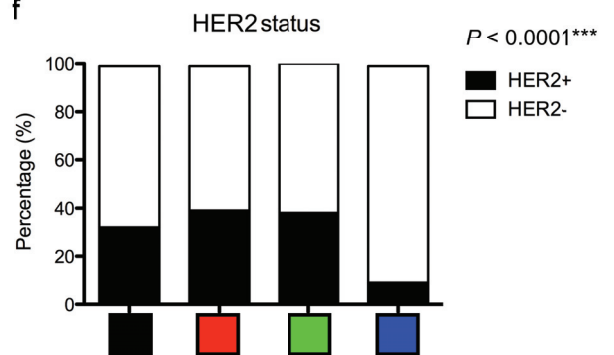
rRNA 2'OMe sites

a



b



a**b****c****d****e****f**

D Article III : Ribosome biogenesis and rRNA 2'O-ribose methylation : Uncovering specific alterations in glioma subtypes

Ce travail, dont je suis premier auteur, est en cours de rédaction et sera soumis d'ici la fin de l'année. Le manuscrit et les figures supplémentaires sont inclus après cette page introductive.

D.1 Contexte

Les gliomes sont les tumeurs primaires du système nerveux central les plus fréquentes et les plus mortelles chez l'adulte. Des études récentes ont révélé l'existence d'une hétérogénéité notable au sein de ces tumeurs, ce qui conduit à une maladie très résistante et agressive pour laquelle il existe peu de thérapies ciblées efficaces. Il est donc essentiel d'identifier de nouveaux marqueurs distinctifs des différents sous-types de gliomes et de découvrir de nouvelles cibles thérapeutiques.

Comme nous l'avons vu dans le deuxième chapitre d'introduction, il est maintenant établi que dans certaines cellules cancéreuses, la biogenèse des ribosomes est hyper-activée pour maintenir un taux de prolifération cellulaire élevé et ainsi favoriser la croissance tumorale. Cette hyper-activation est notamment utilisée comme cible thérapeutique anti-cancer. De plus, des données récentes suggèrent que des altérations de la composition du ribosome, en particulier la méthylation en 2' du ribose des ARN ribosomiques (2'Ome des ARNr), favorisent la traduction d'ARNm qui codent pour des oncogènes, démontrant l'existence d'un nouveau mécanisme impliquant la 2'Ome des ARNr pour comprendre le développement tumoral. Enfin, dans le deuxième article de cette thèse, nous avons montré que le niveau de 2'Ome des ARNr diffère entre les sous-types de cancer du sein et les grades de ces tumeurs. Par conséquent, le ribosome constitue un composant cellulaire peu exploré mais très intéressant pour comprendre la pathogénicité des cancers et découvrir des stratégies thérapeutiques innovantes.

Ainsi les objectifs de cette étude étaient d'étudier les deux facettes du ribosome dans les gliomes :

- (i) l'altération de la quantité du ribosome en analysant l'altération de la biogenèse des ribosomes
- (ii) l'altération de la qualité du ribosome en analysant l'altération de la 2'Ome des ARNr

D.2 Résultats principaux

Nous avons analysé une cohorte composée de 40 échantillons de gliomes diffus primaires de haut grade (III-IV) (13 Astrocytomes, 14 Oligodendrogliomes et 13 Glioblastomes), ainsi que 6 échantillons non-tumoraux provenant de donneurs non atteints d'un cancer. A partir de l'ARN total extrait des tissus congelés, le statut de la biogenèse des ribosomes a été évalué en quantifiant les niveaux d'ARNm de 25 facteurs clés de la biogenèse des ribosomes par la technologie de RTqPCR à moyen débit (Fluidigm). Nous avons observé une différence significative du niveau d'expression de certains facteurs de la biogenèse des ribosomes entre les différents sous-types de gliomes (Figure 1 et 2). D'une manière générale, les oligodendrogliomes et les astrocytomes surexpriment les facteurs de biogenèse des ribosomes par rapport aux glioblastomes et aux tissus non-tumoraux.

D'autre part, la composition des ribosomes a été évaluée en analysant les niveaux de 2'Ome des ARNr à l'aide de l'approche RiboMethSeq. Le séquençage a été réalisé par la plateforme de génomique des cancers du CRCL sur un séquenceur Illumina NovaSeq. Dans un premier temps, ce manuscrit montre les dernières améliorations du protocole de RiboMethSeq qui ont été réalisées au cours de ma thèse. Il s'avère que le RiboMethSeq peut être réalisé en utilisant une purification automatisée de l'ARN total à partir de tissus congelés. En effet, le RiboMethSeq est une méthode reproductible et robuste pour quantifier l'ARNr 2'Ome, même en utilisant différents protocoles d'extraction d'ARN. De plus, l'utilisation du séquençage Illumina NovaSeq a permis non seulement d'augmenter le nombre total de lectures alignées de manière unique sur des séquences d'ARNr (moyenne de 28 millions au lieu de 7 millions précédemment), mais aussi d'augmenter le nombre d'échantillons séquencés en un seul séquençage (48 échantillons pour cette étude mais théoriquement jusqu'à 96) (Figure S4A). La dernière amélioration concerne les ajustements de l'effet batch à partir des données de RiboMethSeq. Nous démontrons en effet l'efficacité de l'ajustement de l'effet batch à

l'aide de l'algorithme de ComBat-seq (Figure S4B).

Dans un second temps, le manuscrit se focalise sur l'analyse de la variation de la 2'Ome des ARNr dans les gliomes. Nos données révèlent des différences significatives entre les profils de 2'Ome des différents sous-types de gliome (Figure 3, 4 et 5). En particulier, les glioblastomes se distinguent des autres sous-types avec une diminution de la méthylation sur plusieurs sites par rapport aux oligodendriogliomes et astrocytomes. De plus, un sous-ensemble de sites d'ARNr montre des niveaux de 2'Ome statistiquement différents entre les sous-types.

En conclusion, cette étude révèle que l'hétérogénéité des gliomes est récapitulée par une altération de la quantité et de la qualité des ribosomes. Ainsi, la biogenèse et les modifications chimiques des ribosomes offrent de nouveaux biomarqueurs et des possibilités de cibles thérapeutiques pour améliorer la prise en charge clinique des patients atteints de gliomes.

D.3 Ma participation

Lors de cette étude, j'ai participé au design et à la planification expérimentale (nombre d'échantillons à séquencer, profondeur de séquençage, présence d'échantillons contrôle dans les différentes librairies de séquençage...). J'ai réalisé le traitement bio-informatique des données de séquençage de RiboMethSeq (contrôle qualité, alignement, comptage des extrémités 5'...) jusqu'au calcul des Cscores comme définis dans l'article II (Virginie MARCEL, KIELBASSA et al., 2020). J'ai suggéré et mis en place l'algorithme de ComBat-seq afin de pouvoir ajuster l'effet batch observé dans les données de RiboMethSeq. Enfin, j'ai réalisé l'analyse des données de Cscore, les tests statistiques, la mise en forme des figures et l'écriture de l'article.

D.4 Article

Glioma subtypes reveal disconnection between alterations in ribosome quantity and quality

Hermes PARAQUINDES^{1,2}, Nour-EI-Houda MOURKSI¹, Samantha BALLESTA^{1,3}, Fleur BOURDELAIS^{1,4}, Emilie THOMAS^{1,2}, Janice KIELBASSA², Jessie PERROSSIER^{1,5}, Marjorie CARRERE^{1,5}, Valéry ATTIGNON^{1,5}, Laurie TONON^{1,5}, Alain VIARI^{1,5,6}, Jean-Jacques DIAZ¹, David MEYRONET^{1,7}, Mathieu GABUT¹, François DUCRAY^{1,8}, Virginie MARCEL^{1,#,*}, Sébastien DURAND^{1,#,*}

Affiliations

1. Centre de Recherche en Cancérologie de Lyon, Inserm U1052, CNRS UMR5286, Université de Lyon, Université Claude Bernard Lyon 1, Centre Léon Bérard, CEDEX 08, F-69373 Lyon, France

2. Synergie Lyon Cancer, Gilles Thomas Bioinformatics Platform, Centre Léon Bérard, CEDEX 08, F-69373 Lyon, France

3. Organoid Platform, Centre de Recherche en Cancérologie de Lyon, CEDEX 08, F-69373 Lyon, France

4. Inovarion

5. Cancer Genomics Platform, Centre de Recherche en Cancérologie de Lyon, CEDEX 08, F-69373 Lyon, France

6. INRIA Grenoble Rhône-Alpes, 38330 Montbonnot-Saint-Martin, France

7. Hospices Civils de Lyon, Laboratoire de biologie médicale et d'anatomie pathologique

8. Hospices Civils de Lyon, Hôpital Pierre Wertheimer

Equal contribution as co-last authors

* Corresponding authors: virginie.marcel@lyon.unicancer.fr and sebastien.durand@inserm.fr

Keywords

rRNA 2'O-methylation, RiboMethSeq, ribosome biogenesis factors

Abstract (191 words)

Adult diffuse gliomas are the most common and deadly primary tumors of the central nervous system in adults. Due to their high heterogeneity, a low number of effective therapies is used in clinic and identifying new hallmarks of gliomas remains essential. The recent advances in the field of ribosome might open new perspectives. Indeed, cancer cells are addictive to hyperactivation of ribosome biogenesis and display alteration in ribosome composition affecting translational activity of ribosome. Using a cohort of 40 high grade (III-IV), primary diffuse gliomas and 6 non-tumoral samples, we show that glioma subtypes display distinct alterations in ribosome biogenesis and composition. While astrocytoma and oligodendroglioma subtypes are characterized by an over-expression of ribosome biogenesis factors compared to non-neoplastic tissues, glioblastoma multiforme (GBM) subtype displays a specific alteration of 2'O-ribose methylation levels at distinct rRNA positions. These data reveal that the heterogeneity in gliomas is recapitulated by specific alterations in either ribosome quantity or quality, indicating that alteration in ribosome quantity and quality can be disconnected in cancer. Accordingly, astrocytoma and oligodendroglioma cell lines are more sensitive to rRNA synthesis inhibitors, CX5461 and BMH21, compared to GBM. Thus, the ribosome biogenesis and chemical modifications offer novel biomarkers and therapeutic target opportunities to improve management of glioma patients in clinic.

Introduction

The ribosome is a ribonucleoprotein complex, which translates mRNAs into proteins. While ribosome has been extensively studied for a long time, its role in cancer has only recently starts to emerge (Pelletier et al, Nat Rev Cancer 2018; Truitt et al, Nat Rev Cancer 2016). In cancer cells, it indeed appears that the “quantity” but also the “quality” of ribosome can be altered compared to non-neoplastic cells providing novel promising biomarkers and therapeutic targets in clinics.

The last 10 years first revealed that during carcinogenesis, ribosome biogenesis is abnormally increased to fulfil protein needs to support the unregulated cell growth and highly proliferative rate of cancer cells. Hyperactivation of ribosome biogenesis mostly relies on the increased activity of the RNA polymerase I (RNA pol I) leading to ribosomal RNA (rRNA) synthesis in response to oncogenic activation or tumor suppressor loss (Ferreira et al, Cells 2020; Bywater et al, Nat Rev Cancer 2013). Several studies support the notion that cancer cells are addicted to ribosome biogenesis. First, hyperactivation of ribosome biogenesis is pivotal for the oncogenic activity of Cmyc in mice (Barna et al, Nature 2008). Second, inhibition of RNA pol I activity is sufficient to specifically kill cancer cells (Bywater et al, Cancer Cell 2012; Peltonen et al, Cancer Cell 14). Third, the activity of several chemotherapies such as oxaliplatin depends on ribosome biogenesis activation (Bruno et al, Nat Med 2017). Such discoveries had resulted in the development of some RNA pol I inhibitors, which showed promising response in clinical trials (Catez et al, Bioch Pharmacol 2019; Kho et al, Cancer Discov 2019). However, as usually observed for many cancer treatments, during clinical assays using Pol I inhibitors, half of the patients showed progressive disease indicating that a better patients' stratification is required. At present, it is crucial to identify novel cancer types which could benefit from these innovative therapies. Moreover, it remains key to determine right away whether ribosome biogenesis hyperactivation is a common or a specific molecular trait of cancer subtypes to avoid patients' stratification issue, by using biomarkers compatible with clinical routine.

Another decisive discovery in the ribosome field emerges in the last 10 years. For many decades, it was assumed that all ribosomes of a cell or organism display similar composition. However, it appears that ribosomes exhibit different compositions, either at ribosomal protein or rRNA levels (Genuth et al, Nat Rev Genet 2018). The current hypothesis suggests that ribosomes with distinct composition display specialized intrinsic translational activities to specify particular phenotypes (Xue et al, Nat Rev Mol Cell Biol 2012). In addition to ribosomal proteins, diversity in ribosome composition involves variation in rRNA chemical modifications catalysed by snoRNP complex, such as rRNA 2'Ome and pseudo-uridylation catalysed by snoRDs and snoRAs respectively (Jafaar et al, Cells 2021). We and others report that alteration in rRNA 2'Ome occurs in cancer and affects both translation of particular mRNA

subsets and cell proliferation (Marcel et al, Cancer Cell 2013; Eroles et al, PNAS 2017; Jansson et al, Nat Struc Mol Biol 2021; Jafaar et al, Cells 2021). In particular, we recently demonstrated by profiling 195 tumor samples using the RiboMethSeq approach, that rRNA 2'Ome profiles associate with cancer subtype and tumor grade in breast cancer (Marcel et al, NAR Cancer 2020). Similarly, alteration of rRNA 2'Ome has been reported in a small series of 17 diffuse large B-cell lymphoma (Krogh et al, NAR Cancer 2020). However, whether rRNA 2'Ome is altered in additional cancers remain to be determined. Moreover, it remains to clarify whether alterations in ribosome quality occurs independently of alterations in ribosome quantity in cancer.

Deregulation of expression of several ribosomal proteins has been reported in the most aggressive subtype of adult diffuse glioma subtype, the multiforme glioblastoma (GBM) (Hide et al, Cells 2022). Moreover, overexpression of the ribosomal protein RPS6 has been shown to promote acquisition of glioma stem cell properties, an hallmark of GBM (Shirakawa et al, Cancer Sci 2020; Gabut et al, Cells 2020; Hide et al, Cells 2022). Thus, management of patients bearing adult diffuse glioma might benefit from the recent discovery in the ribosome field. Interestingly, adult diffuse glioma are brain and spinal tumours originated from glial cells displaying heterogeneous prognosis and treatment responses. The different main subtypes (astrocytoma, oligodendroglioma and GBM) require specific therapeutic protocols. Diagnostic precision is thus crucial and currently relies on both histological and molecular criteria. In particular, the mutation status of the *isocitrate dehydrogenase (IDH)* 1 et 2 discriminates GBM (*IDH-wt*) from astrocytoma and oligodendroglioma (*IDH-mt*), the latter ones being distinguished based on the 1p19q co-deletion. Despite the recent advances in the molecular and prognostic classifications, adult diffuse gliomas are still associated with a poor survival rate that can be explained by the high rate of disease recurrence, at least in part due to the lack of alternative therapies. For instance, GBM patients treated with the conventional combination of surgery, radiotherapy and temozolomide (TMZ) chemotherapy, display a median survival of only 15 months. Identification of novel molecular mechanism deregulated in the different glioma subtypes remains crucial to improve the current therapeutic arsenal.

Here, we investigate whether alterations in ribosome quantity and quality, individually or concomitantly, are associated with the main subtypes of adult diffuse gliomas using tumor sample cohorts to the purpose of identifying novel vulnerability of adult diffuse glioma.

Materials and methods

Human grade III-IV diffuse gliomas and healthy donors' samples

A test cohort of 40 samples from adult high grade (III-IV) primary diffuse gliomas was collected from 2010 to 2016 (Hospices Civils de Lyon, Hôpital Neurologique, Neurobiotec, Lyon) and is composed of 13 glioblastoma (GBM), 13 astrocytoma (A) and 14 oligodendroglioma (O) (WHO 2017 classification). Samples were associated to clinical data, including the *IDH* mutation status, the mitotic index and survivals (Table 1). Overall survival (OS) is defined as the length of time from the date of diagnosis to the date of death or last follow-up. Progression-free survival (PFS) corresponds to the time length from the date of diagnosis to the date of relapse. Additionally, 6 healthy donor (HD) samples were provided from patients without diagnosed brain neoplastic event between 2003 to 2015 (HCL, Lyon).

A validation cohort of 11 samples, including 8 glioblastoma (GBM) and 3 healthy donors (HD), was collected at HCL (Lyon).

All patients were informed and written consent of all participants were obtained in accordance to legal and local ethical approvals.

Total RNA was extracted from frozen tissue using Maxwell or Trizol protocol as indicated by the manufacturer.

Gene expression analysis

Gene expression in healthy donors and glioma samples was evaluated by medium throughput real-time quantitative PCR (RT-qPCR) using the Biomark™ HD system (Fluidigm) as previously described (Nguyen et al, Cancers 2018). Briefly, total RNA (200 ng) was used for cDNA synthesis using the Prime Script RT Reagent kit (Takara). For each sample, two independent retro-transcription reactions were performed. Using cDNA, a PCR multiplex was first carried out to amplify gene of interest thus facilitating their detection (Supplementary Table 1). Second, after Exonuclease I treatment, qPCR reaction was performed on a 96.96 Dynamic Array™ IFC (Fluidigm) using Master Mix 2X EvaGreen (Biorad) according to manufacturer recommendations. Each cDNA was analysed in triplicate. The Fluidigm Real Time PCR Analysis software (v 4.5.2) was used for the calculation of relative fold-changes by the $2^{-\Delta\Delta CT}$ method based on a total of 6 Ct values per samples. Median Ct value of 5 housekeeping genes (*Actin*, *GAPDH*, *HPRT1*, *PGK1*, *PPIA*) was used as reference gene for normalization. Data were then normalized to Human Xpress Ref Universal Total RNA (Qiagen) as standard RNA.

RiboMethSeq

RiboMeth-seq was performed as previously described using the Illumina sequencing technology (Marchand et al, *Methods Mol Biol* 2017; Marcel et al, *NAR Cancer* 2020). Briefly, total RNA was partially and randomly fragmented under alkaline conditions. Libraries of 24 samples were prepared simultaneously using NEBNext Multiplex Small RNA Library kit. A calibrated source of total RNA (Human Xpress Ref Universal Total RNA, Qiagen) was introduced in each library. In contrast to previous studies, sequencing was performed using Illumina NovaSeq sequencer in single-end mode (SR50). Usage of this device allows to sequence up to 48 samples (equivalent to 2 libraries) per run with a median sequencing depth of 40 million of reads per sample. NovaSeq sequencing was performed by the Cancer Genomics Platform at Léon Bérard Center.

Sequencing data were processed as previously described (Marchand et al, *Methods Mol Biol* 2017; Pichot et al, *Frontiers in Genetics* 2020; Marcel et al, *NAR Cancer* 2020). FastQC was used to verify sequence quality and adapter removal was performed using Trimmomatic. The trimmed reads were aligned by Bowtie2 on the 7.2 kb-long rRNA sequence of reference (NR_046235) and the 5' read-ends were computed using bedtool genomcov utility. The C-score, which reflects the rRNA 2'Ome level, was calculated using a normalization against the median raw counts of neighbouring +/- 6 nucleotide window (Marcel et al, *NAR Cancer* 2020). The C-score varies from 0 to 1: 0 meaning that no rRNA molecule is 2'Ome; 1 that all rRNA molecules are 2'Ome; and]0 : 1[corresponding to a mix of non-2'Ome and 2'Ome rRNA molecules.

The C-scores of the 106 admitted rRNA 2'Ome sites were extracted for further analysis, either as a rRNA 2'Ome profile or as a site-by-site comparison. The most variable rRNA 2'Ome sites were identified using previously described approach based on the comparison of the interquartile range (IQR) of each site across the 40 glioma samples of the test cohort (Marcel et al, *NAR Cancer* 2020).

Batch effect adjustment of RiboMethSeq data using ComBat-Seq

Batch effect adjustment of RiboMethSeq data was performed using 5' read-end count matrix as input for ComBat-seq (Yuqing Zhang, *NAR genomics and bioinformatics* 2020). Indeed, ComBat-seq tool was developed to adjust batch effect in RNA-seq data following a negative binomial distribution and both its input and output are integer counts. ComBat-seq was used with default parameter settings. This process produces an adjusted 5' read-end count matrix then used for the C-score calculation as described above.

Statistical methods

Data and statistical analyses of both medium-throughput RT-qPCR and RiboMethSeq as well as graphical representation, were performed using R (v 4.1.2). Unsupervised data analysis was performed via hierarchical clustering and principal component analysis (PCA). Hierarchical clustering was generated via pheatmap package using Manhattan distance and Ward's linkage method. PCA was computed using ade4 package with the default parameters and visualised via factoextra package. To correlate the first 5 principal components (PCs) of the PCA on gene expression or rRNA 2'Ome profiles with the clinical features (OS, PFS, mitotic index), the eigencorplot function of PCAtools package was used. As parameters, pearson correlation coefficient and the False Discovery Rate (FDR) method for the adjusted p-values were used.

Pairwise comparison between groups was performed using non-parametric Wilcoxon rank-sum test, or Kruskal-Wallis test when more than 2 groups were identified. The FDR method was used for p-value adjustment when multiple tests were undertaken. A p-value < 0.05 was considered to be statistically significant. Significance symbols of adjusted p-value were indicated as *: p < 0.05; **: p < 0.01; ***: p < 0.001; ****: p < 0.0001. Regarding the rRNA 2'Ome, in addition to a significant p-value, a Δ C-score greater than 0.05 was considered for identifying differentially methylated rRNA 2'Ome sites, the Δ C-score corresponding to the difference between the highest and the lowest mean C-score of the groups of interest.

Availability of datasets

The generated RiboMethSeq data are available in GEO profile (GEOxxx). The datasets include the fastQC of each sample, a C-score matrix for each cohort after batch effect adjustment using ComBat-seq and a matrix summarizing the clinical data.

Results

The expression profile of ribosome biogenesis factors discriminates adult diffuse glioma subtypes

To investigate whether ribosome biogenesis (RiBi) is dysregulated in adult diffuse gliomas, we measured the expression of a set of 20 genes (referred as “RiBi-gene set”). To reflect alteration in this multistep process involving more than 200 factors (Catez et al, *Biochem Pharmacol* 2019), we selected the RiBi-gene set among the main steps of RiBi, including rRNA transcription (*NCL*, *NPM*, *POLR1A*, *TAF1A*, *TAF1B*, *TAF1C* and *UBTF*), rRNA maturation (*BOP1*, *PES1* and *WDR12*) and snoRNA biogenesis (*RUVBL1*, *PIDH1D1* and *RUVBL2*), or composing H/ACA (*DKC1*, *GAR1*, *NHP2* and *NOP10*) and C/D box snoRNPs (*NOP56*, *SNU13* and *FBL*) (Figure 1A). Of note, several selected genes (*PIH1D1*, *RUVBL2* and *FBL*) are located on the long arm of the chromosome 19 (19q), which undergoes a heterozygous deletion in oligodendrogliomas. A readout of RiBi steady state was determined by quantifying mRNA expression levels of these selected genes by medium throughput RT-qPCR in tissue samples from glioblastoma (GBM, n=13), astrocytoma (A, n=13), oligodendroglioma (O, n=14) and cerebral cortex of healthy donors (HD, n=6) (Table 1). mRNA levels were normalized to the median mRNA expression of 5 housekeeping genes, which expectedly did not significantly vary among the four groups (Figure S1A).

To examine the association between expression levels of RiBi factors and glioma subtypes, we first performed an unsupervised approach using a principal component analysis (PCA) based on the RiBi-gene set profile (Figure 1B). Interestingly, three main clusters were observed. A first large cluster composed of both healthy donors and glioblastoma (HD/GBM cluster) was distinct from two other clusters, corresponding to oligodendrogliomas (O cluster) and astrocytomas (A cluster). In particular, the principal component 1 (PC1, variance: 67.8%) split HD/GBM cluster from A/O clusters, while the PC2 (variance: 9%) distinguished the O cluster from others. These data suggest that the expression profile of only 20 RiBi factors allows a clear distinction between the three main glioma subtypes.

To gain more insight into connections between the RiBi-gene set expression profile and the molecular and clinical characteristics of gliomas, we first performed Pearson correlations between the first 5 PCs and the overall survival (OS) and progression-free survival (PFS) (Figure 1C and Supplementary Figure S2A). We observed a significant association only for PC1 and PC2 axes, indicating that the clustering based on the RiBi-gene set profiles provided by these two axes is sufficient to explain all clinical data of interest. OS and PFS were significantly correlated with both PC1 and PC2 in a negative manner, suggesting that samples

clustering at the right part of the PC1 and PC2 graph display lower OS and PFS. This is consistent with glioblastoma patients (located at right of PC1) exhibiting the poorest OS (less than 30 months) and PFS (under 30 months), followed by astrocytoma patients (located left of PC1 but right of PC2) and oligodendroglioma patients (left of both PC1 and PC2). Concerning the mitotic index, it was slightly correlated with the PC1 only, albeit significantly, suggesting that, while the clustering based on the RiBi-gene set can explain differences in mitotic index between GBM and other gliomas, it failed to define mitotic index differences between oligodendrogliomas and others (Figure 1C and Supplementary Figure S2A). This observation is consistent with the fact that glioblastoma located at right of PC1 are known to be highly proliferative compared to both astrocytoma and oligodendroglioma. Thus, the expression profiling of the the RiBi-gene set allows to distinguish glioma samples with distinct survivals and to a lesser extent with different mitotic index.

We then evaluated the correlation between the first 2 PC axes and the *IDH1/2* mutation or 1p/19q co-deletion statuses (Supplementary Figure S2B-C). Interestingly, the PC1 (HD/GBM vs A/O clusters) but not PC2 axis significantly correlates with the *IDH1/2* mutation status, while PC2 (A/GBM vs O clusters) but not PC1 succeeded in segregating glioma tumours based on the 1p/19q codeletion status. These observations support the notion that the RiBi-gene set profiling distinguishes glioma samples with distinct genomic alterations. Altogether, our results suggest that the ribosome biogenesis pathway is differentially dysregulated in glioma and that the profiling of only 20 genes involved in ribosome biogenesis is sufficient to discriminate the three main subtypes of glioma.

Ribosome biogenesis factors are highly expressed in astrocytomas and oligodendrogliomas

To better describe the dysregulation of the ribosome biogenesis pathway occurring in glioma subtypes, we compared gene expression levels of the 20 ribosome biogenesis factors individually in healthy donors and the three glioma subtypes (Figure 2 and Supplementary Figure S1B)[HP1]. A significant difference was observed in the expression of all tested RiBi factors between the four groups (Kruskal-Wallis test, Supplementary Figure S1B). Quite surprisingly, no significant difference in mRNA expression was observed in glioblastoma compared to non-neoplastic samples, except for *GAR1* and *RUVBL1* (Figure S1B, lane "HD vs. GBM"). [HP2] However, glioma samples carrying *IDH1/2* mutations (*i.e.*, astrocytoma and oligodendroglioma) significantly overexpressed most of these genes compared to wild-type *IDH* samples (*i.e.*, healthy donors and glioblastoma). For instance, expression of *NCL*, which encodes for a key factor in rRNA synthesis, is increased 2- and 3-fold in astrocytoma and

oligodendroglioma, respectively, compared to healthy donors (Figure 2A). In contrast, *NCL* expression levels in glioblastoma samples were similar to those of non-neoplastic tissues. Likewise, the use of a second distinct set of primers to analyse *NCL* expression provided the exact same trend (Supplementary Figure S1C), confirming our observations do not arise from technical caveat. Therefore, the expression levels of factors involved in ribosome biogenesis are higher in *IDH1/2* mutated astrocytoma and oligodendroglioma samples than in glioblastoma and non-neoplastic samples.

Interestingly, genes located on chromosomes 19q (*FBL*, *PIH1D1*, *RUVBL2*) displayed a particular expression pattern (Figures 2C, 2E and Supplementary Figure S1B). Like other genes involved in ribosome biogenesis, *FBL*, *PIH1D1* and *RUVBL2* are significantly overexpressed in astrocytoma compared to both glioblastoma and non-neoplastic samples. However, mRNA expression levels in oligodendroglioma were lower than in astrocytoma and displayed expression profiles rather resembling those observed in glioblastoma and non-neoplastic samples. Thus, *FBL* was highly expressed exclusively in astrocytoma while *FBL* expression levels were equivalent in oligodendroglioma and glioblastoma, as confirmed by a second set of primers (Supplementary Figure S1D). Considering that oligodendroglioma samples exhibit a heterozygous deletion of both chromosomes 1p and 19q (Table 1), the latter carrying *FBL*, *PIH1D1* and *RUVBL2* genes, these data indicate that the specific expression profile observed for these genes in oligodendrogliomas is rather due to an haploinsufficiency caused by genetic alterations than a regulation of expression similar to that of glioblastomas and non-neoplastic samples. Altogether, these data clearly demonstrate that, compared to non-neoplastic tissues, the ribosome biogenesis pathway is enhanced in *IDH1/2* mutant gliomas (*i.e.*, astrocytoma and oligodendroglioma) but remains relatively unaffected in *IDH* wild-type glioblastoma.

RiboMethSeq optimisations for analysing large scale series of human samples

Considering the difference in levels of the rRNA 2'O-methyl-transferase *FBL* and others components of the C/D snoRNP complex among the three main adult diffuse glioma subtypes, we investigated whether alteration in rRNA 2'Ome differentially occurs in glioma. We quantified rRNA 2'Ome level using the recently described RiboMethSeq approach to simultaneously measure 2'Ome levels on the 106 known sites. To first optimize RiboMethSeq for large-scale study of human samples, we used a small cohort (n=11) to compare the impact of manual (Trizol protocol) and automatized (Maxwell kit) purification methods of total RNA on RiboMethSeq data. RNA Integrity Number and RNA-seq read numbers did not show any significant differences between both purification methods (data not shown). We then compared

the C-scores, which reflects rRNA 2'Ome levels at specific positions, using an unsupervised clustering approach (Figure S3A) and a two-by-two Pearson correlation (Figure S3B) and uncovered a strong correlation between the C-scores among the same sample, regardless of the method for RNA purification. These data demonstrate that RiboMethSeq can be performed using automatized purification of total RNA, and that RiboMethSeq is a reproducible and robust method to quantify rRNA 2'Ome even when using different RNA extraction protocols.

Up to date, RiboMethSeq was mostly performed using HiSeq Illumina device. However, it has main limitations for analysing large-scale series of human samples. Indeed, increasing the number of samples per run to reduce putative batch effect decreases the sequencing depth and 5' read-end coverage across the rRNA as well. Therefore, as the C-score corresponds to the ratio of the 5' read-end counts at a position to the local 5' read-end count coverage, increasing the number of samples per run ultimately decreases the reproducibility and precision of C-score and the robustness of rRNA 2'Ome level quantification. To overcome these limitations, we used the NovaSeq Illumina sequencing utility and evaluated Combat-Seq as a tool to adjust batch effect (Supplementary Figure S4).

To established the rRNA 2'Ome profile of our glioma cohort, we randomly segregated RNA samples and prepared two independent library batches of 23 samples (40 gliomas and 6 healthy donors), each batch also containing a commercially-available "reference" total RNA that we systematically included in each RiboMethSeq library. All samples were sequenced together on a single run using NovaSeq device therefore resulting in a high level of read per samples (Supplementary Figure S4A). The median of the total number of sequenced reads reached 40 million, with about 28 million of reads uniquely aligning to the rRNA sequence or ference, which is 2-fold higher than the minimal number of reads required to determine stable C-score estimation. As a comparison, usage of HiSeq device in a previous study allowed to reach a median of total number of reads of 7.2 million (Marcel at al, NAR Cancer 2020). Thus, the utilization of NovaSeq platform not only permitted to increase the total number of reads that uniquely aligned to rRNA sequences but also allowed to increase the number of samples sequenced in a single run.

Unsupervised analysis of the entire cohort using PCA based on C-scores of either all rRNA positions (7055 sites) or the 106 positions corresponding to known rRNA 2'Ome sites allowed to clearly distinguished samples depending on the library origin, suggesting a batch effect between the two libraries (Supplementary Figure S4B, left panels). Since no study ever addressed batch effects in RiboMethSeq approach, we evaluated adjustment of RiboMethSeq data using ComBat-seq algorithm, one of the most routinely used tool for adjustment of RNA-seq data (Yuqing Zhang, NAR genomics and bioinformatics, 2020). ComBat-seq indeed

uses a Bayesian linear regression model to adjust the batch effect in count data following a negative binomial distribution. ComBat-seq was applied to the 5' read-end counts and the adjusted 5' read-end counts were used to re-calculate C-scores. In contrast to PCA on C-scores issued from unadjusted data, no distinction between samples based on their library origin was observed in PCA on C-scores issued from adjusted data (Supplementary Figures S4B right panels). More importantly, the reference total RNA included in the two libraries almost completely overlaps based on the comparison of the CombatSeq adjusted C-scores of the 106 rRNA 2'O-methylated sites (Supplementary Figure S4B bottom right panel). These data demonstrate the efficiency of the batch effect removal using ComBat-seq algorithm on RiboMethSeq data. Overall, automatized purification of total RNA from human tissues, usage of NovaSeq Illumina device and raw data adjustment using ComBat-seq appear as useful tools for rRNA 2'Ome profiling of large-scale human series using the RiboMethSeq approach.

rRNA 2'Ome profiles discriminate glioblastomas from astrocytomas and oligodendrogliomas

To now gain insight into alterations of rRNA 2'Ome occurring in adult diffuse gliomas, we analyzed in more details rRNA 2'Ome in our cohort by comparing C-scores obtained using the RiboMethSeq approach implemented with optimizations described above (automated RNA purification, NovaSeq sequencing device and Combat-seq adjustment). Using unsupervised hierarchical clustering analysis (HCA), we first evaluated rRNA 2'Ome levels at the 106 known sites in the 46 non-neoplastic and glioma samples (Figure 3). While a vast majority of sites show a C-score close to 1 in a single sample, indicating that all rRNA molecules of the sample are 2'O-methylated at this specific position, few sites displayed in contrast a C-score below 0.9, indicating that a mix of 2'O-unmethylated and 2'O-methylated rRNA molecules are present in the same sample. Similar results were observed in recent studies in human samples from diffuse large B cell lymphoma and breast cancer (Krogh et al, NAR Cancer, 2020; Marcel et al, NAR Cancer, 2020). These data further confirmed observations by us and others that some rRNA molecules can exist without 2'Ome at some particular sites. In addition, these observations sustain the emerging notion that rRNA 2'Ome is altered in cancers, including now in adult diffuse gliomas.

Interestingly, in the sample dendrogram of the HCA, all glioblastoma samples form a branch completely separated from other samples (left), suggesting that the rRNA 2'Ome profile clearly distinguished glioblastoma samples from astrocytoma, oligodendroglioma and non-neoplastic tissues (Figure 3). PCA analyses indicated that PC2 but not PC1 axis unambiguously separates glioblastoma samples from the other groups (Supplementary Figure 5A). As

expected, while PC1 was not associated with a significant correlation to any of the tested clinical features, PC2 significantly correlated with the *IDH1/2* mutation status, the mitotic index and the OS/PFS (Supplementary Figure S5B-D). To validate these observations, we analyzed rRNA 2'Ome levels using RiboMethSeq on an independent validation cohort composed of 8 glioblastomas and 3 non-neoplastic tissues. Once again, unsupervised PCA based on the rRNA 2'Ome profile showed that PC1 clearly separates glioblastomas from non-neoplastic tissues (Supplementary figure S5E). Altogether, our data revealed that rRNA 2'Ome levels varies in glioma and that rRNA 2'Ome profiles discriminate glioblastomas not only from other glioma subtypes but also from healthy donor tissues.

Glioblastomas display a site-specific decrease in rRNA 2'Ome level

To better characterize differences in rRNA 2'Ome profiles between glioma subtypes, we performed a site-by-site analysis. At first, we focused on the most variable rRNA 2'Ome sites among gliomas. To identify such sites, we compared the variability of C-scores at each single site among all glioma samples (n=40) using the inter-quartile range (IQR) value as previously described (Marcel et al, NAR Cancer 2020). A set of 35 sites showed a clear discontinuity in the linear increase of IQR values and were considered as the most variable rRNA 2'Ome sites among gliomas (Figure 4A), in agreement with our recent findings that only a subset of rRNA sites (about one third) displays plasticity in 2'Ome levels (Marcel et al, NAR Cancer 2020). Interestingly, PCA analyses using the rRNA 2'Ome profile based on this 35-site set is sufficient to discriminate glioblastomas from astrocytoma and oligodendroglioma (Figure 4B). Actually, PCA analyses based on the 35-site set could discriminate glioblastomas from others on PC1 unlike PCA analyses based on rRNA 2'Ome of the 106 known sites. These data further emphasize the specificity of glioblastoma rRNA 2'Ome profiles compared to other glioma subtypes and suggest that glioblastomas display strong alterations of rRNA 2'Ome levels at some specific sites.

To identify rRNA sites whose 2'Ome level is significantly affected between glioma subtypes and non-neoplastic tissues, we performed systematic analysis for each of the 106 rRNA 2'Ome sites using Kruskal-Wallis test with adjusted p-values. From this screening, only 16 rRNA 2'Ome sites displayed a significant variation, *i.e.*, Δ C-scores >0.05 , in at least one of the glioma subtypes or healthy donors (Figure 5A). Of note, 15 of these 16 sites correspond to most variable sites, further supporting a site-specific alteration of rRNA 2'Ome level in gliomas. Pairwise comparisons revealed that rRNA 2'Ome levels were differentially affected among glioma subtypes (Figure 5). First, almost no significant differences in 2'Ome level was observed when comparing non-neoplastic tissues to either astrocytomas or

oligodendrogliomas. Second, rRNA 2'Ome levels appeared to be significantly increased in 6 sites in astrocytomas compared to oligodendroglioma that can be explained, at least in part, by locations of C/D box snoRNAs targeting these sites in the chromosome 1p, which undergoes a heterozygous deletion in oligodendrogliomas. Finally, the main difference relied on glioblastomas that displayed an overall lower rRNA 2'Ome levels at some specific sites (Figure 5). Indeed, rRNA 2'Ome levels were significantly decreased in 8 out of the 16 analyzed sites in glioblastomas compared to astrocytoma, oligodendroglioma and healthy donors, and reached 11 out of 15 sites when comparing glioblastomas and astrocytomas (Figure 5B). In contrast, 1 site (18S-Am576), 2 sites (28S-Gm4593, 28S-Cm2848) and 3 sites (18S-Cm174, 28S-Gm4593 and 28S-Am4560) were significantly more 2'O-methylated in glioblastomas compared respectively to others or to astrocytoma or to both to healthy donors and oligodendroglioma (Figure 5B). The same overall decrease in rRNA 2'Ome level was observed in glioblastoma compared to non-neoplastic tissues using the second validation cohort (Supplementary Figure S6).

Taken altogether, our data showed the rRNA 2'Ome on several specific sites is altered in gliomas. Interestingly, rRNA 2'Ome of these sites is mostly decreased in glioblastomas compared to other glioma subtypes and non-neoplastic tissues.

Discussion

Adult diffuse glioma are heterogeneous tumours associated with distinct, but still poor survival rates due to the lack of effective targeted therapies, in particular for the most aggressive subtype, the multiform glioblastoma (GBM). Identifying novel molecular vulnerabilities is thus crucial to improve glioma patient management. Here, we report that glioma subtypes display distinct alterations in ribosome. Moreover, our data support for the first time the notion that alterations in ribosome quantity can be disconnected from alteration in ribosome quality, at least in glioma tumours.

Alteration in rRNA 2'Ome profiles has been reported in numerous cellular models and only in two cancer types, breast cancer and diffuse large-B cell lymphoma (DLBLL) (Jafaar et al, Cells 2021; Marcel et al, NAR Cancer 2020; Krogh et al, NAR Cancer 2020). By optimizing the RiboMethSeq technology, we show that rRNA 2'Ome profile also vary in adult diffuse glioma. As demonstrated previously in breast cancer and DLBLL, we observe that only a third of the known rRNA 2'Ome sites display variability in their 2'Ome levels in glioma. However, among the rRNA sites that can tolerate absence of 2'Ome, only few are common to all cancer types, most of them being cancer-specific. Due to the difference in RiboMethSeq protocols (sequencing platform, batch effect, C-score computation, identification of variable rRNA sites...), rigorous methodologies will be required to perform meta-analysis on the available RiboMethSeq datasets to clearly address this central question (Krogh et al, Methods 2019; Pichot et al, Front Genet 2020; Marcel et al, NAR Cancer 2020). Interestingly, by comparing glioma subtypes, we observed that rRNA 2'Ome profiles is different in GBM compared to astrocytoma, oligodendroglioma and non-neoplastic tissues, supporting the notion that rRNA 2'Ome reflects the inter-tumor variability (Marcel et al, NAR Cancer 2020). At least 6 rRNA sites can distinguish GBM from other glioma subtypes based on their 2'Ome levels. Among them, 2 were specifically altered in glioma but not in breast cancer and DLBLL (28S_Cm1227, 28S_Cm1868), the others having also been associated with distinct breast cancer subtypes. In particular, similar increase in rRNA 2'Ome levels at 18S_Am576 and decrease at 18S_Gm1447 were observed in the most aggressive subtypes of glioma (GBM) and of breast cancer (triple negative or TNBC), suggesting a link between these sites and cancer aggressiveness (Marcel et al, NAR Cancer 2020). Whether these rRNA 2'Ome sites contributes in the acquisition of cancer cell characteristics remains to be deciphered. Recent study indeed demonstrates that alteration of 2'Ome level at a single rRNA site is sufficient to affect cell proliferation, one hallmark of cancer aggressivity (Jansson et al, Nat Struc Mol Biol 2021).

In DLBLL, the global decrease in rRNA 2'Ome profile was directly associated with Ki67-estimated high proliferative rate of tumours, suggesting that decrease in rRNA 2'Ome level results from a side-effect of the hyperproliferative rate of cancer cells associated with increase in rRNA synthesis, the component of the rRNA 2'Ome machinery becoming the limiting factor in cancer cells (Krogh et al, NAR Cancer 2020). Similar association was observed in glioma, global decrease in rRNA 2'Ome level being associated with mitotic index. However, several observations in our study suggest that the ratio rRNA:rRNA 2'Ome machinery is not sufficient to explain alteration in rRNA 2'Ome profile in glioma. First, astrocytoma and oligodendroglioma over-expressed ribosome biogenesis factors, used here as a surrogate of rRNA synthesis, compared to non-neoplastic and GBM tissues. Second, GBM express as much as the rRNA 2'O-methyltransferase mRNA levels as oligodendroglioma. Third, only GBM display global decrease in rRNA 2'Ome at some sites compare to non-neoplastic samples. Finally, levels of rRNA 2'Ome increase in GBM compared to other glioma subtypes at some sites. Thus, additional molecular mechanisms should be explored to identify the origin of rRNA 2'Ome alteration in cancer, and understand the rRNA site- and cancer-specificity observed. In particular, combined analysis with expression of C/D snoRNAs which guides FBL on the rRNA sites would be of high interest. The hypothesis that decrease in rRNA 2'Ome levels mainly results from a passive effect, implies that alteration in ribosome quantity is intimately connected to alteration in ribosome quality. Our findings that the main glioma subtypes are associated with alteration in either ribosome quantity and quality challenge this vision.

The specificity of alterations regarding ribosome opens novel perspectives in clinics. Our data reveal that *IDH-mt* glioma including both astrocytoma and oligodendroglioma display the highest expression of ribosome biogenesis factors, suggesting an increase in ribosome biogenesis. Whether *IDH-mt* is directly involved in the over-expression of ribosome biogenesis factors remain to be deciphered, as well as whether *IDH-mt* effect is limited to the RNA pol II activity or affects the activity of the three RNA pol to finely coordinate ribosome biogenesis. At present, no literature has linked *IDH* mutation status to ribosome biogenesis. In clinics, glioma subtype-specific alteration in ribosome might first allow identification of additional biomarkers to refine glioma subtypes. Although we restricted our analysis to mRNA levels to compare ribosome quantity/quality, most of the ribosome biogenesis factors analysed could be easily transferred to immunohistochemistry, as shown for *FBL* or *NCL* for example (Nguyen et al, BMC Cancer 2021; Mourksi et al, in revision). Moreover, the specificity of alteration in ribosome biogenesis in astrocytoma and oligodendroglioma compared to GBM, as well as non-tumoral tissue, suggest that such tumours might be sensitive to the newly develop RNA pol I inhibitors, whose were successfully evaluated in clinical trial (Catez et al, Bioch Pharm 2019).

Altogether, our data indicate that alterations in ribosome quantity and quality are disconnected in glioma. Thus, each subtype would benefit from the recent discoveries in the ribosome field, either as novel biomarker or as therapeutic target.

Conflict of interest

The authors have no conflicts of interest to declare.

Ethical Approval

All the experiment protocol for involving human was in accordance with the guidelines of national and institutional entities based on Declaration of Helsinki in the manuscript. All experimental protocols were approved by an independent French Ethic Committee (CPP Sud Est IV). Written informed consent was obtained from all patients at their inclusion.

Fundings

This work was supported by INCa (PLBio 2019-138 MARACAS), the French association ARTC (n°), the Cancéropôle Lyon Auvergne Rhône-Alpes (AAP international 2021 MARACAS.v3.0) and the SIRIC program (INCa-DGOS-Inserm_12563, LyRICAN). HP and NEHM were recipients of PhD fellowships from Ligue Nationale Contre le Cancer.

Author contribution

HP, NEHM, SB, FB, JP, MC, VA, LT performed and analyzed experiments. HP, NEHM, EM, JK performed statistical description. DM and FD provide human samples and clinical data. HP, NEHM, FC, VM and SD interpreted the data. VM and SD shaped the clinical and research question and supervised the project coordination. AV, JJD, MG, FC, VM and SD provide financial supports. HP, NEHM, VM and SD wrote the first draft of the manuscript. All authors have read and approved the manuscript.

Acknowledgements

We would like to thank all the people involved in this study, including the patients and their families. We thank platform' staffs that have not been referred as co-authors: CRB platform (HCL), C3D platform (CRCL), Cancer genomics platform (CRCL), Gilles Thomas Bioinformatic platform (CRCL).

References

Table 1. Clinical characteristics of patients with high grade primary diffuse glioma

	Glioblastom a (n=13)	Astrocytoma (n=13)	Oligodendroglioma (n=14)
Age (years)	60 [52-73]	34 [25-52]	53 [35-66]
Gender			
Female	4 (31%)	4 (31%)	5 (36%)
Male	9 (69%)	9 (69%)	9 (64%)
Tumour location			
Temporal	3 (23.1%)	3 (23.1%)	2 (14.3%)
Frontal	7 (53.8 %)	10 (76.9%)	12 (85.7%)
Orbital	1 (7.7%)	0 (0%)	0 (0%)
Parietal	2 (15.4%)	0 (0%)	0 (0%)
Genetic alteration			
mutation IDH1/2	0 (0%)	13 (100%)	14 (100%)
del 1p/19q	0 (0%)	0 (0%)	14 (100%)
ATRX-	NA	12 (92.3%)	0 (0%)
Progression			
Yes	12 (92.3%)	5 (38.5%)	3 (21.4%)
No	0 (0%)	8 (61.5%)	11 (78.6%)
NA	1 (7.7%)	0 (0%)	0 (0%)

Figure Legends

Figure 1. Expression profile of ribosome biogenesis factors distinguishes main glioma subtypes. (A) Panel of analyzed genes involved in ribosome biogenesis. Expression levels of 20 genes involved in rRNA transcription, rRNA maturation, snoRNA biogenesis and coding for H/ACA or C/D boxes snoRNAs associated proteins have been analysed using medium throughput RT-qPCR. Genes located on short arm of chromosome 1 (1p) or long arm of chromosome 19 (19q) are indicated. (B) Expression of factors involved in ribosome biogenesis enables gliomas subtype clustering. Principal component analysis (PCA) based on the expression profile of the 20 genes involved in ribosome biogenesis. Each dot represents a sample. Ellipsis shows 90% confidence interval around the centroid of each group. Percentage of variance explained by PC1 and PC2 are indicated. PC1 dissociates healthy donors (HD) and glioblastoma (GBM) from astrocytoma (A) and oligodendroglioma (O). (C) Expression profile of factors involved in ribosome biogenesis associates with main clinical features. Correlation heatmap of PC1 to PC5 axes with patient clinical data. OS, PFS and mitotic index correlate with PC1 and PC2. Coefficients of correlation (R) are indicated and highlighted by different colours, blue when positive and red when negative. Significant correlations are indicated with asterisk: **p < 0.01; ***p < 0.001. OS, overall survival; PFS, progression-free survival.

Figure 2. Astrocytoma and oligodendroglioma display increased expression of ribosome biogenesis factors compared to glioblastoma. Medium throughput RT-qPCR analysis of genes expression in glioma and healthy donor samples was performed. Expression level of genes coding for factors involved in rRNA transcription (A), rRNA maturation (B), snoRNAs biogenesis (C), H/ACA box snoRNAs associated proteins (D) and C/D box snoRNAs associated proteins (E) were normalized to the median expression value of 5 genes of reference (*Actin*, *GAPDH*, *HPRT1*, *PGK1*, *PPIA*) and to a standard RNA. Factors involved in ribosome biogenesis are overexpressed in mutant IDH gliomas (i.e., A and O) compared to both wild-type IDH glioma (GBM) and non-neoplastic tissues (HD). *FBL*, *PIH1D1* and *RUVBL2* are located on chromosomes 1p or 19q.

Figure 3. rRNA 2'Ome level varies in glioma. Unsupervised hierarchical clustering of C-score at 106 sites (rows) was performed in the cohort of 40 high grade (III-IV) glioma and 6 healthy donors (columns). C-score is represented by colour scale from 0 to 1 (colour scale panel) and sample origin is given (top colour panel). Mean C-score for each site among the 46 samples is shown at the right panel. rRNA 2'Ome level exhibit high variability in glioma, among

a single sample (i.e., intra-variability) but also among the cohort (i.e., inter-variability). 86 sites have a mean C-score higher than 0.9 (black) and 20 sites lower than 0.9 (red). Glioblastoma samples are distinguished from the other sample types based on their rRNA 2'Ome profiles (sample dendrogram).

Figure 4. The most variable rRNA 2'Ome sites are sufficient to distinguish the different glioma subtypes. (A) Identification of the most variable rRNA 2'Ome sites in glioma. The C-score variability was measured among the glioma samples using the interquartile range (IQR) for each site. After ranking rRNA 2'Ome sites by increasing IQR value, sites whose IQR does not increase linearly were assigned as "the most variable sites". A group of 35 sites exhibits the most variable rRNA 2'Ome level in glioma (red box). **(B)** The rRNA 2'Ome profiles at the most variable sites set apart the different glioma subtypes. Unsupervised Principal Component Analysis (PCA) based on the C-scores of the 35 most variable sites was performed on glioma samples. Glioblastoma (GBM) have a distinct rRNA 2'Ome profile at the 35 sites from astrocytoma (A) and oligodendroglioma (O). Percentage of variance explained by PC1 and PC2 are indicated. Samples are coloured according to their subtype and the 95 % confidence ellipsoids around the centroid of each group are drawn.

Figure 5. Glioblastoma display lower rRNA 2'Ome level than other glioma subtypes. (A) Identification of significantly deregulated rRNA 2'Ome sites in glioma. rRNA 2'Ome sites whose level shows both significant difference among glioma and healthy donors (Kruskall-Wallis, adjusted $p < 0.05$) and an absolute difference of mean C-score between the highest and lowest 2'Ome groups above $|0.5|$ ($\Delta C\text{-score} > |0.5|$), were considered as significantly deregulated in the cohort. Boxplots show the distribution of C-score for each glioma subtype (A, O and GBM) and healthy donor (HD) in the 16 deregulated rRNA 2'Ome sites. Adjusted p-value: bottom-left of each panel; colour: sample origin. **(B)** Specific alteration of rRNA 2'Ome level on glioma subtypes at a restricted number of sites. Pairwise comparison between the different groups for the 16 significantly deregulated rRNA 2'Ome sites. Compared to astrocytoma (A), oligodendroglioma (O) and healthy donors (HD), most rRNA sites have a lower level of 2'Ome in glioblastoma. rRNA sites, whose 2'Ome is dependent upon C/D boxe snoRNAs located in the chromosome 1p, are indicated (SNORD). Color: $\Delta C\text{-score}$ difference (UP: $\Delta C\text{-score} > 0$; DOWN: $\Delta C\text{-score} < 0$); number: adjusted p-values with *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$.

Supplementary Table 1: RT-qPCR primers sequence

Genes	Forward primer	Reverse primer
<i>Actin</i>	5'-CCAACCGCGAGAAGATGA-3'	5'-TCCATCACGATGCCAGTG-3'
<i>BOP1</i>	5'-GACGATCCTGACTACTGGCG-3'	5'-ACCTGCTCATCCGTCAGTCT-3'
<i>DKC1</i>	5'-CCCTTTGGAAAAGCTGTTGA-3'	5'-TAATCTTGGCCCCATAGCAG-3'
<i>FBL #1</i>	5'-CCTGGGGAATCAGTTTATGG-3'	5'-CCAGGCTCGGTACTCAATTTT-3'
<i>FBL #2</i>	5'-CCTGCGTAATGGAGGACACT-3'	5'-GCTGAGGCTGTGGAGTCAAT-3'
<i>GAPDH</i>	5'-AGCCACATCGCTCAGACAC-3'	5'-GCCCAATACGACCAAATCC-3'
<i>GAR1</i>	5'-CAAGACCAAGGACCTCCAGA-3'	5'-TCATCTTCACAGGGATGCAG-3'
<i>HPRT1</i>	5'-TGACACTGGCAAAACAATGCA-3'	5'-GGTCCTTTTCACCAGCAAGCT-3'
<i>NCL #1</i>	5'-CCAGAACCAAAATGGCAAAT-3'	5'-CTGATTGCTCTGCCCTCAAT-3'
<i>NCL #2</i>	5'-GTCAGCAAGGATGGGAAAAG-3'	5'-TAGATCGCCCATCGATCTCT-3'
<i>NHP2</i>	5'-GGTCAACCAGAACCCCATC-3'	5'-TTCTGCTTCACCGCTTTCTT-3'
<i>NOP10</i>	5'-GAAGAAATTTGACCCGATGG-3'	5'-TGAAGCGTTTCTTGATGGTG-3'
<i>NOP56</i>	5'-AGGCTATTCTGGATGCCTCA-3'	5'-GTAGGCTCTGGCGGTATTCA-3'
<i>NPM</i>	5'-TTGTTGAAGCAGAGGCAATG-3'	5'-TATTTCAAAGCCCCCAAGG-3'
<i>PES1</i>	5'-GGCAAGAGGCGAAAAATCCG-3'	5'-TTCTTCTCAGACCTCACCGC-3'
<i>PGK1</i>	5'-AAGTGAAGCTCGGAAAGCTTCTAT-3'	5'-AGGGAAAAGATGCTTCTGGG-3'
<i>PIH1D1</i>	5'-TGAGTCTGGGAGAGCCTCAT-3'	5'-AAATCGCTGTTCTGCATCCT-3'
<i>POLR1A</i>	5'-CTGAGCCCCTGGGAATTGAG-3'	5'-CCTTCATTCTTCCACAGGGCA-3'
<i>PPIA</i>	5'-GTCAACCCCACCGTGTTCTT-3'	5'-CTGCTGTCTTTGGGACCTTGT-3'
<i>RUVBL1</i>	5'-AAGGGGATGTGC ACA AAA AG	5'-CACATCCAAGTCATGCAAGG-3'
<i>RUVBL2</i>	5'-CGCTCTTCTCAGGTGACACA-3'	5'-CAGCACTCCAGGGATGATCT-3'
<i>SNU13</i>	5'-GCTACTGGACCTCGTTCAGC-3'	5'-ACTCAGAGATGCCCTGTTG-3'
<i>TAF1A</i>	5'-TCCTGGAGTTTGGGACCCTT-3'	5'-TGGTGAGTACCTCTTGGGCT-3'
<i>TAF1B</i>	5'-CGAGGAGGCGGAAGAGTTTA-3'	5'-TCTCTCTGTAACATTGTGGCAAGA-3'
<i>TAF1C</i>	5'-CGGAGTGAAGATGCTGGACA-3'	5'-GCCCCCAAACGAAAAAGCAA-3'
<i>UBTF</i>	5'-TGTGGAACGACCTGTCTGAG-3'	5'-CTCTCCGACTGAGCCTTGAG-3'
<i>WDR12</i>	5'-TAAAGGGGCAGAGGAATGGAT-3'	5'-CAACATCCGTATGTCCCACAA-3'

Supplementary Figure Legends

Figure S1. RiBi gene expression quantification by medium throughput RT-qPCR strategy revealed disparity within glioma subtype. (A) Validation of housekeeping genes stability among tissues. Median Ct value of 5 reference genes (*Actin*, *GAPDH*, *HPRT1*, *PGK1*, *PPIA*) among each sample and standard RNA obtained by medium throughput RT-qPCR (Biomark HD, Fluidigm). Median Ct value remains constant within studied samples allowing RT-qPCR data normalization. **(B)** Comparison of RiBi gene expression in diffuse adult gliomas and healthy donors. Adjusted p-value heat map corresponding to Kruskal-Wallis or Wilcoxon Mann Whitney statistical test of expression level of 20 genes involved in ribosome biogenesis within the different sample groups. Compared to healthy donors (HD) and glioblastoma (GBM), astrocytoma (A) and oligodendroglioma (O) exhibit overall similar RiBi expression pattern. **(C-D)** Validation of *NCL* and *FBL* expression pattern in glioma and healthy donors. *NCL* (C) and *FBL* (D) mRNA levels was assessed using two different set of primers (#1 and #2) in healthy donors (HD), glioblastoma (GBM), astrocytoma (A) and oligodendroglioma (O) by medium throughput RT-qPCR. *NCL* and *FBL* are differentially expressed in gliomas.

Figure S2. Glioma clustering based on the expression profile of 20 genes involved in ribosome biogenesis correlates with patient clinical data. (A) PC1 and PC2 of Figure 1B are associated with glioma clinical data. Correlation scatter plots of PC1 (left panel) or PC2 (right panel) and patient clinical data (OS, overall survival; PFS, progression-free survival; mitotic index). Each dot corresponds to a sample: pink, glioblastoma (GBM); green, astrocytoma (A) and violet, oligodendroglioma (O). Pearson correlation coefficient (R) is indicated. Significant negative correlation was found between both PC1 and PC2 and clinical data of OS and PFS. **(B-C)** Expression profile of 20 genes involved in ribosome biogenesis allows distinction of *IDH*-WT and -MT gliomas or 1p19q deleted region gliomas. **(B)** *IDH* mutation status (wild-type, WT or mutant, MT) according to PC1 (top panel) or PC2 (bottom panel) of Figure 1B. PC1 splits samples according to their *IDH* mutation status. P-values of Wilcoxon test are indicated. **(C)** 1p19q region status (WT or deleted) according to PC1 (left panel) PC2 (right panel) of Figure 1B. PC2 splits gliomas which have 1p19q region deleted from the rest of gliomas.

Figure S3. Reproducibility of RiboMethSeq using two different methods of RNA extraction. (A) Comparison of RNA extraction methods based on C-scores of the whole cohort. Total RNA of 11 samples (S1-S11) were extracted in duplicate using either manual Trizol (blue) or automated Maxwell (red) protocols. Samples were compared based on their C-

scores at the 106 rRNA 2'Ome sites using unsupervised Principal Component Analysis (PCA). Total RNA of 11 (S1-S11) samples were extracted using Trizol and Maxwell protocol. The two methods of RNA extraction give similar results as the duplicates are close to each other in the 2D PCA plot. Percentage of variance explained by PC1 and PC2 are indicated. **(B)** Comparison of the two RNA extraction methods based on C-scores for each individual sample. Person correlation coefficient for each sample (from S1 to S11) was calculated between the two RNA extraction methods. A strong correlation between C-scores at the 106 rRNA 2'Ome was observed for all samples.

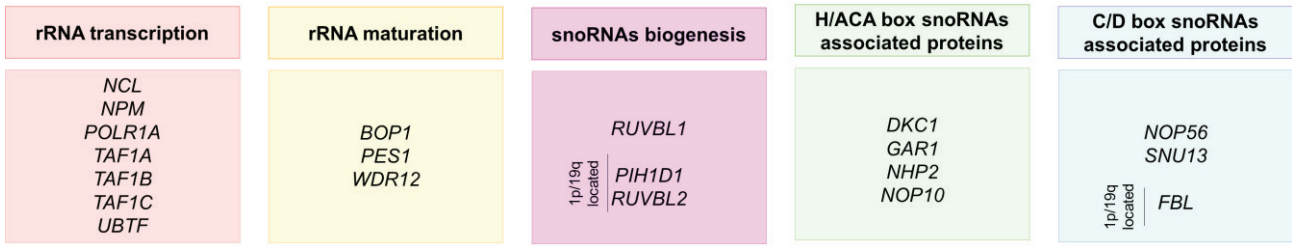
Figure S4. RiboMethSeq optimization using Illumina NovaSeq device and ComBat-seq algorithm. **(A)** Illumina NovaSeq device allows to increase number of sequenced samples per run with high depth. increase number of sequences and uniquely mapped reads. Distribution of the number of total sequenced reads (red) and uniquely mapped reads (blue) on rRNA reference is given. Using Illumina NovaSeq device to sequence 48 samples in a single run, the median of total sequenced reads is about 40 millions per samples and the median of uniquely mapped reads is about 28 millions per samples. **(B)** Batch effect adjustment of RiboMethSeq data using ComBat-seq method. Unsupervised Principal Component Analysis (PCA) was performed on the 48 samples prepared as two distinct libraries (L1, L2) based on their C-scores at all the 7055 genomic positions of rRNAs (upper panels) or at the 106 rRNA 2'Ome sites (down panels) calculated using either on non-adjusted (left panels) or adjusted (right panels) 5'-end read counts. Using the non-adjusted 5'-end read counts, data samples were separated by libraries according to the PC1 axis (left panels), indicating a batch effect. After adjustment of 5'-end read counts using ComBat-seq algorithm, the batch effect was removed (right panels). Moreover, the two control samples (RNA-ctrl) corresponding to a commercial total RNA added as a control sample in each library, displayed similar C-score profiles using adjusted 5'-end read counts (right panels).

Figure S5. Glioblastoma have a distinct rRNA 2'Ome profile, which correlates with clinical features. **(A)** rRNA 2'Ome profile discriminates glioblastoma from other glioma subtypes. Unsupervised Principal Component Analysis (PCA) on adjusted C-scores at the 106 rRNA 2'Ome sites was performed on the test cohort of 46 samples composed of glioblastoma (GBM), astrocytoma (A), oligodendroglioma (O) and healthy donors (HD). Glioblastoma display a rRNA 2'Ome profile different from the rest of the samples according to the principal component 2 (PC2). Percentage of variance explained by PC1 and PC2 are indicated. Samples are coloured according to their subtype and the 95 % confidence ellipsoids around

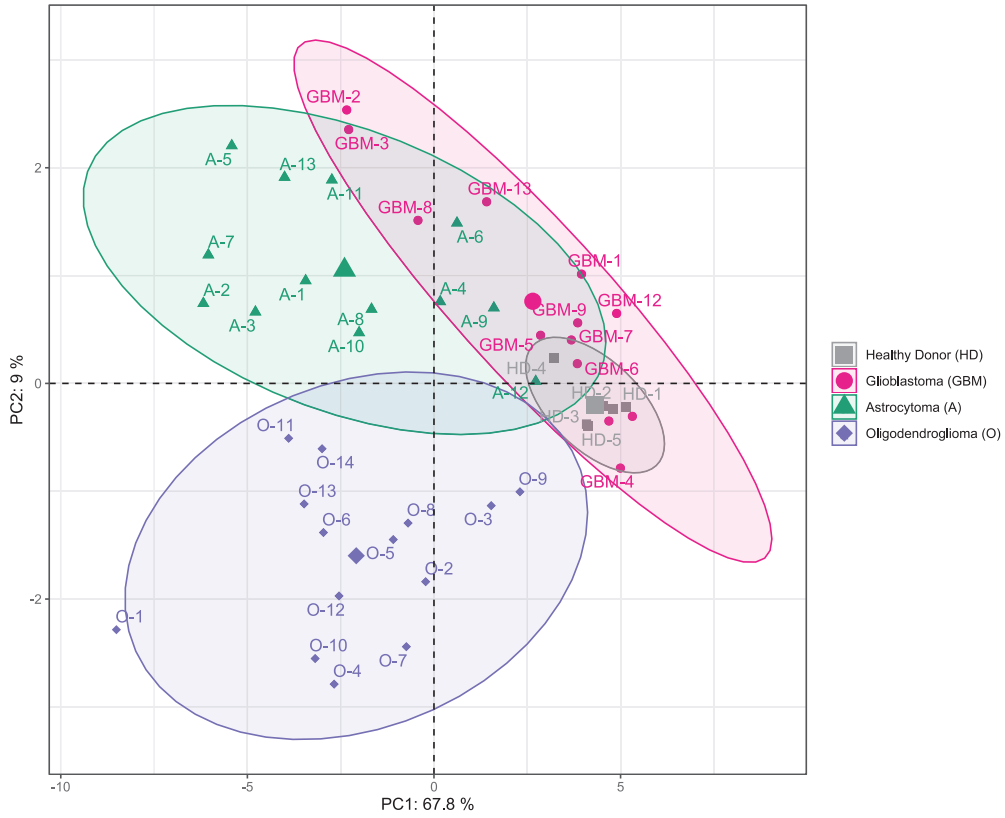
the centroid of each group are drawn. **(B)** rRNA 2'Ome profile correlates with IDH mutation status in glioma (n=40). Comparison of the distribution of the first two principal components (PC) according to the *IDH* mutation status was performed using non-parametric Wilcoxon rank-sum test. A significant difference was observed regarding *IDH* mutation status on PC2, suggesting that the upper samples of the PCA are wild-type *IDH* samples. **(C-D)** rRNA 2'Ome profile correlates with survival and mitotic index. Correlation of clinical features (OS: overall survival; PFS: progression-free survival; and mitotic index) with the first two principal components (PC) was investigated using Pearson correlation (R: correlation coefficient) (C) in the test cohort of glioma (n=40) and a summary plot of the correlation with the first five PCs is given (D). While PC1 was not associated with any clinical features, PC2 was significantly associated with mitotic index, OS and PFS, suggesting that the upper samples of the PCA display the poorest OS/PFS and the highest mitotic index. Each point represents a single sample and is colored according to the glioma subtype. Red-blue color scale: Pearson correlation coefficient (R from 1 to -1); number: adjusted p-values with *: p < 0.05; **: p < 0.01; ***: p < 0.001; and ****: p < 0.0001. **(E)** Analysis of a validation cohort composed of 3 healthy donors and 8 glioblastoma samples. PCA was performed on the C-score of the 106 rRNA 2'Ome sites and each dot corresponds to a single sample. Like in the test cohort, the rRNA 2'Ome profile discriminates glioblastoma from healthy tissues. Pink, glioblastoma (GBM); green, astrocytoma (A); violet, oligodendroglioma (O); grey: healthy donors (HD).

Figure S6. Comparison of 2'Ome levels at the 11 significantly deregulated rRNA sites in a validation cohort. At the 11 significantly deregulated rRNA sites in Glioblastoma compared to healthy donors (Figure 5), distribution of the 2'Ome levels was compared between glioblastoma (GBM, n=8) and healthy donors (HD, n=3) using box plots. Compared to healthy tissues, glioblastoma samples displayed a lower rRNA 2'Ome level at 8 sites and a higher one in only 3 sites, supporting the notion that glioblastoma tumours exhibit a global decrease in rRNA 2'Ome levels compared to non-neoplastic tissues.

A



B



C

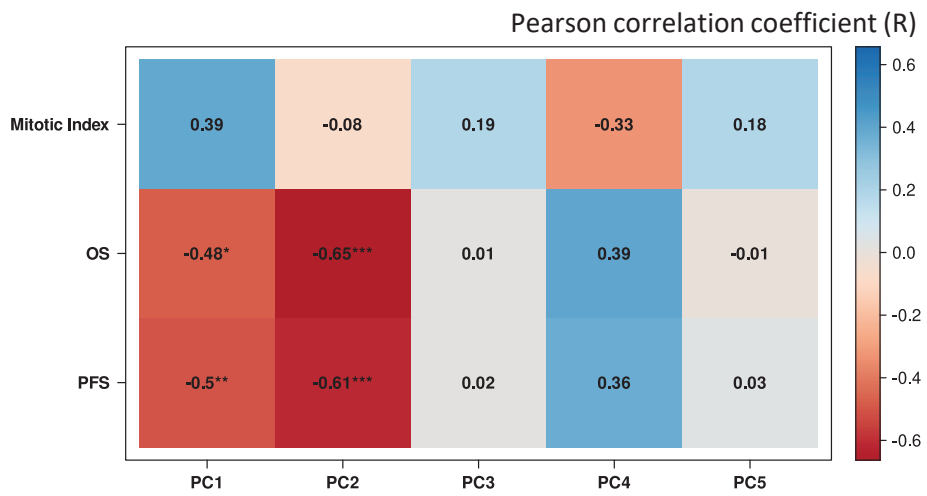


Figure 1

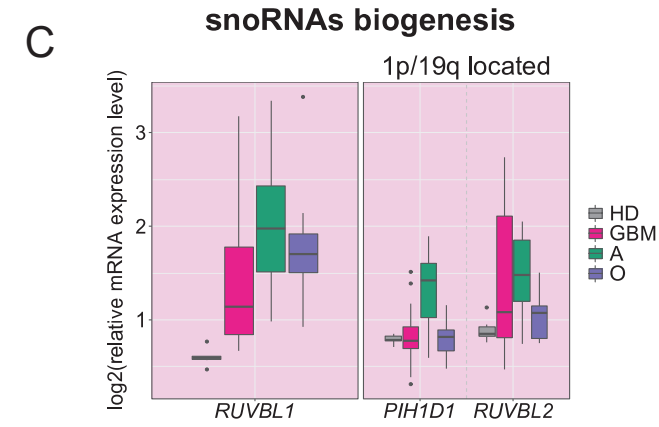
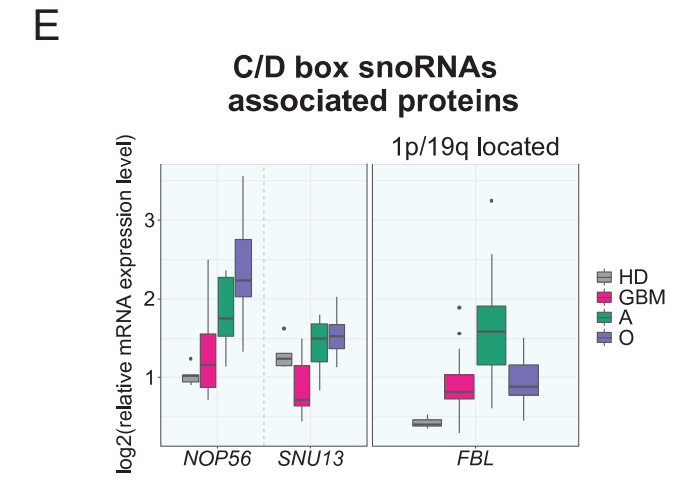
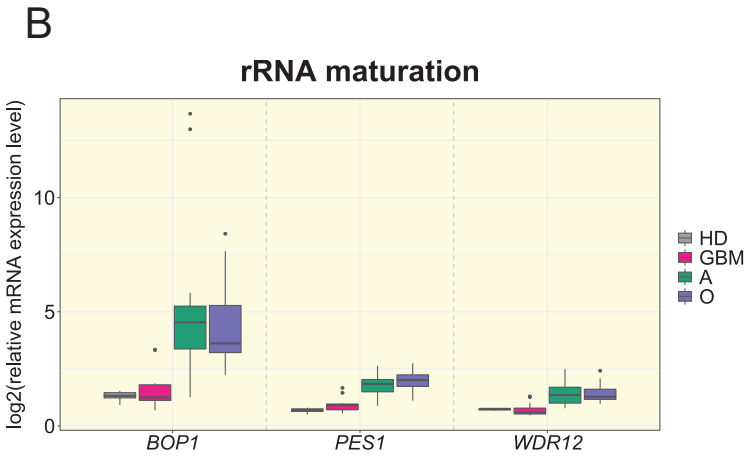
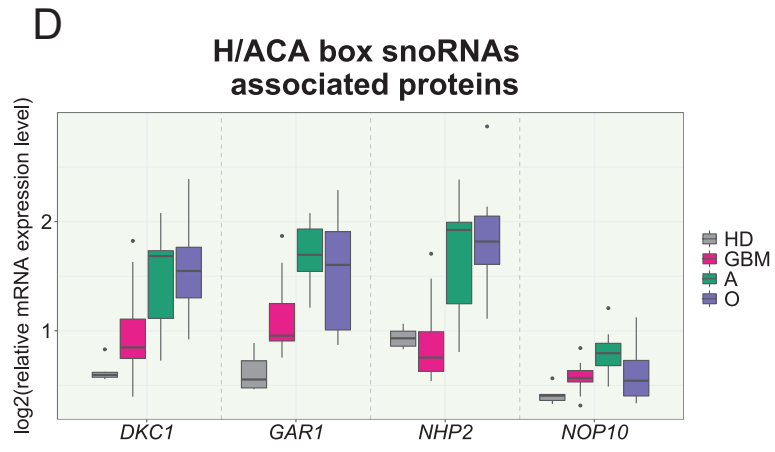
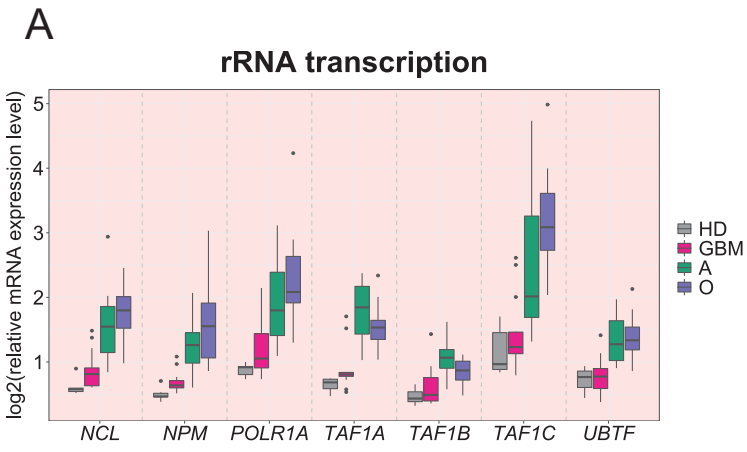


Figure 2

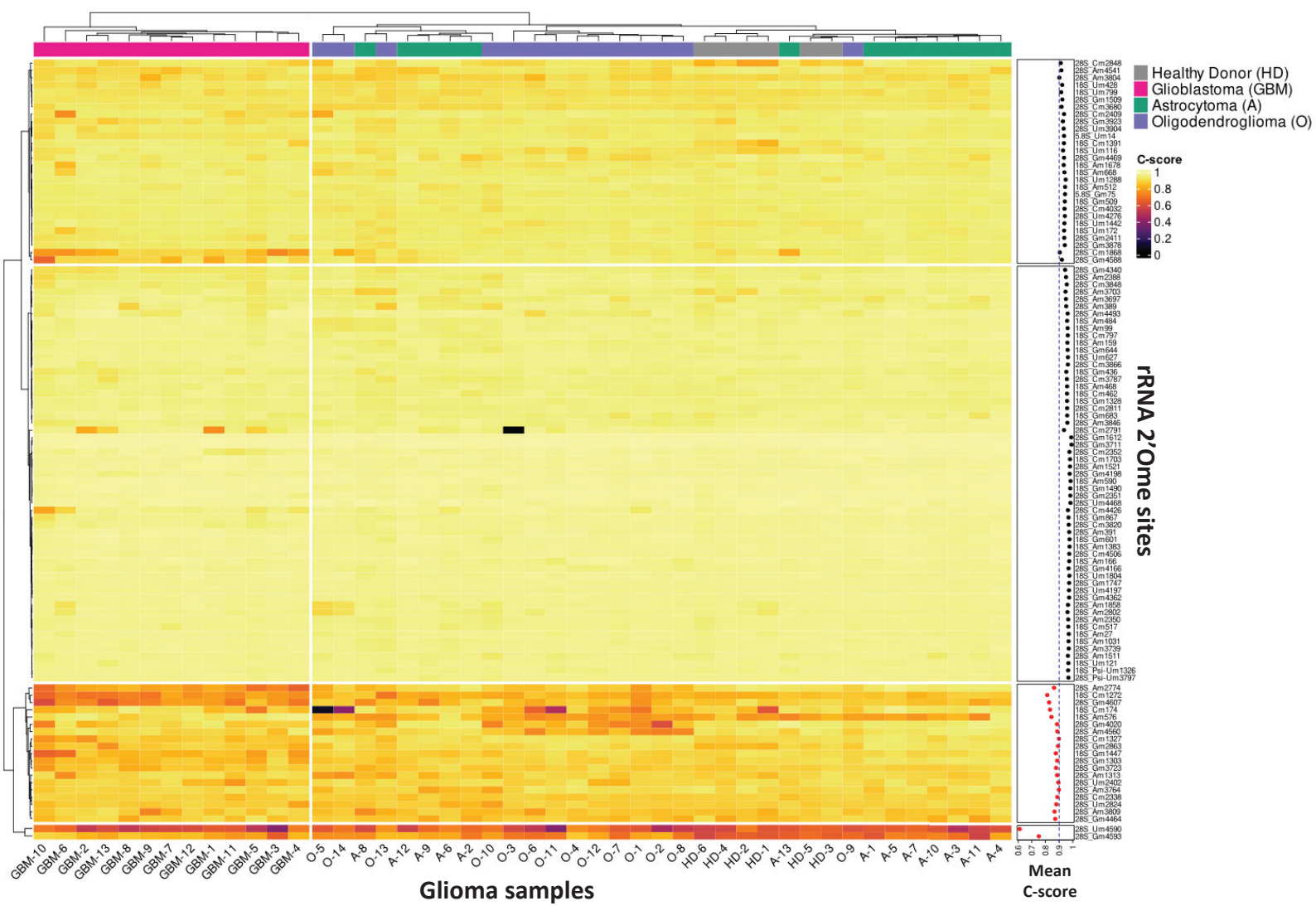
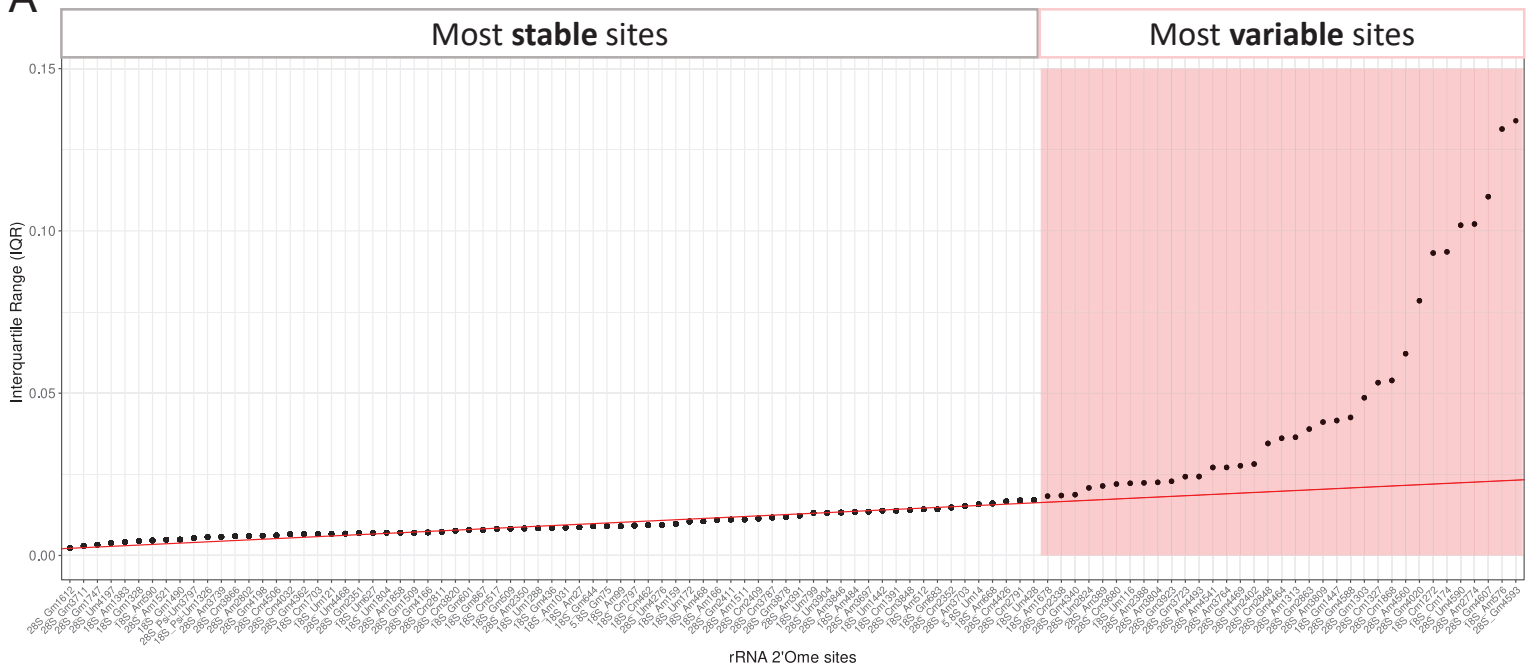


Figure 3

A

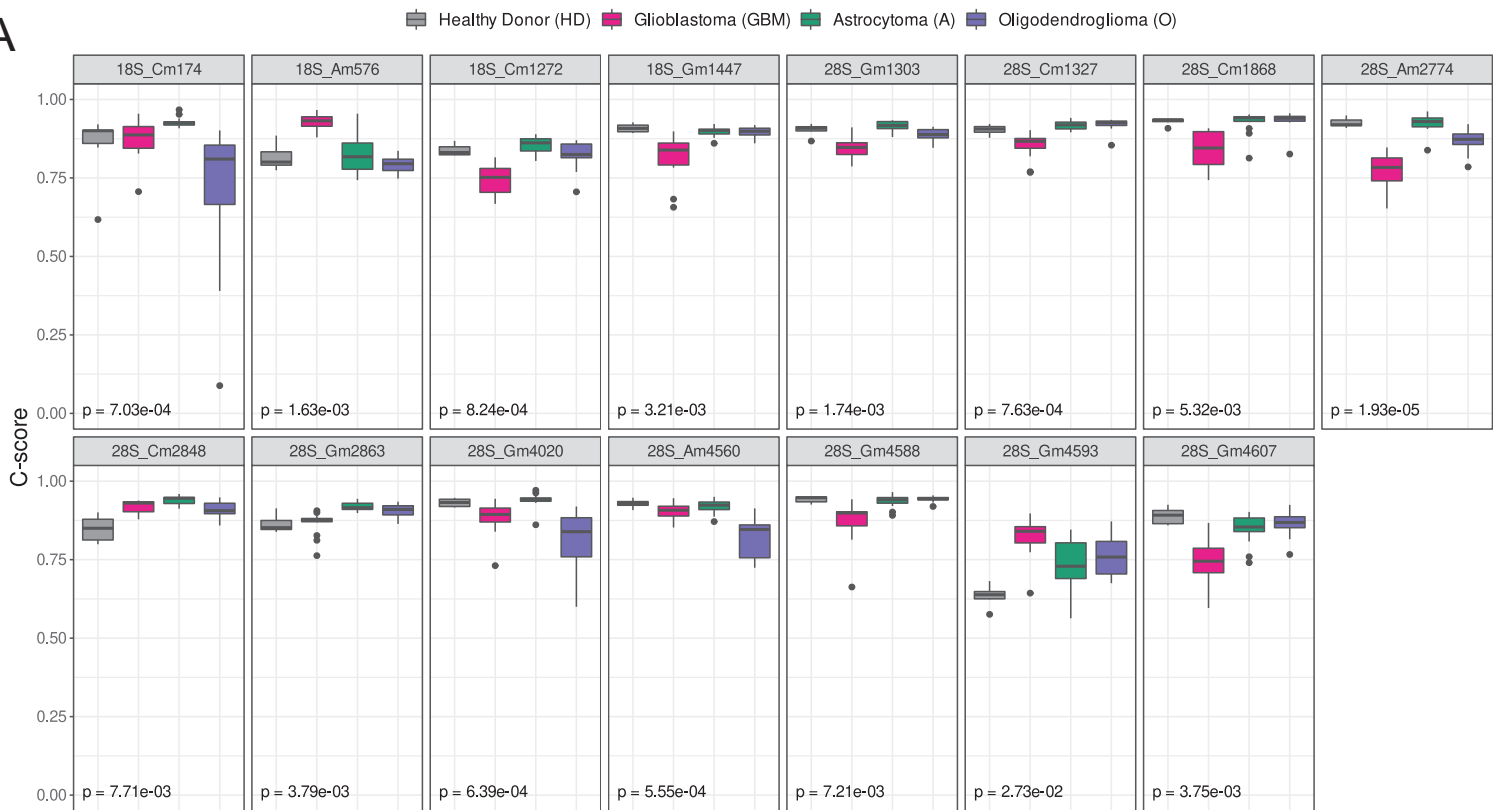


B



Figure 4

A



B

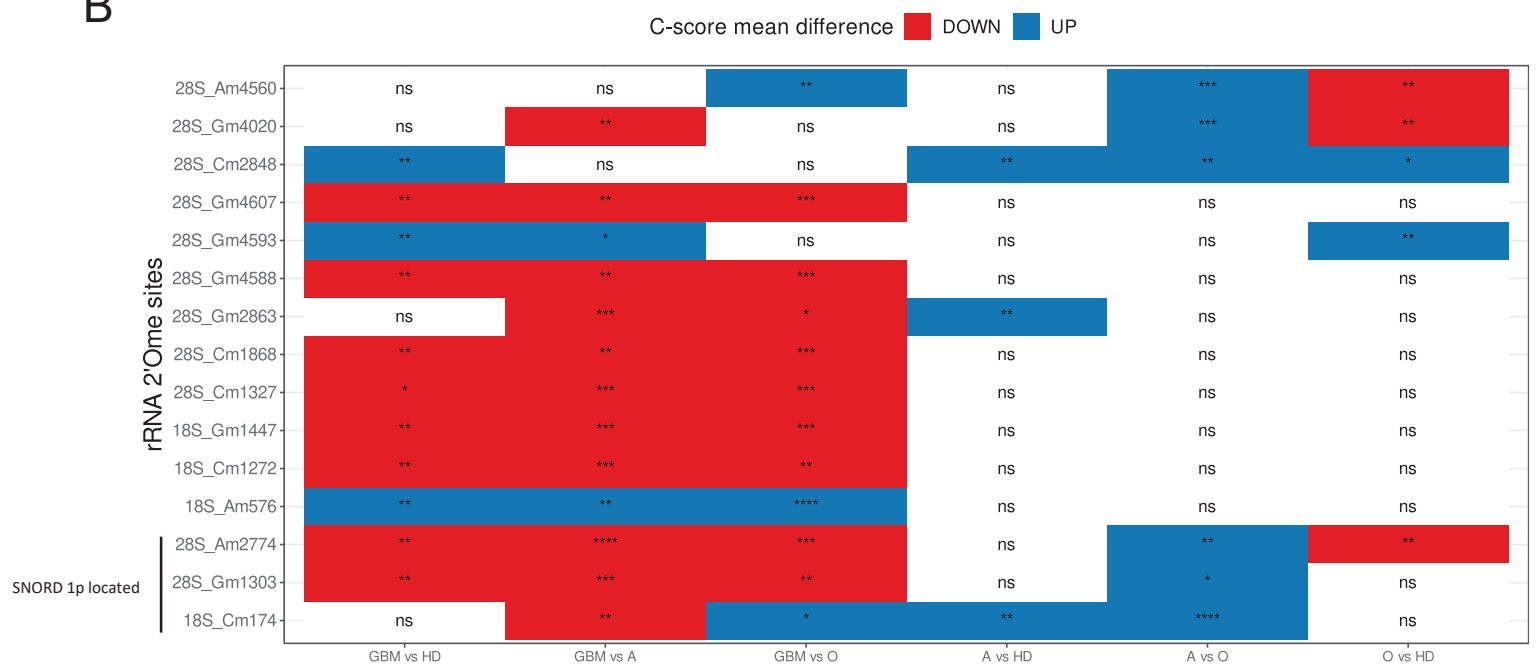
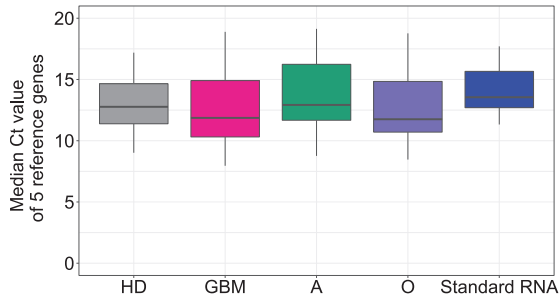
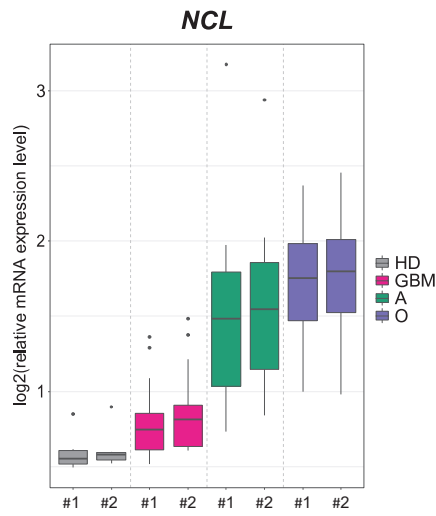


Figure 5

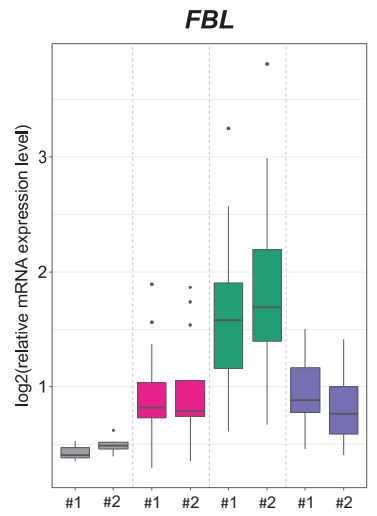
A



C



D



B

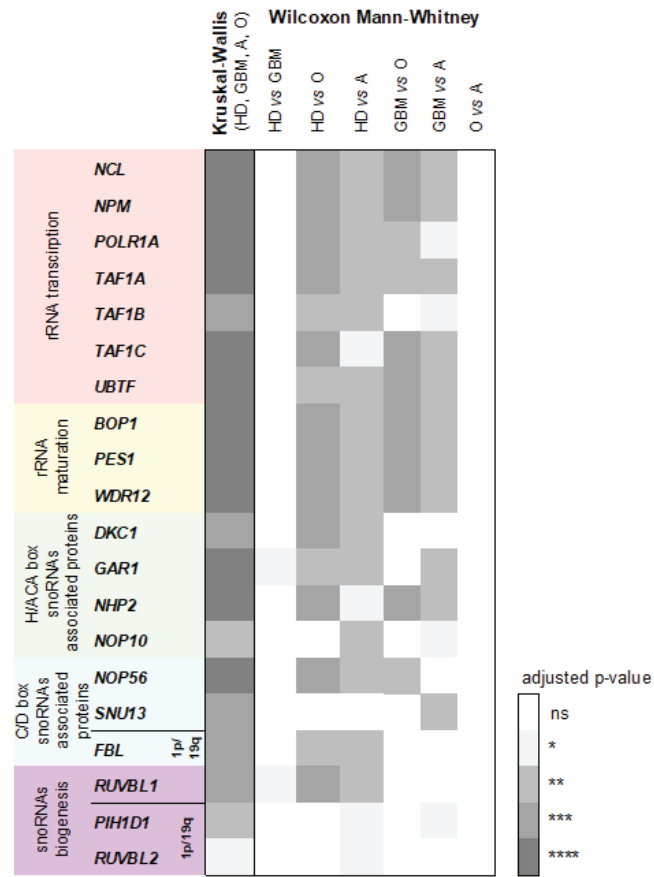


Figure S1

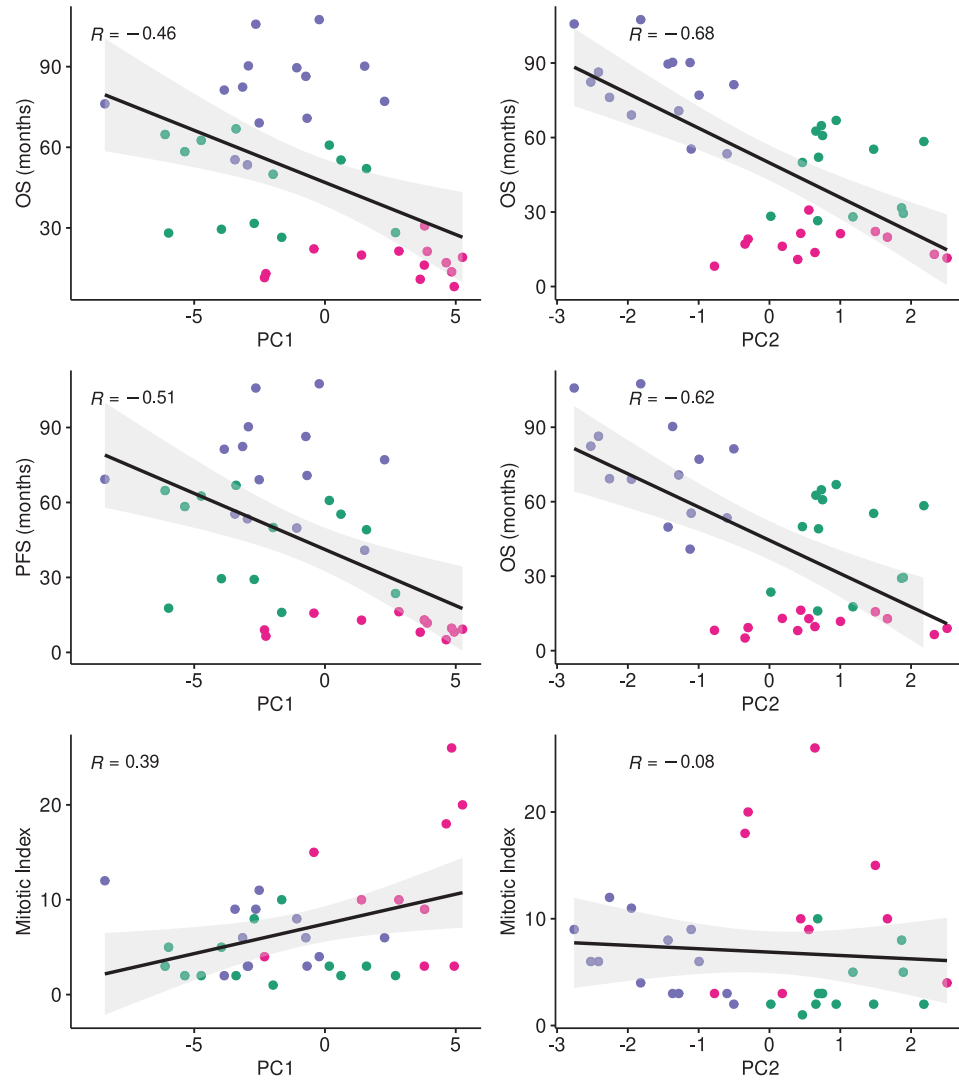
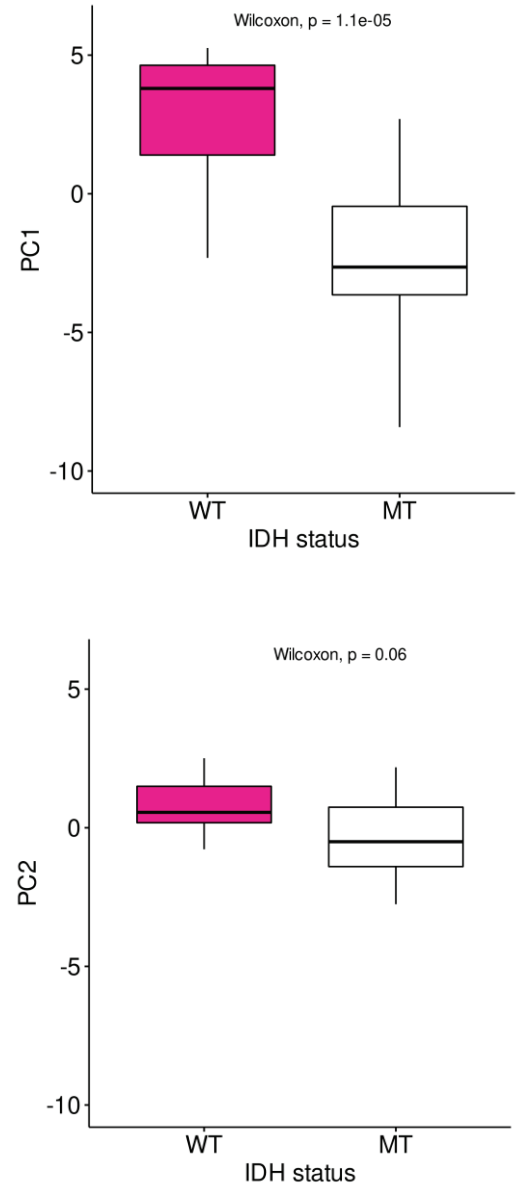
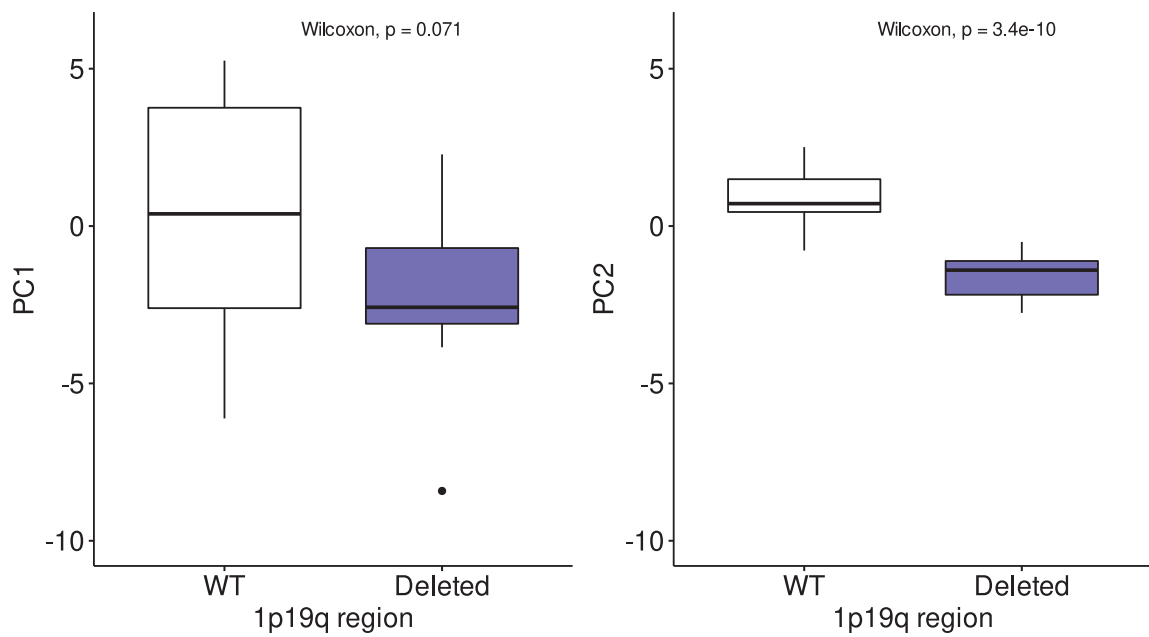
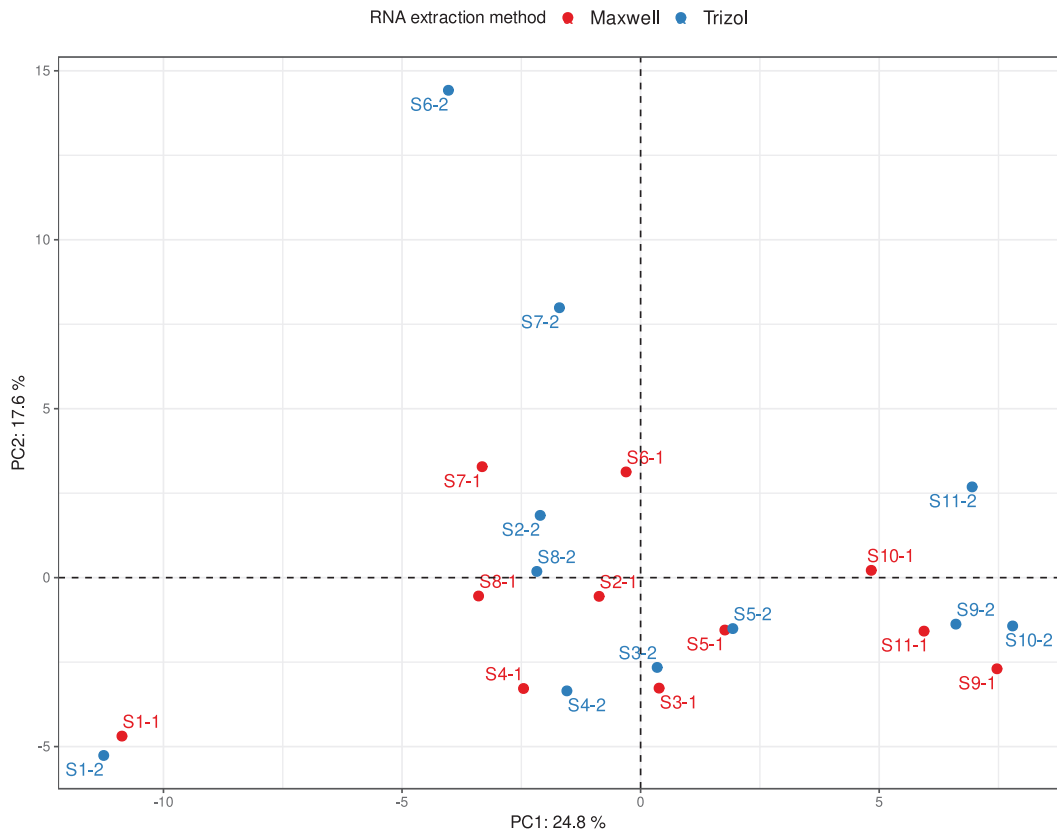
A**B****C**

Figure S2

A



B

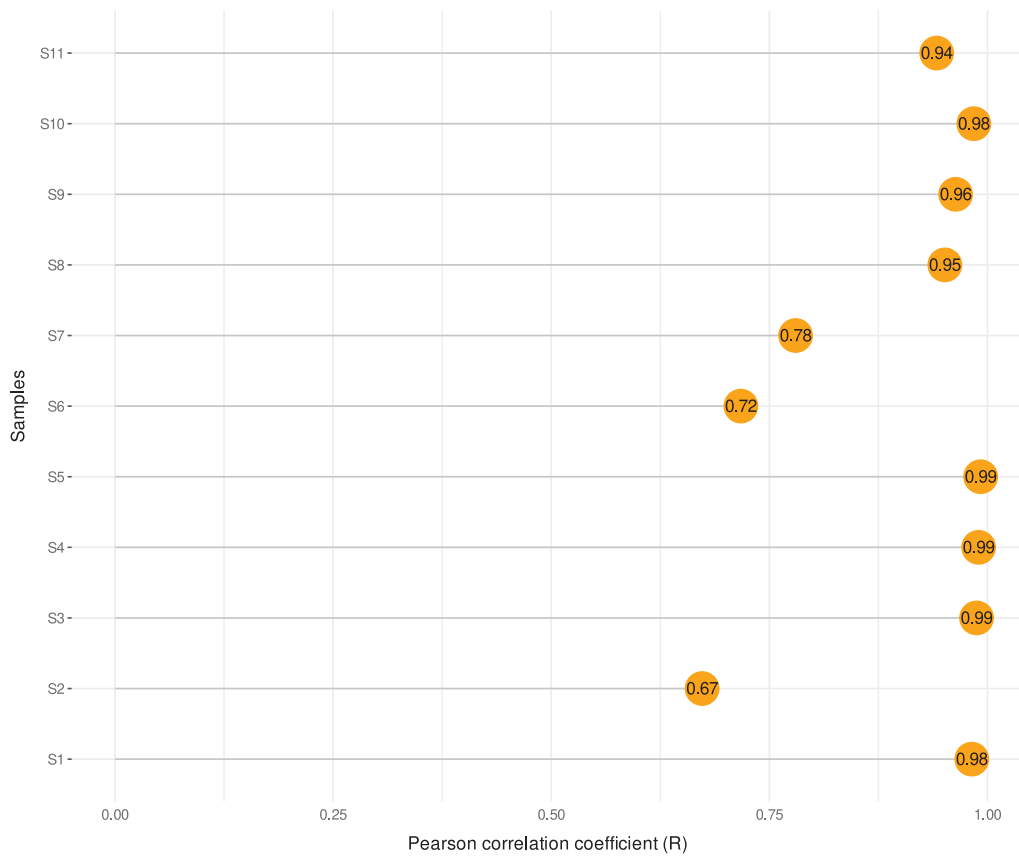


Figure S3

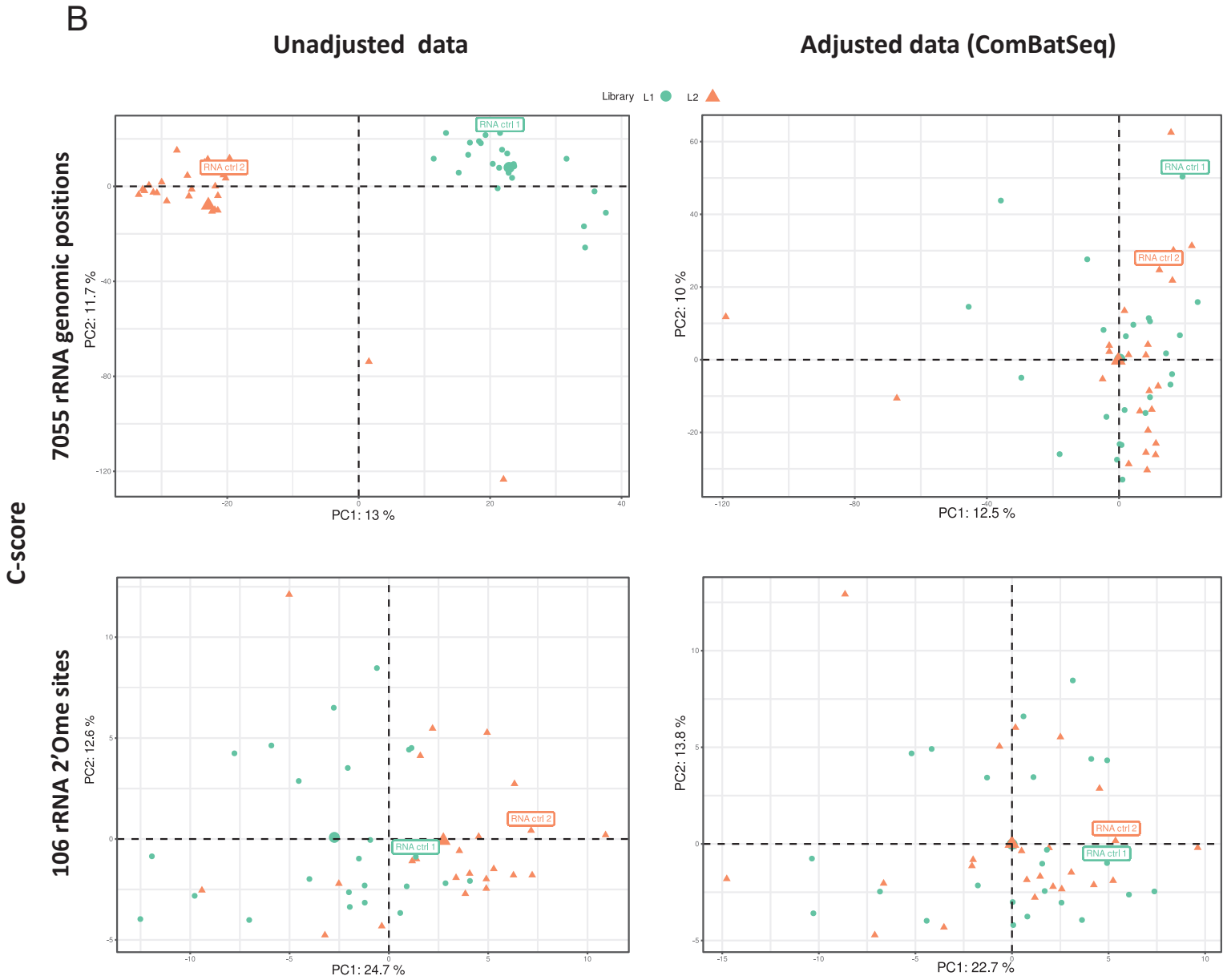
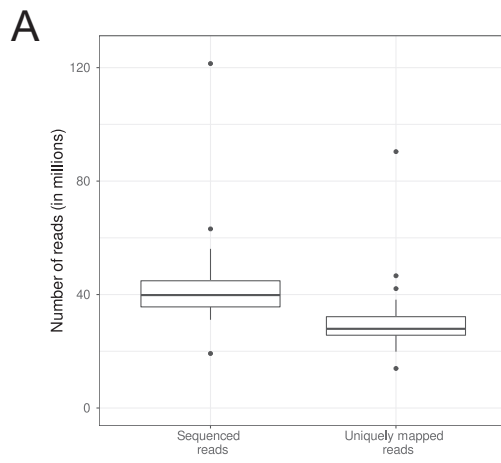
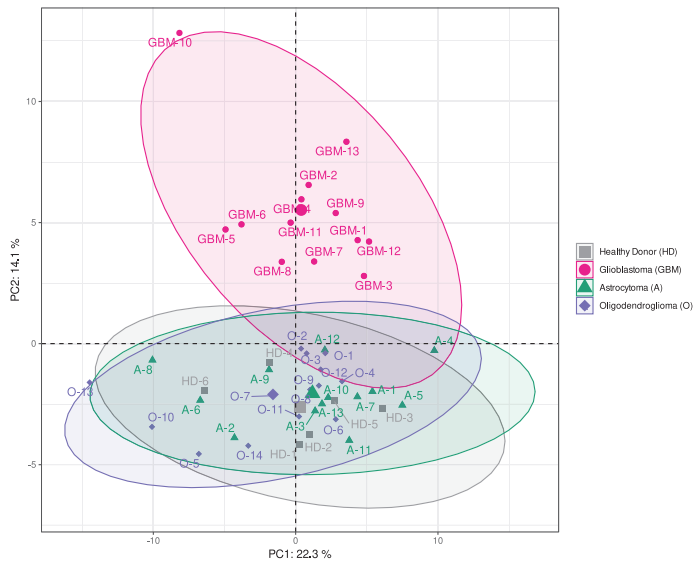
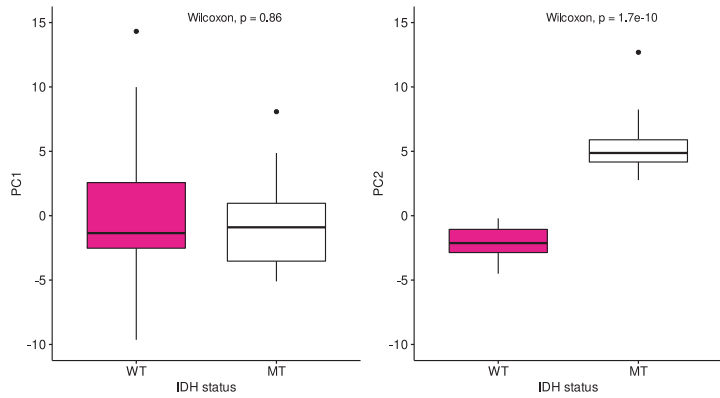
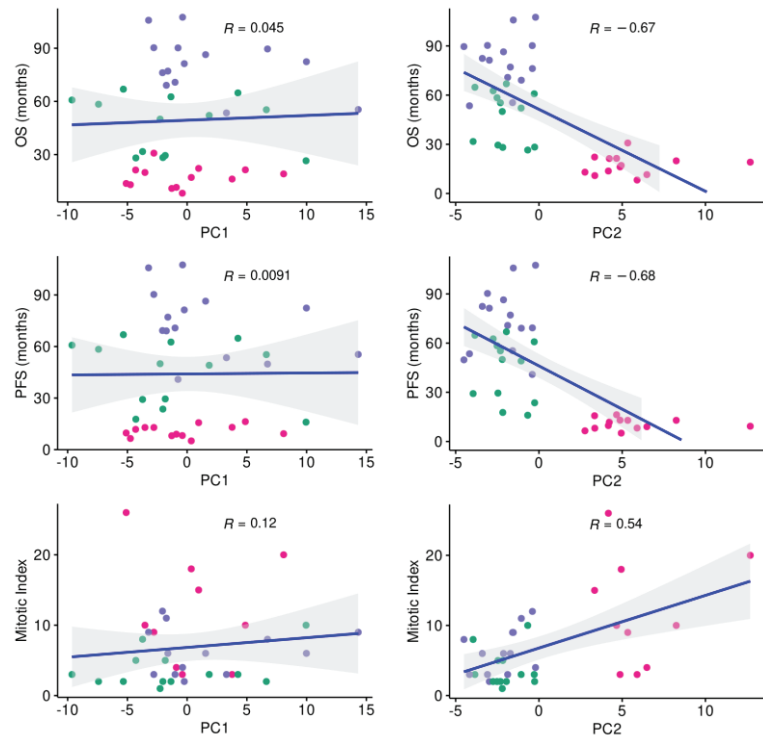
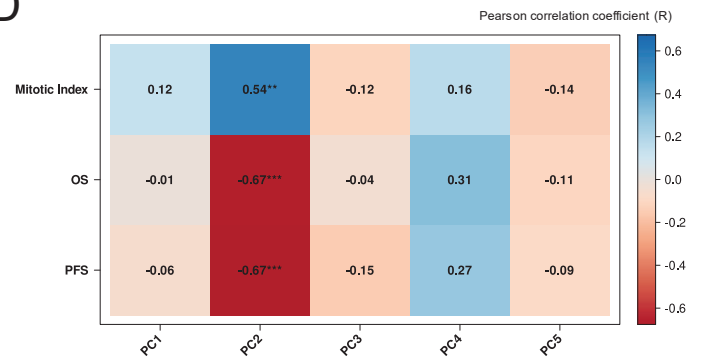
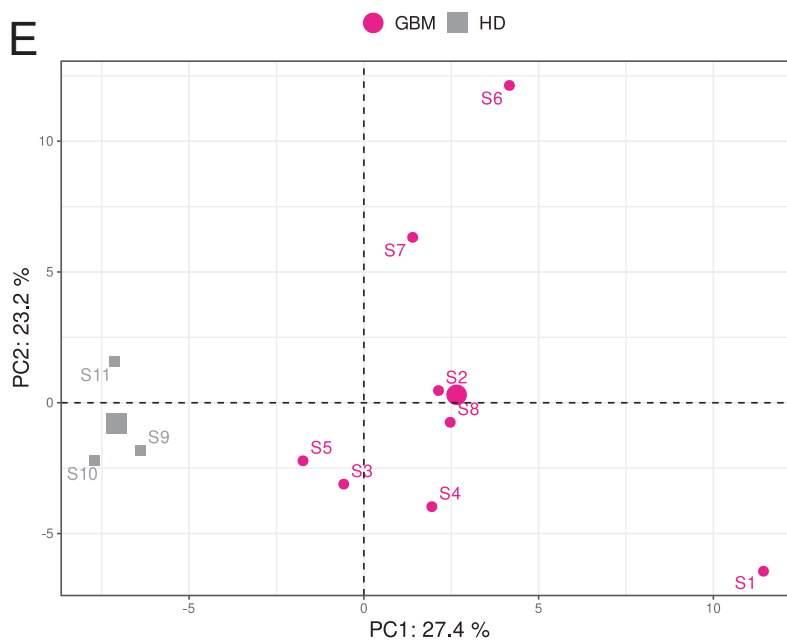


Figure S4

A**B****C****D**

n = 106 rRNA 2'Ome sites

E

HD GBM

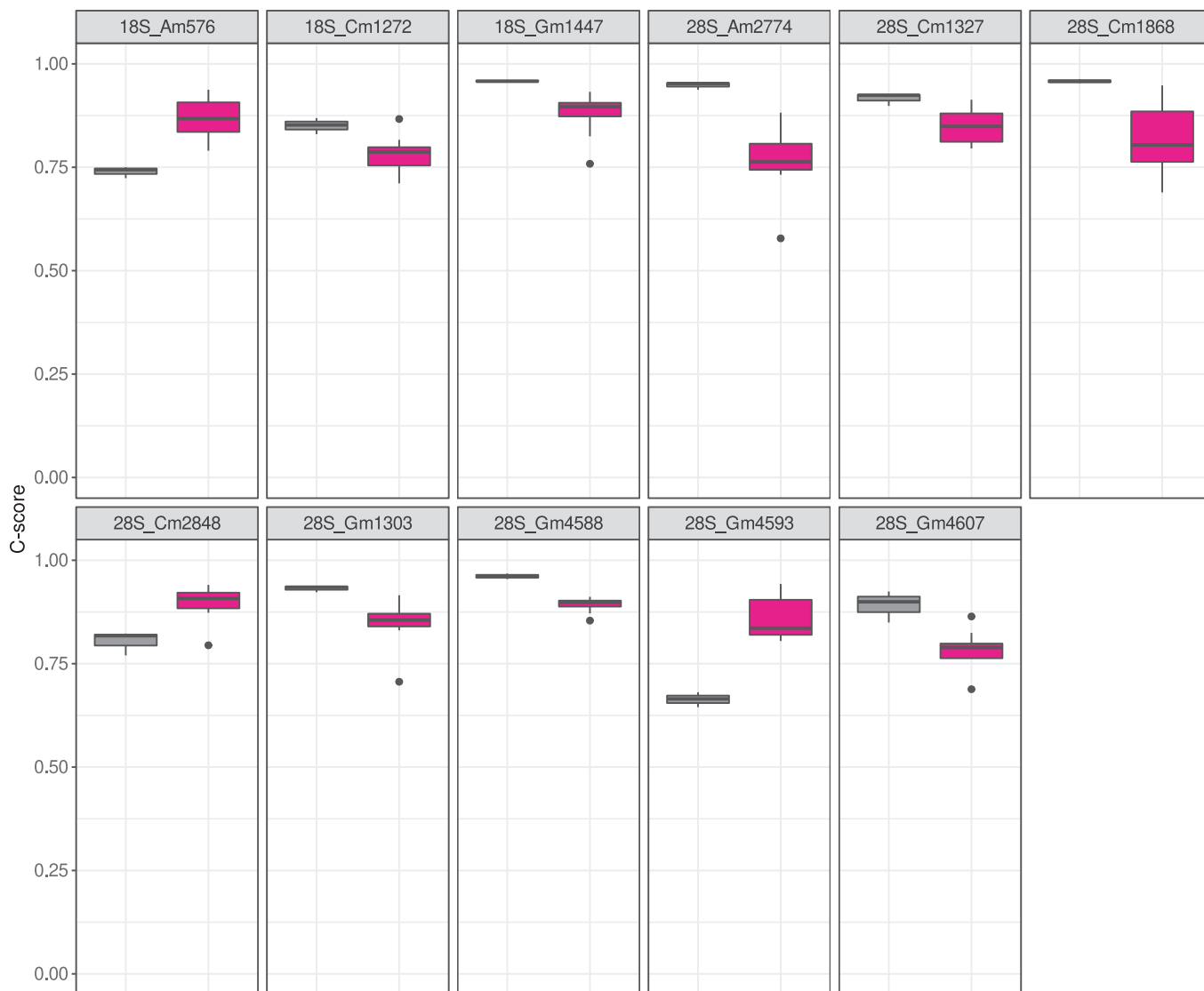


Figure S6

6 Discussion

A Développement de méthodologies et d'outils dédiés au RiboMethSeq

A.1 Évolution de la technique de RiboMethSeq et de ses outils

Pendant de nombreuses années, l'analyse de la 2'Ome des ARNr a été limitée par un manque de techniques disponibles capables de détecter et de quantifier le taux de méthylation sur la centaine de sites connus des ARNr. Le développement de nouvelles technologies omics dédiées à la 2'Ome a répondu à un besoin grandissant au sein de la communauté scientifique, mais a aussi permis d'ouvrir un nouveau champ de recherche grâce aux découvertes qui en ont découlées. En effet, pour mieux comprendre le rôle de la 2'Ome des ARNr et tester des hypothèses, la détection et la quantification de la 2'Ome des ARNr restaient impératives. Les techniques omics dédiées à la 2'Ome des ARNr ont commencé à se développer à partir de la fin de l'année 2015 et sont donc relativement nouvelles, encore aujourd'hui (Birkedal 2015). En 2015, le RiboMethSeq basé sur un séquençage Ion Torrent permettait l'analyse de la 2'Ome chez la levure (BIRKEDAL, CHRISTENSEN-DALSGAARD et al., 2014). En 2016, l'adaptation du protocole pour un séquençage Illumina chez la levure a permis de rendre cette technologie plus accessible (MARCHAND, BLANLOEIL-OILLO et al., 2016). En collaboration avec la plateforme de séquençage EpiRNA-Seq de l'Université de Lorraine, nous avons contribué à adapter l'utilisation du RiboMethSeq pour l'analyse de lignées cellulaires humaines en 2017 puis d'échantillons humains en 2021 (ERALES et al., 2017; Virginie MARCEL, KIELBASSA et al., 2020).

Parmi les techniques haut-débit disponibles pour analyser la 2'Ome des ARNr, le RiboMethSeq semble être la seule technique ayant pris de l'ampleur et

continuant à être utilisée dans l'analyse de la 2'Ome par de nombreux laboratoires (JAAFAR et al., 2021). En effet, l'adaptation du protocole pour du séquençage Illumina a permis une utilisation en routine et une analyse de la 2'Ome dans des échantillons d'origine diverse (MARCHAND, BLANLOEIL-OILLO et al., 2016). Jusqu'au début de ma thèse, la plupart des études dédiées à la 2'Ome, se sont focalisées sur un petit nombre d'échantillons provenant principalement de lignées cellulaires ou de comparaison entre deux conditions biologiques. Un de nos objectifs était d'utiliser le RiboMethSeq en routine pour analyser de grandes cohortes d'échantillons humains afin de déterminer la diversité de la variation de la 2'Ome dans les cancers. Pour faciliter la prise en charge de grandes cohortes, du traitement des échantillons aux traitements des données de séquençage et l'analyse des données, nous avons commencé par implémenter le RiboMethSeq au sein du Centre de Recherche en Cancérologie de Lyon (CRCL) grâce aux savoirs-faire des plateformes de Génomique des Cancers et de Bioinformatique Gilles Thomas.

Outils de contrôle qualité (QC) : développement d'un pipeline de processing des données.

Dès le début de ma thèse, je me suis concentré sur l'amélioration de l'approche de RiboMethSeq et le développement d'outils bioinformatiques dédiés à l'analyse de grandes séries d'échantillons humains. Nous avons couplé pour la première fois l'approche RiboMethSeq avec un séquençage haut-débit de dernière génération de type Illumina NovaSeq 6000 (Paraquindes et al, en préparation – D). Par rapport aux études actuelles, cela nous a permis d'augmenter la profondeur de séquençage d'environ 5x (30 millions de lectures alignées sur ARNr) et de multiplier par 2 le nombre d'échantillons séquencés en même temps (48 échantillons dans un seul run). L'utilisation du NovaSeq 6000 a permis de stabiliser la technologie et de permettre la détection d'une faible variabilité du niveau de 2'Ome (Virginie MARCEL, KIELBASSA et al., 2020).

De plus, nous avons mis en place des étapes de contrôle qualité à différents niveaux qui sont actuellement inexistantes dans les publications. Ces étapes sont effectivement nécessaires lors d'analyses de grandes cohortes qui reposent sur des échantillons biologiques complexes (ex : dégradation, faible quantité, fixation, . . .) et un jeu de données issu de plusieurs runs de séquençage.

1) Pré-séquençage afin d'identifier les échantillons ayant un état de dégradation avancé. En effet, les échantillons humains sont souvent limités en quantité et la qualité peut varier fortement. Pour le moment, nos études se sont limitées à l'utilisation de tissus congelés pour pallier cette problématique, mais il faut savoir que la plupart des collections tumorales sont conservées sous forme FFPE, une fixation chimique qui accélère la dégradation des tissus. Ainsi, nous avons identifié des seuils pour certains paramètres au niveau wetlab, comme le RIN number, pour identifier avant le séquençage des problèmes de quantité et de qualité de l'échantillon afin d'éviter de perdre des ressources dans des échantillons de mauvaise qualité.

2) Post-séquençage afin d'identifier toute anomalie issue du séquençage. Nous avons utilisé des outils comme FastQC et multiqc pour identifier des paramètres propres aux données de RiboMethSeq. Nous avons ainsi examiné la qualité du séquençage, mis en évidence des problématiques de contamination des lectures et identifié des paramètres comme le nombre de lectures alignées uniquement une fois sur l'ARNr. Par exemple, les échantillons présentant des problèmes de séquençage ou de contamination ont souvent peu de lectures qui s'alignent sur l'ARNr de référence.

L'ensemble des paramètres que j'ai identifiés est inclus dans un pipeline bio-informatique qui réalise l'étape de contrôle qualité, l'alignement sur l'ARNr de référence et le comptage des extrémités des lectures en 5'-end et 3'-end. Le pipeline utilise le gestionnaire nextflow afin de mieux gérer les fichiers d'entrée et de sortie ainsi que l'exécution en parallèle des différentes étapes. Chaque étape possède son propre environnement conda, ce qui facilite l'exportation et l'utilisation du pipeline par d'autres utilisateurs sur d'autres systèmes. Cet outil devrait être disponible au grand public d'ici à la fin de l'année 2022.

3) Pré-analytique afin d'identifier les échantillons « outliers », c'est-à-dire d'identifier les échantillons très différents des autres au sein d'une série. Pour cela, j'ai identifié un ensemble de paramètres permettant d'identifier les outliers et de pouvoir les exclure objectivement de l'analyse. Ceci repose par exemple sur des défauts de comptage des extrémités des lectures, ou de la valeur des Cscores à l'ensemble des positions génomiques. Cette étape sera intégrée dans la seconde version du package (cf. [A](#)).

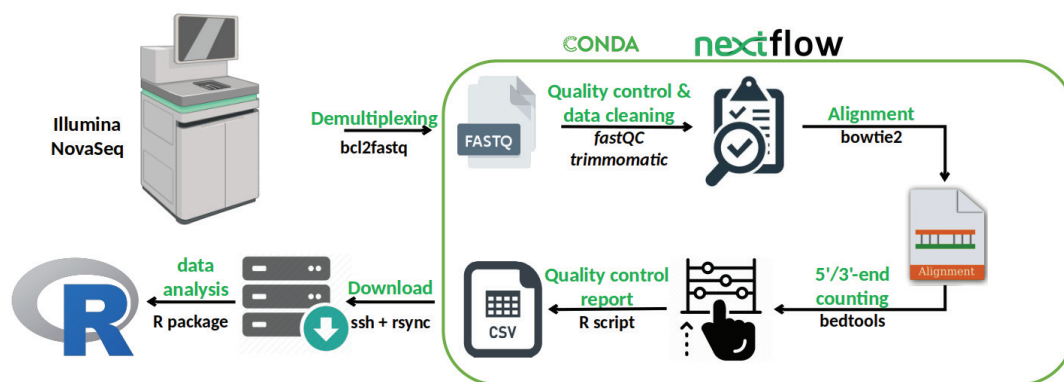


FIGURE 6.1 – Schéma représentant le pipeline de traitement des données de séquençage afin d’obtenir les données de comptages des extrémités 5’ et 3’.

Outils d’analyse des données de RiboMethSeq : développement du Package R.

L’ensemble de ces développements était nécessaire pour l’analyse de grandes cohortes, mais aussi parce qu’aucun outil bioinformatique dédié au RiboMethSeq n’était disponible au début de ma thèse. Aujourd’hui encore, aucun outil clé-en-main n’est disponible pour l’étape post-séquençage. Seulement un outil sous forme de package R est disponible : RNAmoDR. Ce package contient un module appelé RNAmoDR.RiboMethSeq qui se limite au calcul du Cscore comme défini par Birkedal et al pour l’ensemble des positions génomiques d’un ARN d’intérêt. Cependant, ce package nécessite en entrée (i) des fichiers d’alignement au format bam, qui sont souvent très volumineux (jusqu’à 1Go) et difficile à manipuler avec le logiciel R, (ii) des fichiers d’annotation des ARN au format gff ou objet TxDb créé avec le package GenomicFeatures et (iii) finalement, de séparer les échantillons en deux groupes (contrôle vs traité). D’après notre expérience, l’utilisateur a souvent accès aux données de comptage qui sont stockées dans des fichiers texte. En sortie, ce module RNAmoDR.RiboMethSeq permet seulement deux types de visualisation des Cscores : une heatmap avec les Cscores des sites par ARN (sans regroupement hiérarchique) et les Cscores par échantillon à des localisations précises.

Dans sa globalité, ce package ne peut pas être utilisé pour analyser la 2’Ome dans une cohorte et comparer plus de deux conditions. C’est une des raisons pour laquelle nous avons commencé à compiler dans un package R

les différentes fonctions et méthodes que j'ai développées et testées pendant ma thèse. Nous nous sommes inspirés des packages R couramment utilisés dans l'analyse de données de type RNA-seq comme DESeq2. Les données de comptages nécessaires pour notre package peuvent être obtenues en utilisant notre pipeline de traitement des données de séquençage ou par d'autres méthodes. La première version du package permet de lire ces données de comptages, calculer un Cscore pour chaque position génomique, explorer les données à l'aide d'une ACP, normaliser d'éventuels effets batch et exporter les résultats dans un fichier texte.

Nous avons conçu ce package pour permettre des évolutions futures. Dans ses prochaines versions, le package permettra de réaliser des étapes pré-analytiques supplémentaires (ex : identification des échantillons outliers) et analytiques (ex : analyse de la variation de la 2'Ome comme utilisée dans nos publications, regroupement hiérarchiques, tests statistiques non-paramétriques). Cet outil permettra une homogénéisation de la définition des sites stables et variants, et ainsi de comparer les données de RiboMethSeq entre les différentes études.

A.2 Utilisation du RiboMethSeq dans d'autres espèces

L'ensemble des données disponibles dans la littérature et issues de cette thèse soutiennent l'hypothèse d'une hétérogénéité de la 2'Ome des ARNr et d'une association de la 2'Ome avec des contextes biologiques particuliers. Ces données proviennent de l'analyse de nombreux modèles reflétant le règne du vivant. En effet, la cartographie des sites de 2'Ome des ARNr commence à être établie pour de nombreuses espèces modèles comme le poisson zèbre ou la plante *Arabidopsis Thaliana* (JAAFAR et al., 2021). Chez l'Homme cette cartographie semble être bien établie, même si certains sites spécifiques à une condition biologique ou un tissu peuvent encore être manquants. Mais certaines problématiques restent à abordées.

Cartographie des sites 2'Ome des ARNr chez la souris. L'ensemble des données disponibles dans la littérature et issues de cette thèse soutiennent (i) l'hypothèse d'une hétérogénéité intra- et inter-tumorale de la 2'Ome des ARNr dans les cancers et (ii) l'association de la 2'Ome avec des caractéristiques biologiques et cliniques des cancers. Aujourd'hui, il est important de comprendre comment cette hétérogénéité est mise en place et comment elle

contribue au processus de tumorigenèse. Pour cela, il est nécessaire de développer des approches *in vivo*.

Un des modèles les plus utilisés pour comprendre les processus biologiques *in vivo* reste le modèle murin. Néanmoins, aujourd'hui, il n'existe aucune cartographie des sites 2'Ome des ARNr chez la souris, empêchant ainsi l'utilisation des modèles murins pour mieux comprendre le rôle biologique de la 2'Ome. Il faut effectivement rappeler que cette cartographie est une connaissance indispensable pour utiliser le RiboMethSeq qui permet de calculer un Cscore pour toutes les positions génomiques des ARNr avant extraction des Cscores des sites méthylés connus.

Dans la littérature, il existe deux études dédiées à l'analyse de la 2'Ome chez la souris (HEBRAS et al., 2019; INCARNATO et al., 2016). Ces deux études reposent sur l'hypothèse que les sites cartographiés chez l'Homme existent tous chez la souris. Cependant, il est fort probable que certains sites 2'Ome soient spécifiques à l'Homme, et que d'autres n'existent que chez la souris.

Pendant ma thèse, j'ai commencé un travail *in silico* de cartographie des sites méthylés chez la souris. Comme les deux études précédentes, je suis parti d'une hypothèse de conservation de la localisation des sites méthylés au cours de l'évolution entre la levure et l'Homme, deux organismes qui ont des cartographies bien établies (TAOKA, NOBE, YAMAKI, SATO et al., 2018; TAOKA, NOBE, YAMAKI, YAMAUCHI et al., 2016). J'ai identifié les correspondances des sites entre les espèces (levure, humain et souris) par alignement de séquence. Dans un second temps, et contrairement aux études disponibles, j'ai postulé que comme la 2'Ome des ARNr est guidée par des snoARNs, l'analyse de ceux-ci me permettrait de raffiner les sites murins. Pour cela, j'ai donc extrait les snoARN murins ainsi que leurs sites prédits de 2'Ome des ARNr à partir de la base de données snOPY, en me focalisant seulement sur les snoARNs identifiés chez la souris (YOSHIHAMA, NAKAO et KENMOCHI, 2013). Cela m'a permis de générer une liste de 150 sites potentiellement méthylés chez la souris. Par la suite, j'ai utilisé nos savoirs-faire et outils (pipeline de traitement des données de séquençage et package R) ainsi que des données de RiboMethSeq réalisées chez la souris par mon équipe et nos collaborateurs pour raffiner cette liste. Les données disponibles provenaient d'un large spectre de conditions biologiques (tissus de cerveau de souris adulte, lignées cellulaires murines cancéreuses pancréatiques et murines embryonnaires, $n = 44$). Ainsi, j'ai pu éliminer 20 positions présentant

un niveau de 2'Ome nul dans toutes les conditions biologiques, suggérant que ces nucléotides ne sont pas 2'Ome.

La liste définie in silico de 130 sites potentiellement méthylés est en train d'être validée par des méthodes biochimiques (spectrométrie de masse et extensions d'amorces). En attendant, la liste peut déjà être utilisée pour étudier la 2'Ome chez la souris à l'aide du package R que nous avons développé. À l'avenir, la liste des sites méthylés validés, ainsi que les Cscores associés, seront intégrés dans la base de données snoDB (cf [B](#) qui a pour objectif d'élargir son champ d'application qui est actuellement restreint aux données humaines).

Nomenclature des sites 2'Ome. Une des problématiques rencontrées au cours de ma thèse concerne les différentes séquences humaines d'ARNr qui existent dans la littérature et qui sont utilisées pour l'étude de la 2'Ome. On peut compter jusqu'à trois versions majeures : (i) celle utilisée par snoRNABase qui continue à être une base de données de référence, même si elle n'est plus mise à jour depuis 2007 (LESTRADE, 2006), (ii) la séquence proposée par la base de donnée snOPY qui est basée sur une ancienne version provenant de NCBI (YOSHIHAMA, NAKAO et KENMOCHI, 2013) et (iii) la version proposée par NCBI et utilisée pour la structure du ribosome (H. KHATTER et al., 2015; Heena KHATTER et al., 2015). Cette dernière est la plus récente et sert de référence dans nos analyses afin de faciliter la modélisation de nos données sur la structure du ribosome. Afin de pallier la problématique induite par l'existence de plusieurs ARNr de référence humains, une table de conversion est implémentée dans la nouvelle version de la base de donnée snoDB qui contient la position dans chacune des séquences d'ARNr ainsi que le nucléotide à cette position spécifique.

La problématique d'avoir plusieurs séquences de référence disponibles est que les différences entre les séquences entraînent une différence de localisation des sites de 2'Ome, et donc de désignation de ces sites. Ces différences de localisation de la 2'Ome impactent peu les ARNr 5.8S et 18S qui sont les plus conservés au cours de l'évolution, mais beaucoup plus l'ARNr le plus long, l'ARNr 28S (jusqu'à 40 nucléotides de plus en fonction des séquences de référence). Par exemple, le site méthylé de l'ARNr 28S en position 1303 selon snoRNABase se situe à la même position dans la séquence de snOPY, mais en position 1316 dans la séquence de la structure du ribosome. La différence

de localisation des sites 2'Ome complique la comparaison des sites d'intérêt d'une étude à l'autre. La nomenclature proposée par snoRNABase continue à être utilisée dans la plupart des études, même si cette nomenclature se limite à indiquer la nature et la position du nucléotide modifié (ex. Gm1447, Um14). Pour pallier ce problème de diversité des séquences de référence des ARNr humains, tout au long de ce manuscrit j'ai utilisé une nomenclature des sites 2'Ome qui explicite l'ARNr, la nature du nucléotide modifié et sa position dans la structure du ribosome (18S_Gm1447 ou 5.8S_Um14). Cette nomenclature contient toutes les informations nécessaires à l'identification d'un site 2'Ome.

Néanmoins, cette proposition ne concerne que les sites 2'Ome chez l'Homme. À long terme, il sera nécessaire d'implémenter une nomenclature universelle des sites 2'Ome qui prenne en compte la conservation (et la non-conservation) des sites au cours de l'évolution afin de faciliter les comparaisons entre espèces. Ce type de nomenclature a déjà été mis en place pour les protéines ribosomiques (BAN et al., 2014). Par exemple, les protéines présentes dans les ribosomes des trois domaines (bactéries, archées et eucaryotes) sont désignées par le préfixe "u" (pour universel). La mise en place d'une nomenclature unique ne résoudra pas entièrement la problématique de nomenclature des sites 2'Ome, elle devra inclure l'identification d'une séquence de référence des ARNr universels, et ce, pour chaque espèce. De plus, cette nomenclature demande la collaboration d'un grand nombre d'acteurs du domaine du ribosome, mais aidera dans le futur à clarifier et à améliorer notre compréhension de la 2'Ome des ARNr.

A.3 Utilisation du RiboMethSeq pour d'autres applications ?

À ce jour, seul le comptage des extrémités des lectures issues du RiboMethSeq est utilisé. Or, nos derniers développements et notamment l'utilisation d'un séquenceur de type NovaSeq 6000 permet d'obtenir une grande profondeur de lecture et des lectures de bonne qualité. On peut donc se demander si ces informations non utilisées aujourd'hui pourraient permettre d'utiliser le RiboMethSeq pour d'autres applications.

Quantification des snoARN à partir des données de RiboMethSeq. La variation du niveau de 2'Ome est un phénomène commun à différents organismes et survient dans diverses conditions physiopathologiques. Toutefois,

l'origine de cette variabilité est peu connue et peu étudiée. Une des hypothèses serait que l'altération de l'expression des snoARN qui guident la 2'Ome des ARNr serait à l'origine de la variabilité du niveau de 2'Ome. Comme discuté dans la partie [C.2.3](#), l'expression de plusieurs snoARN à boîtes C/D est dérégulée dans les cancers. Il serait intéressant de déterminer s'il existe une corrélation entre l'expression des snoARN et le niveau de 2'Ome. Cela permettrait de mettre en évidence les sites dont la 2'Ome dépend directement de l'expression du snoARN associé.

Nous avons commencé des analyses afin de répondre à cette question en analysant l'expression des snoARN à l'aide des données issues de RiboMethSeq. En effet, l'objectif principal de l'approche RiboMethSeq est de quantifier le niveau de 2'Ome des ARNr, mais l'approche se base sur un séquençage RNA-seq de l'ARN total, qui contient, en plus des ARNr, des petits ARN non codant comme les snoARN. Des études reposant sur le RiboMethSeq ont en effet déjà utilisé cette approche (SHARMA et al., [2017](#)). Souvent, ces analyses restent discutables, car reposant sur une faible profondeur de lecture, sur une annotation des snoARN non-explicitée et sans validation par une autre méthode.

En collaboration avec l'équipe de Michelle Scott de l'université de Sherbrooke (Canada) spécialisée dans la quantification des snoARN et leurs annotations dans le génome (DESCHAMPS-FRANCOEUR et al., [2019](#); FAFARD-COUTURE et al., [2021](#)), nous avons quantifié l'abondance des snoARN dans les données de RiboMethSeq de l'article III ([D](#)). Ces données ont été séquencées en Illumina NovaSeq et chaque échantillon a une profondeur de séquençage très importante. Les données d'abondance des snoARN semblent être de bonne qualité, peuvent être utilisées pour une analyse différentielle et semblent cohérentes avec les données de la littérature. Par exemple, nous avons identifié SNORD115 comme étant sous-exprimé dans les échantillons de glioblastomes comparé aux échantillons non tumoraux (données non publiées). Or, SNORD115 est connu comme étant sous-exprimé dans les glioblastomes (JHA et al., [2015](#); VELÁZQUEZ-FLORES et al., [2020](#)). Pour valider ces résultats issus du RiboMethSeq, nous avons quantifié l'expression de 12 snoARN avec une autre méthode (RT-qPCR) (Fig : [6.2](#)). Les résultats montrent une bonne corrélation des valeurs absolues issues de la quantification par ces deux méthodes, à l'exception de trois snoARN (SNORD114, SNORD119, SNORD127) (Fig : [6.2](#)). Ces validations démontrent que les données de séquençage de

RiboMethSeq peuvent être utilisées pour la quantification des snoARN. Cependant, la quantification semble être limitée pour certains snoARN.

Par la suite, nous nous sommes intéressés à la corrélation entre l'expression des snoARN et le taux de 2'Ome (Cscore). Comme dans de nombreuses études, nous n'avons pas détecté de corrélation stricte entre le niveau d'expression des snoARN et le taux de 2'Ome (JAAFAR et al., 2021; KROGH, JANSSON et al., 2016; RAJAN et al., 2020; SHARMA et al., 2017). Il serait intéressant de comparer les données de RiboMethSeq avec une autre méthode de séquençage dédiée aux petits ARN non codants de type TGIRT-seq. L'ensemble de ces données montre que la régulation du niveau de 2'Ome des ARNr est un processus complexe, qui repose au moins en partie sur des variations d'expression des snoARN. Néanmoins, d'autres hypothèses restent à être explorées. On peut, par exemple, supposer qu'une différence d'expression ou d'activité d'un des composants protéiques du complexe snoRNP (FBL, NHP2L1, NOP56 ou NOP58) peut suffire à affecter la 2'Ome. En effet, de nombreuses études montrent une corrélation entre altération d'expression de FBL et altération de la 2'Ome (Wu2021; ERALES et al., 2017; Virginie MARCEL, GHAYAD et al., 2013). De même, l'altération d'interactant de ces protéines peuvent affecter la 2'Ome, par exemple EZH2 qui, en interagissant avec Nop56, empêche la formation du snoRNP et affecte ainsi la 2'Ome (YI et al., 2021). Ainsi, d'autres approches seront nécessaires pour mieux comprendre l'origine de la variation de 2'Ome des ARNr.

Analyse de la variation des séquences nucléotidiques des ARNr à partir des données de RiboMethSeq.

En plus de la quantification du niveau d'expression des snoARN, la séquence des lectures pourrait être utilisée pour analyser la variabilité des séquences nucléotidiques des ARNr. En effet, des variations nucléotidiques intra et inter-individus ont déjà été identifiées dans les ARNr chez la souris et l'Homme (PARKS et al., 2018).

Nous avons initié l'exploitation des données de séquençage générées par le RiboMethSeq. Pendant ma thèse, j'ai pu encadrer plusieurs étudiants en projet ou en stage dont l'objectif était d'identifier les variations nucléotidiques des ARNr à partir des données de RiboMethSeq. Grâce aux différents pipelines et approches mis en place (et qui sont toujours en train d'évoluer), nous

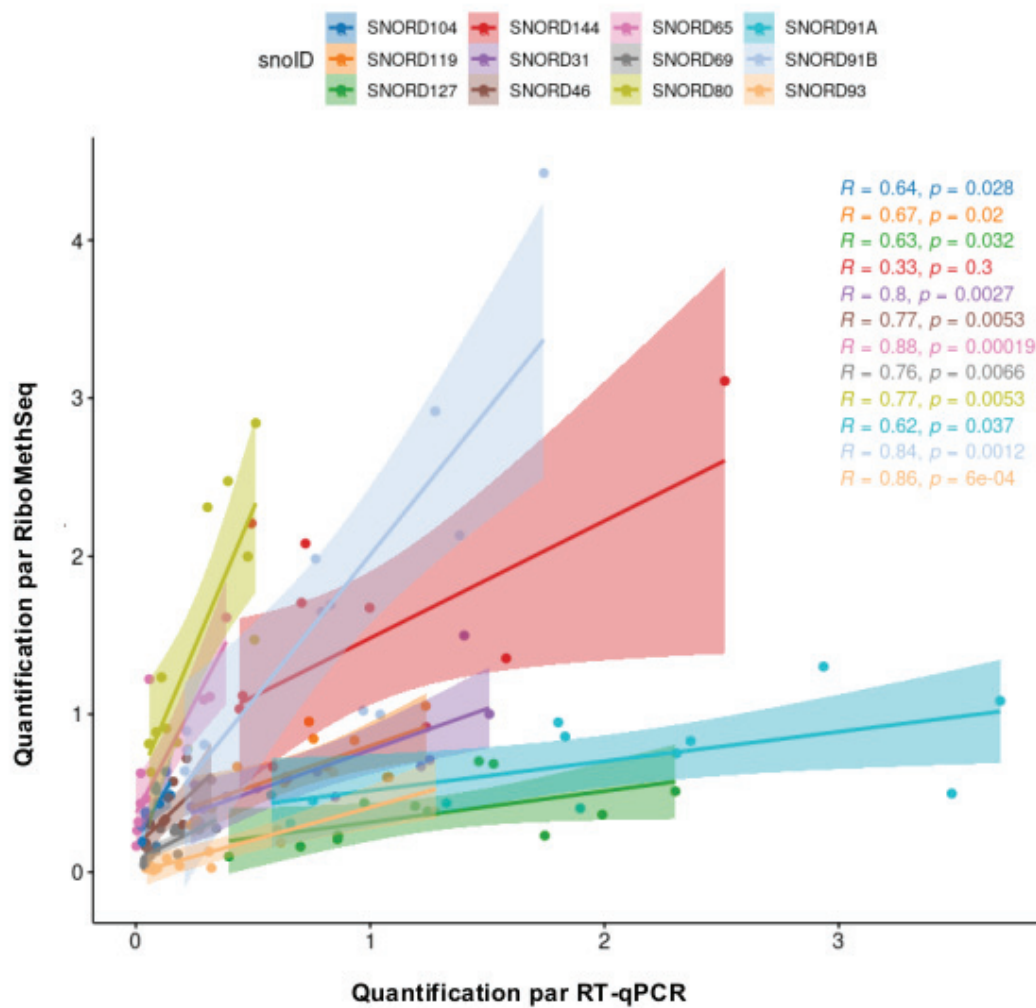


FIGURE 6.2

avons pu identifier des variations nucléotidiques par rapport à la séquence d'ARNr humaine de référence que nous utilisons.

Ces analyses nous permettront dans un premier temps d'améliorer la séquence de référence en incluant les variations qui sont communes à tous les échantillons analysés. Dans un second temps, ces analyses nous permettront d'identifier des variations nucléotidiques qui sont spécifiques à une condition biologique.

Les données de séquences issues du RiboMethSeq pourraient permettre d'aborder de nombreuses questions relatives à l'étude des ARNr. En revanche, ces nouvelles approches demanderont des validations pour démontrer leur utilisation. En plus de l'analyse de l'expression des snoARNs ou de la variation nucléotidique des ARN, de nombreuses études commencent à utiliser

les données de RiboMethSeq pour analyser la 2'Ome d'autres types d'ARN comme les ARNr mitochondriaux (SANCHEZ et al., 2020) ou les ARNt (GALVANIN et al., 2018).

B Hétérogénéité de la 2'Ome des ARNr dans les cancers

B.1 Démonstration de l'existence d'une hétérogénéité de la 2'Ome des ARNr dans les cancers

Avant le début de ma thèse, plusieurs études avaient montré qu'il existe une hétérogénéité de la 2'Ome dans des lignées cellulaires cancéreuses, mais aucune étude n'avait démontré cet aspect chez l'Homme (JAAFAR et al., 2021). Pour répondre à cette question, nous nous sommes tout d'abord concentrés sur les cancers du sein, qui présentent une forte hétérogénéité entre les patientes. En utilisant l'approche RiboMethSeq, nous avons analysé les profils de 2'Ome de 195 échantillons de tumeurs mammaires primaires (Virginie MARCEL, KIELBASSA et al., 2020). Nos résultats montrent que les molécules d'ARNr dans les cancers du sein ne sont pas toutes méthylées à certaines positions. Une variabilité du taux de 2'Ome est tout d'abord observée au sein d'une tumeur (variabilité intra-tumorale), comme cela avait déjà été montré dans les lignées cellulaires cancéreuses (JAAFAR et al., 2021 ; KROGH, JANSSON et al., 2016 ; SHARMA et al., 2017). La comparaison entre les tumeurs a aussi révélé une variabilité inter-tumorale de la 2'Ome des ARNr. Sur la base de cette variabilité inter-tumorale du taux de 2'Ome entre les patientes, nous avons identifié deux classes de sites 2'Ome d'ARNr. Les sites les plus « stables » présentent une faible variabilité inter-patientes bien que la cohorte analysée représente le spectre complet des sous-types de cancer du sein. Ces sites se situent proches des centres fonctionnels du ribosome et sont potentiellement essentiels au bon fonctionnement du ribosome. Les sites les plus « variables » présentent eux une forte variabilité inter-patientes. Etant donné les différences entre les méthodes de séquençage et d'analyse du RiboMethSeq, nous n'avons pas pu comparer de manière exhaustive ces données avec les sites identifiés comme variants dans les lignées cellulaires, bien qu'un grand nombre de sites semblent être identiques. De manière intéressante, nous avons identifié 4 sites 2'Ome « variables » (18S_Gm1447, 28S_Gm1303,

28S_Gm4588, and 18S_Am576) dont le niveau de méthylation est significativement différent entre les sous-types de cancer du sein, les statuts œstrogène et progestérone, ainsi que les grades tumoraux, suggérant que les sites les plus « variables » n'ont pas été identifiés par hasard, mais qu'ils semblent associés à des caractéristiques biologiques particulières. Dans notre deuxième étude (cf **D**), nous avons voulu déterminer si la 2'Ome des ARNr est altérée dans d'autres cancers et analysé une cohorte d'échantillons humains présentant les principaux sous-types de gliomes diffus de l'adulte. Les résultats confirment que certaines molécules d'ARNr peuvent exister sans 2'Ome sur certains sites particuliers et que la 2'Ome est altérée dans différents cancers. Par ailleurs, comme pour les cancers du sein, l'altération de la 2'Ome se fait sur des sites spécifiques qui permettent de distinguer les sous-types de gliomes. L'ensemble de ces données soutient l'hypothèse que la 2'Ome des ARNr reflète la variabilité inter-tumorale.

À ce jour, d'autres études ont analysé la variabilité de la 2'Ome des ARNr dans les cancers, toutes s'étant principalement focalisées sur l'étude des cancers hématologiques. Une étude réalisée à l'aide de l'approche RiboMethSeq couplée au séquençage Ion Torrent sur une cohorte de 17 échantillons de lymphomes diffus à grandes cellules B montrent des résultats similaires avec environ un tiers des sites présentant une variabilité de la 2'Ome (KROGH, ASMAR et al., 2020). Plus récemment, une étude a montré que la 2'Ome des ARNr variait dans les leucémies aiguës myéloïdes (ZHOU et al., 2022). En raison des différences entre les protocoles RiboMethSeq (plateforme de séquençage, effet batch, calcul du Cscore, identification des sites d'ARNr variables, etc.), il est difficile de comparer les valeurs absolues entre toutes les études. En revanche, toutes ces données soutiennent l'existence d'une hétérogénéité de 2'Ome des ARNr dans les cancers.

Au cours de ma thèse, pour répondre à la question de la spécificité de l'altération de la 2'Ome des ARNr dans les cancers sans utiliser les données de la littérature pour éviter des biais techniques, j'ai analysé d'autres jeux de données provenant d'autres types de cancer comme la leucémie aiguë myéloïde (LAM) ou le lymphome à cellules B, qui confirment encore une fois la variabilité de la 2'Ome dans les cancers (Fig : 6.3). J'ai appliqué notre méthodologie d'identification des sites dits « variants » et j'ai pu comparer leur identité entre les cancers. Parmi les sites d'ARNr qui peuvent tolérer l'absence de 2'Ome, seuls quelques-uns sont communs à tous les types de cancer

(n=6), la plupart étant spécifiques à un type cancer (Fig : 6.3). Les outils méthodologiques et bioinformatiques que j'ai développés au cours de ma thèse permettront de réaliser des méta-analyses afin d'identifier rigoureusement les sites qui acceptent l'absence de 2'Ome, qui sont communs et spécifiques à chaque cancers ou conditions physiopathologiques.

L'ensemble des données actuelles soutient que la 2'Ome est une des sources de l'hétérogénéité de composition du ribosome chez l'Homme. Cela semble être accepté par la communauté scientifique, notamment grâce à un nombre de publications croissant montrant une variabilité du niveau de 2'Ome dans différents contextes physiopathologiques et le développement d'approches innovantes comme le RiboMethSeq qui permettent de mesurer cette variabilité. Cependant, les rôles biologiques et fonctionnels de cette hétérogénéité restent encore soumis à débat.

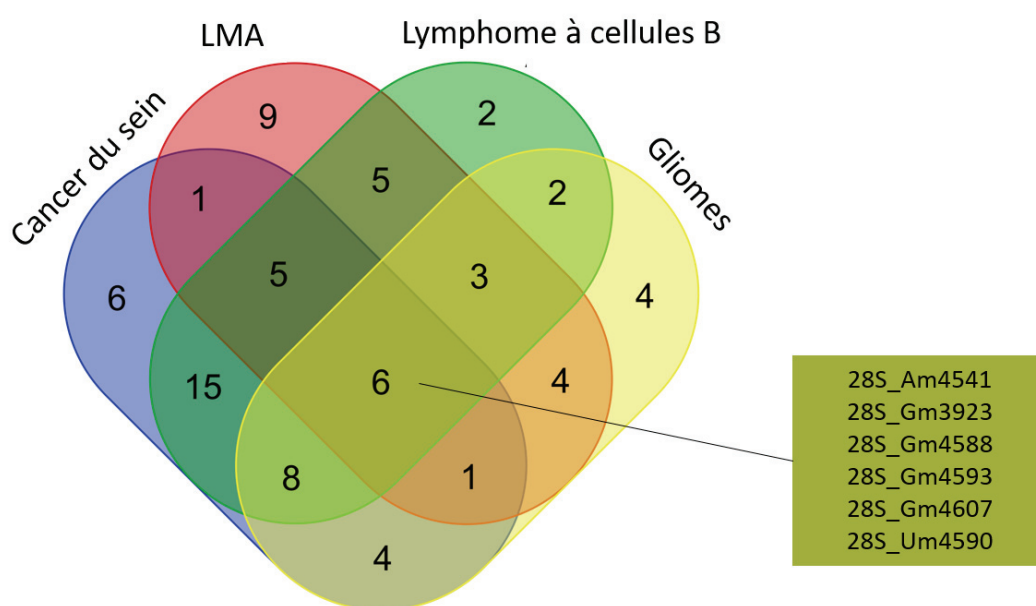


FIGURE 6.3 – Diagramme de Venn montrant quels sites 2'Ome les plus variants sont partagés entre les différents cancers

B.2 Quel rôle de la 2'Ome dans les fonctions biologiques ?

Pour essayer de donner un début de réponse à cette question, nous avons construit une troisième cohorte composée de 290 échantillons provenant de patientes atteintes de cancer du sein correspondant au sous-type le plus agressif, les tumeurs dites triple-négatives ou TNBC. Ces échantillons ont été

séquencés à la fois en RiboMethSeq et en RNA-seq avec l'objectif de confronter les variations de 2'Ome des ARNr avec des données d'expression, pour identifier des fonctions biologiques associées à des profils de 2'Ome des ARNr particuliers. Dans un premier temps, nous avons analysé les profils de 2'Ome dans les 290 échantillons. Comme montré dans les articles II (C) et III (D), l'information biologique est portée par les sites les plus variants. A l'aide de notre méthode d'identification des sites « variants » via l'analyse de l'IQR (Fig : 6.3), j'ai identifié 24 sites « variants » dans cette cohorte et utilisé leur taux de 2'Ome pour créer quatre groupes d'échantillons ayant des profils similaires à l'aide d'algorithmes de clusterisation (Fig : 6.5). À partir de ces quatre groupes, j'ai commencé à analyser les profils d'expression géniques. L'analyse d'enrichissement en groupes de gènes (GSEA, Gene Set Enrichment Analysis) a montré qu'il existe des processus biologiques spécifiques à chacun des groupes identifiés sur la base de 2'Ome des ARNr. Par exemple, selon les signatures Hallmark, les processus biologiques liés à la transition épithélio-mésenchymateuse (EMT) ou à la myogenèse sont enrichis dans le deuxième groupe (cluster) par rapport aux autres groupes (Fig : 6.6). Ces résultats suggèrent que la 2'Ome des ARNr peut être associée à des fonctions biologiques particulières. Des analyses supplémentaires sont nécessaires pour mieux exploiter ces données, d'autant plus qu'aujourd'hui, aucune étude n'a comparé le taux de 2'Ome et l'expression génique dans une cohorte d'échantillons humains.

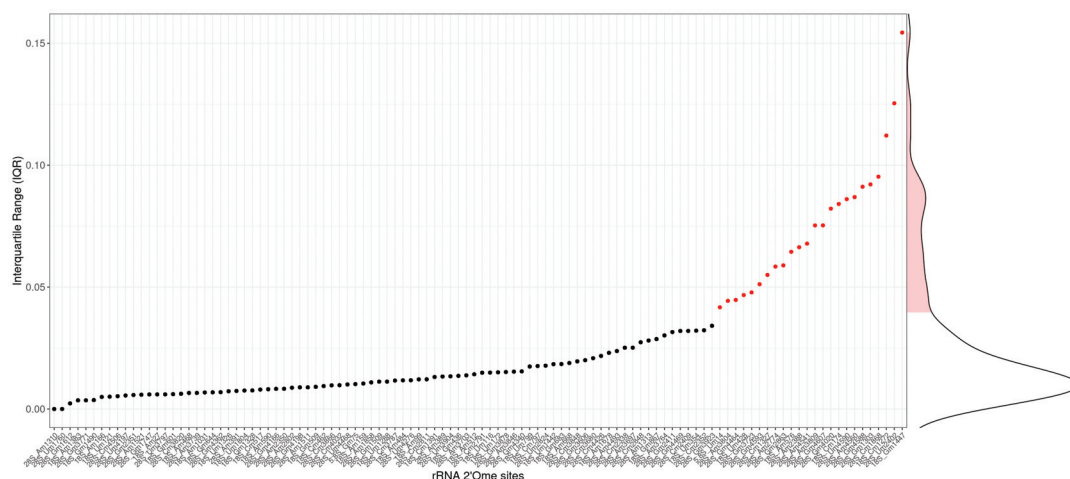


FIGURE 6.4 – Variabilité de la 2'Ome dans les cancers du sein TNBC. Les sites sont ordonnés en ordre croissant de IQR. Les points colorés en rouge représentent les sites les plus variants.

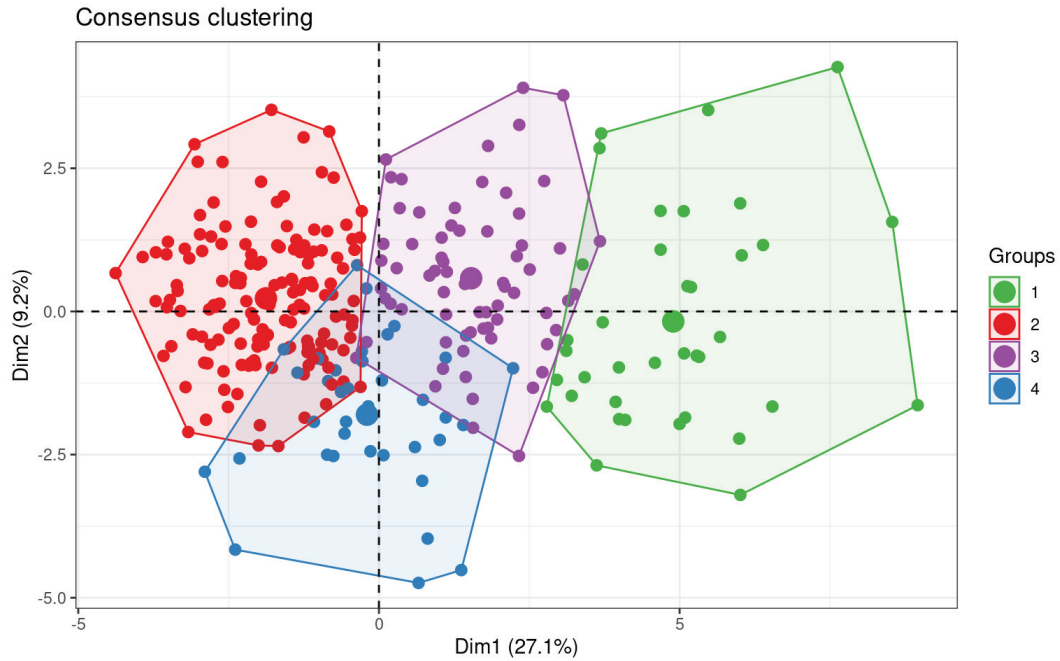


FIGURE 6.5 – Clusterisation des échantillons TNBC sur la base des profils de 2'Ome. L'ACP basée sur les Cscores des 24 sites les plus variants est visualisée. Les échantillons sont colorés selon les groupes de clusterisation

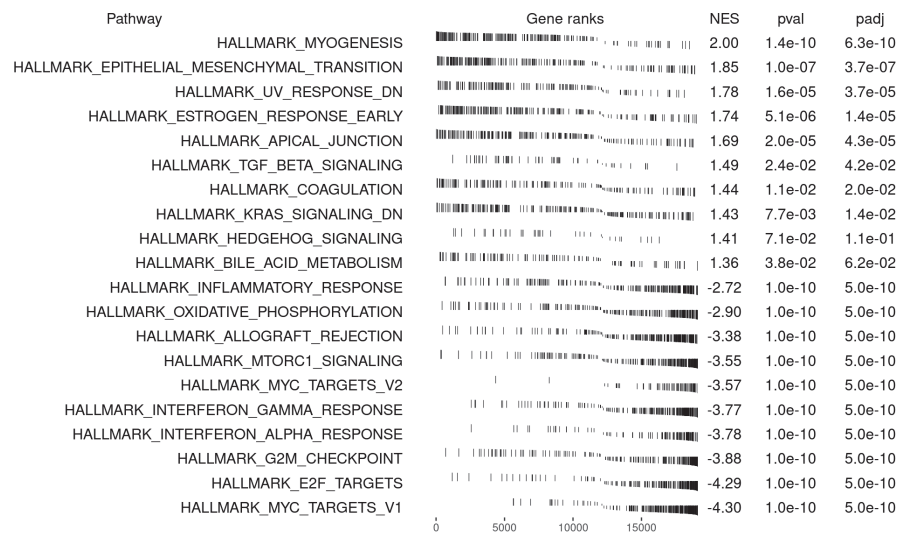


FIGURE 6.6 – Analyse de GSEA. Visualisation de l'analyse GSEA qui compare les échantillons du deuxième groupe (cluster) par rapport à tout le reste. Les voies biologiques Hallmark les plus significatives sont montrées.

Cette première analyse a été basée sur le profil des sites les plus variants. Cependant, certaines données suggèrent que même l'altération d'un seul site

2'Ome peut avoir un impact biologique important. Comme nous l'avons montré dans les articles II (C) et III (D), le taux de 2'Ome de certains sites seulement est associé à des caractéristiques biologiques et cliniques dans les cancers. Le site 18S_Am576 présente une augmentation du taux de 2'Ome dans les sous-types les plus agressifs de gliome (glioblastome multiforme ou GBM) et de cancer du sein (triple négatif ou TNBC), tandis que le site 18S_Gm1447 présente une diminution du taux de 2'Ome dans ces mêmes sous-types. Cela suggère un lien entre ces sites et l'agressivité du cancer. Dans une autre étude plus récente, Jansson *et al* ont montré que la surexpression de l'oncogène cMyc entraîne une altération spécifique de la 2'Ome des ARNr dans des cellules humaines (JANSSON *et al.*, 2021). Cette altération touche particulièrement le site 18S_Cm174. La perte de la méthylation au site 18S_Cm174, induite par une délétion par CRISPR-Cas9 du snoARN (SNORD45C) responsable de sa méthylation, n'affecte ni la synthèse protéique globale, ni la biogenèse des ribosomes. En revanche, elle induit une diminution le taux de prolifération des cellules. Cette étude montre pour la première fois que l'altération d'un seul site de 2'Ome de l'ARNr peut affecter la prolifération d'une lignée cellulaire humaine, la prolifération étant une des caractéristiques de l'agressivité du cancer. D'autres études seront nécessaires pour démontrer le rôle de la 2'Ome des ARNr dans l'acquisition de fonction biologique particulière.

B.3 Quel rôle de la 2'Ome dans la traduction ?

Depuis le début de ma thèse, plusieurs études ont identifié des altérations de la traduction suite à une altération de la 2'Ome des ARNr dans des contextes physiopathologiques variés. Par exemple, Wang *et al* ont identifié une altération du niveau de la 2'Ome sur 12 sites des ARNr suite à l'expression de la protéine Nucléocapside (N) du virus SARS-CoV-2 dans la lignée cellulaire humaine HEK293T (H. WANG *et al.*, 2022). Cette étude a montré que la protéine N augmente la synthèse protéique globale en améliorant l'efficacité de la traduction des ribosomes, en particulier pour les ARNm viraux. Le lien direct entre 2'Ome des ARNr et impact sur la traduction reste à démontrer. Dans une autre étude, Jansson *et al* ont montré que l'absence de la méthylation au site 18S_Cm174 induite par la délétion de son snoARN associé altère non seulement la prolifération cellulaire, mais aussi la traduction de certains ARNm. En particulier, les ARNm enrichis en codons composés

des nucléotides AU sont traduits moins efficacement en l'absence de 2'Ome au site 18S_Cm174, tandis que les ARNm enrichis en codons riches en GC sont traduits plus efficacement. À large échelle, l'absence de la 2'Ome au site 18S_Cm174 réduit notamment la traduction d'ARNm codant des protéines impliquées dans la prolifération cellulaire. En revanche, le mécanisme par lequel l'absence de 2'Ome au site 18S_Cm174 affecte l'activité du ribosome et ainsi la traduction reste inconnu.

Malgré un manque de démonstration d'un lien direct, ces études soutiennent que la 2'Ome des ARNr pourrait jouer un rôle dans la régulation de la traduction en participant à la génération de ribosomes de composition différente. De nombreuses questions restent actuellement ouvertes. Quels sites ont un impact sur la traduction ? Est-ce l'absence d'une seule méthylation ou une combinaison à plusieurs sites qui affecte l'activité du ribosome ? Pour répondre à cette dernière question, des études récentes ont commencé à développer des approches de type « single molecule » grâce à des techniques de séquençage de type nanopore (BEGIK et al., 2021 ; SMITH et al., 2019 ; STEPHENSON et al., 2022). Ces technologies récentes nécessitent le développement de nouvelles méthodes et d'outils bioinformatiques, mais permettront de déterminer les combinaisons de 2'Ome existant sur une même molécule d'ARNr. De plus, elles permettront de donner une vue plus globale des altérations de composition des ARNr, induites par la 2'Ome mais aussi par les autres modifications chimiques des ARNr, voire par les variations nucléotidiques des ARNr. En effet, l'hétérogénéité de composition du ribosome est certainement le résultat de plusieurs types de changements dans les ARNr et les protéines qui le composent. Identifier toute la variabilité des ribosomes dans les tissus requerra certainement des approches intégrées omics.

Ainsi, malgré le nombre d'études croissant rapportant l'existence d'une hétérogénéité de composition du ribosome dans différents contextes biologiques, le mécanisme par lequel ces altérations de composition des ribosomes affectent la traduction reste peu connu. De plus, il faut noter que l'addition des modifications chimiques des ARNr a lieu principalement au cours ou peu après la transcription des ARNr et est irréversible. Ce qui signifie que la régulation de la traduction par les modifications chimiques des ARNr n'est certainement pas impliquée dans une réponse immédiate à un stress mais plutôt dans un contexte bien établi.

7 Conclusion générale

Au cours de ma thèse, j'ai contribué au développement et à l'amélioration de deux outils bioinformatiques qui permettront à la communauté scientifique d'étudier et d'analyser la 2'Ome des ARNr. Tout d'abord des outils sous forme de pipeline bioinformatique et de package R dédiés à l'approche de RiboMethSeq qui permettent de traiter les données de séquençage, de calculer le taux de 2'Ome sur tous les sites d'intérêt, d'explorer les données et d'identifier les altérations de la 2'Ome au niveau des profils ou de sites spécifiques. Dans un deuxième temps, d'un outil de base de données (snoDB) disponible gratuitement en ligne et qui contient des informations précises sur les snoARN qui guident la 2'Ome, leurs expressions dans différentes lignées cellulaires et de tissus, leurs interactions canoniques et non-canoniques, la visualisation des séquences d'ARNr et des données de niveau de 2'Ome.

Grâce à ces outils, j'ai pu montrer pour la première fois que la 2'Ome des ARNr varie chez l'Homme et plus particulièrement dans les cancers. Ces variations de niveau de 2'Ome des ARNr dans les cancers sont associées à des caractéristiques biologiques et cliniques distinctes. De plus, les données issues de ma thèse ont permis d'identifier des sites de 2'Ome d'intérêt en cancérologie.

Ces résultats soutiennent l'hypothèse de l'existence d'une hétérogénéité de composition du ribosome qui pourrait avoir des conséquences biologiques et fonctionnelles.

Aujourd'hui, des études scientifiques interdisciplinaires à l'interface de la physique, biologie, biochimie et bio-informatique, ainsi que le développement de nouvelles technologies haut-débit afin de mieux identifier l'hétérogénéité de composition du ribosome à ses différents niveaux, reste une étape clé pour mieux comprendre le rôle des modifications chimiques des ARNr. Le mécanisme exact par lequel l'hétérogénéité de composition du ribosome est impliquée dans la régulation traductionnelle deviendra plus claire au fur et à mesure que d'autres exemples d'altération de composition du ribosome

associés à différentes conditions biologiques seront identifiés.

Des approches innovantes seront nécessaires, pour séparer les sous-populations ribosomiques sur une base structurale avant de les tester sur le plan fonctionnel. L'élucidation des rôles structuraux et fonctionnels des résidus modifiés au sein du ribosome contribuera sans aucun doute au développement d'une nouvelle génération de thérapies anticancéreuses axées sur le ciblage des ribosomes présentant des schémas de modifications chimiques différenciant dans les cancers et apportera un nouvel éclairage significatif sur le rôle de ces modifications chimiques de l'ARNr connues depuis des décennies dans l'orientation de l'activité du ribosome et de l'activité cellulaire.

A Annexe

A Liste des publications, communications orales, posters et collaborations

A.1 Publications acceptées

1. Jaafar, M., **Paraqindes, H.**, Gabut, M., Diaz, J.-J., Marcel, V., Durand, S. (2021). *2O-Ribose Methylation of Ribosomal RNAs : Natural Diversity in Living Organisms, Biological Processes, and Diseases*. In *Cells* (Vol. 10, Issue 8, p. 1948). MDPI AG. <https://doi.org/10.3390/cells10081948>
2. Bergeron, D., **Paraqindes, H.**, Fafard-Couture, É., Deschamps-Francoeur, G., Faucher-Giguère, L., Bouchard-Bourelle, P., Abou Elela, S., Catez, F., Marcel, V., Scott, M. S. (2022). *snoDB 2.0 : an enhanced interactive database, specializing in human snoRNAs*. In *Nucleic Acids Research*. Oxford University Press (OUP). <https://doi.org/10.1093/nar/gkac835>
3. Marcel, V., Kielbassa, J., Marchand, V., Natchiar, K. S., **Paraqindes, H.**, Van Long, F. N., Ayadi, L., Bourguignon-Igel, V., Monaco, P. L., Monchiet, D., Scott, V., Tonon, L., Bray, S. E., Diot, A., Jordan, L. B., Thompson, A. M., Bourdon, J.-C., Dubois, T., André, F., ... Diaz, J.-J. (2021). *Ribosomal RNA 2 O-methylation as a novel layer of inter-tumour heterogeneity in breast cancer*. In *NAR Cancer* (Vol. 3, Issue 1). Oxford University Press (OUP). <https://doi.org/10.1093/narcan/zcab006>
4. Berthel, E., Vincent, A., Eberst, L., Torres, A. G., Dacheux, E., Rey, C., Marcel, V., **Paraqindes, H.**, Lachuer, J., Catez, F., de Pouplana, L. R., Treilleux, I., Diaz, J.-J., Dalla Venezia, N. (2020). *Uncovering the Translational Regulatory Activity of the Tumor Suppressor BRCA1*. In *Cells* (Vol. 9, Issue 4, p. 941). MDPI AG. <https://doi.org/10.3390/cells9040941>

A.2 Publications en préparation

1. **Paraqindes, H.**, Mourksi N., Ballesta S., Bourdelais, F., Thomas, E., Kielbassa J., Perrossier J., Carrere, M., Attignon V., Tonon L., Viari A., Diaz, JJ., Meyronet, D., Gabut, M., Ducray, F., Marcel, V., Durand, S. *Ribosome biogenesis and rRNA 2'O-ribose methylation : Uncovering specific alterations in glioma subtypes*
2. Combe, T., **Paraqindes, H.**, Kielbassa, J., Thomas, E., Ferrari, A., Marcel, V., *Riboscore : an R package for RiboMethSeq data analysis*

A.3 Communications orales

Mars 2022 - Collaboration avec l'équipe Scott, Sherbrooke, Canada :

- (i) Quantification of 2'O-ribose methylation of ribosomal RNAs by RiboMethSeq
- (ii) Alteration of ribosome biogenesis and 2'O-ribose methylation of ribosomal RNAs in gliomas
- (iii) 2'O-ribose methylation ribosomal RNAs reflects tumor heterogeneity in breast cancers

Avril 2022 - Séminaire de département « Initiation du Cancer et Identité de la cellule Tumorale », Lyon, France :

- (i) Alteration of ribosome biogenesis and 2'O-ribose methylation of ribosomal RNAs in gliomas

Juillet 2022 - Presentation au congrès : Les Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM), Rennes, France

- (i) Uncovering specific alterations of ribosomal RNA 2'O-ribose methylation in glioma subtypes using the latest improvements of RiboMethSeq technology

A.4 Posters

1. Juin 2021 - **Paraqindes H**, Marchand V, Nathiar KS, Bourdon JC, André F, Klaholz BP, Viari A, Diaz JJ, Kielbassa J, Marcel V. *Ribosomal RNA 2'O-methylation as a novel layer of inter-tumour heterogeneity in breast cancer. EACR 2021 virtual congress : Innovative Cancer Science*

2. Juillet 2022 - Combe, T., **Paraqindes, H.**, Kielbassa, J., Thomas, E., Ferrari, A., Marcel, V., *RiboMethSeq platform at CRCL/CLB to profile ribosomal RNA 2'-O-ribose methylation*, **JOBIM 2022**, Rennes, France
3. **Paraqindes, H.**, Mourksi N., Ballesta S., Bourdelais, F., Thomas, E., Kielbassa J., Perrossier J., Carrere, M., Attignon V., Tonon L., Viari A., Diaz, JJ., Meyronet, D., Gabut, M., Ducray, F., Marcel, V., Durand, *Ribosome biogenesis and rRNA 2'-O-ribose methylation : Uncovering specific alterations in glioma subtypes*, **CRCL Symposium 2022**, Lyon, France
4. Novembre 2022 - **Paraqindes, H.**, Kielbassa, J., Thomas, E., Monchiet, D., Ardin, M., Tonon, L., Combe, T., Perrossier J., Carrere, M., Attignon V., Caux, C., Viari, A., Diaz, JJ., Marcel, V. *Ribosomal RNA 2'-O-ribose methylation brings a novel layer of molecular heterogeneity in Triple-Negative Breast Cancer*, **CRCL Symposium 2022**, Lyon, France

B Article : Uncovering the Translational Regulatory Activity of the Tumor Suppressor BRCA1

Ce travail dont je suis auteur en 9^{ème} position a été publié en avril 2020 dans le journal Cells. L'article et les figures supplémentaires sont inclus après cette page introductive.

B.1 Contexte

Breast Cancer 1 (BRCA1) est un gène suppresseur de tumeurs. Des mutations sporadiques et germinales dans ce gène conduisent à une inactivation de la protéine et à un risque accru de développer des cancers chez la femme, en particulier le cancer du sein. En effet la protéine sauvage BRCA1 assure plusieurs fonctions essentielles dans la cellule comme la réparation de l'ADN, la régulation de la transcription, le contrôle du cycle cellulaire ou encore l'apoptose. Les mutations du gène BRCA1 peuvent ainsi rendre la protéine peu ou non fonctionnelle en limitant ses capacités de suppresseur de tumeurs. Deux études montrent que la protéine BRCA1, en plus des fonctions décrites précédemment, est impliquée dans la régulation traductionnelle.

L'objectif de cette étude est de déterminer si le contrôle de la traduction pourrait être une nouvelle fonction de BRCA1 contribuant à ses capacités de suppression des tumeurs.

B.2 Résultats principaux

Nous avons identifié les ARNm associés à la protéine BRCA1 par immunoprécipitation des ARNm couplée à une analyse des ADNc par puces à ADN (n=488). De plus, nous avons démontré que les ARNm associés par BRCA1 sont différentiellement régulés en réponse à une modulation de l'expression de BRCA1 par une analyse de profil des polysomes. Une étude comparative RTqPCR et Western blot nous a permis de confirmer une dérégulation au niveau protéique sans altération au niveau ARNm, suggérant que la traduction d'un sous-ensemble d'ARNm est dérégulée en cas de déficience de BRCA1 dans les lignées cellulaires. Ces résultats ont été confirmés dans l'analyse d'une cohorte de 45 échantillons (26 patients ne présentent pas de mutations BRCA1 et 19 présentent des mutations BRCA1) une association

entre mutation BRCA1 et surexpression de ces protéines d'intérêt. Parmi ces ARNm différenciellement traduits, on peut mentionner ADAT2 qui est ici proposé comme biomarqueur de la présence d'une mutation BRCA1, dont la présence dans les tumeurs repose uniquement sur du séquençage difficilement utilisable en routine en clinique (matériel biologique, coût, ...).

En conclusion, cette étude montre que BRCA1 modifie la traduction d'un sous ensemble d'ARNm impliqués dans les principaux programmes induisant le développement tumoral.










B.3 Ma contribution

J'ai réalisé une analyse de recherche de motifs dans les ARNm associés à la protéine BRCA1. A partir du symbole des 488 gènes d'intérêt, j'ai récupéré leurs identifiants Ensembl (European Bioinformatics Institute and the Wellcome Trust Sanger Institute) à l'aide de la librairie R *ensemldb*. Cela m'a permis de télécharger leurs séquences via REST API de la base de données Ensembl. Par la suite, j'ai effectué l'analyse de recherche de motifs en utilisant le logiciel DREME. J'ai pu identifier 14 motifs communs à au moins 333 des 488 gènes avec des valeurs statistiques significatives. Parmi ces motifs, seulement deux sont connus dans la littérature comme étant associées à des protéines de liaison à l'ARN (Figure S2).

B.4 Article

Article

Uncovering the Translational Regulatory Activity of the Tumor Suppressor BRCA1

Elise Berthel ¹, Anne Vincent ¹, Lauriane Eberst ², Adrian Gabriel Torres ³, Estelle Dacheux ¹, Catherine Rey ⁴, Virginie Marcel ¹, Hermes Paraqindes ¹, Joël Lachuer ⁴, Frédéric Catez ¹, Lluís Ribas de Pouplana ^{3,5}, Isabelle Treilleux ⁶, Jean-Jacques Diaz ^{1,†} and Nicole Dalla Venezia ^{1,†,*}

¹ Inserm U1052, CNRS UMR5286, Centre de Recherche en Cancérologie de Lyon, Université de Lyon 1, Centre Léon Bérard, F-69008 Lyon, France; elise.berthel@inserm.fr (E.B.) ; anne.vincent@lyon.unicancer.fr (A.V.) ; estelle.dacheux@pasteur.fr (E.D.) ; Virginie.MARCEL@lyon.unicancer.fr (V.M.) ; Hermes.PARAQINDES@lyon.unicancer.fr (H.P.) ; frederic.catez@lyon.unicancer.fr (F.C.) ; jeanjacques.diaz@lyon.unicancer.fr (J.-J.D.)

² Centre Léon Bérard, Medical Oncology Department, Université de Lyon 1, F-69008 Lyon, France; lauriane.eberst@lyon.unicancer.fr

³ Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac, 10, 08028 Barcelona, Spain; adriangabriel.torres@irbbarcelona.org (A.G.T.); lluis.ribas@irbbarcelona.org (L.R.d.P.)

⁴ ProfileXpert, UNIV-US7 INSERM-UMS 3453 CNRS, F-69000 Lyon, France; catherine.rey@inserm.fr (C.R.); joel.lachuer@univ-lyon1.fr (J.L.)

⁵ Catalan Institution for Research and Advanced Studies (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain

⁶ Department of Translational Research and Innovation, Centre Léon Bérard, F-69008 Lyon, France; isabelle.treilleux@lyon.unicancer.fr

* Correspondence: nicole.dalla-venezia@lyon.unicancer.fr; Tel.: +33-426-556-745

† These authors contributed equally.

Received: 6 February 2020; Accepted: 4 April 2020; Published: 10 April 2020



Abstract: BRCA1 inactivation is a hallmark of familial breast cancer, often associated with aggressive triple negative breast cancers. BRCA1 is a tumor suppressor with known functions in DNA repair, transcription regulation, cell cycle control, and apoptosis. In the present study, we demonstrate that BRCA1 is also a translational regulator. We previously showed that BRCA1 was implicated in translation regulation. Here, we asked whether translational control could be a novel function of BRCA1 that contributes to its tumor suppressive activity. A combination of RNA-binding protein immunoprecipitation, microarray analysis, and polysome profiling, was used to identify the mRNAs that were specifically deregulated under BRCA1 deficiency. Western blot analysis allowed us to confirm at the protein level the deregulated translation of a subset of mRNAs. A unique and dedicated cohort of patients with documented germ-line BRCA1 pathogenic variant statuses was set up, and tissue microarrays with the biopsies of these patients were constructed and analyzed by immunohistochemistry for their content in each candidate protein. Here, we show that BRCA1 translationally regulates a subset of mRNAs with which it associates. These mRNAs code for proteins involved in major programs in cancer. Accordingly, the level of these key proteins is correlated with BRCA1 status in breast cancer cell lines and in patient breast tumors. ADAT2, one of these key proteins, is proposed as a predictive biomarker of efficacy of treatments recently recommended to patients with BRCA1 deficiency. This study proposes that translational control may represent a novel molecular mechanism with potential clinical impact through which BRCA1 is a tumor suppressor.

Keywords: BRCA1; translation; breast cancer; tumor suppressor

1. Introduction

Breast cancer is the most common cancer worldwide, accounting for 2.1 million new cases and 627,000 deaths worldwide in 2018 [1]. Approximately 5% of these patients carry a germline pathogenic variant in BRCA [2–4] with a risk of developing breast cancer reaching 72% by age of 80 years [5,6]. Additionally, in a significant proportion of sporadic breast tumors, Breast Cancer 1 (BRCA1) promoter hypermethylation, transcription repression, or somatic BRCA1 pathogenic variants are responsible for its inactivation [7,8]. BRCA1 pathogenic variant carriers are more likely to be diagnosed with a triple negative breast cancer (i.e., estrogen-receptor negative, progesterone-receptor negative, and human epidermal growth factor receptor 2 (HER2) negative), which is associated with poor prognosis due to limited therapeutic options [9]. The identification of BRCA1 pathogenic variant carriers is of the most importance since these patients can now benefit from preventive treatments (early detection of contra-lateral breast cancer, prophylactic mastectomy, or oophorectomy) and most importantly from innovative treatments, including poly (adenosine diphosphate-ribose) polymerase (PARP) inhibitors like olaparib [10].

The BRCA1 protein is a tumor suppressor. It maintains genomic integrity through multiple functions; the best documented being those related to DNA repair [11–14]. However, other functions participate in its tumor suppressive capacity including chromatin remodeling, cell cycle regulation, transcription regulation, mRNA splicing, and apoptosis [11,15,16]. Furthermore, we previously showed that BRCA1 was implicated in translational regulation [17,18]. Whether translational control could be a novel function of BRCA1 contributing to its tumor suppressor capabilities remains however to be assessed.

Translational control is a master regulator of gene expression and its deregulation is a hallmark of cancer. Here, we report translational regulation as a novel pivotal function of BRCA1 that may represent a novel mechanism through which BRCA1 exerts its tumor suppressive activity. BRCA1 protein controls the translation of mRNAs with which it interacts. Upon BRCA1 inactivation, Adenosine Deaminase, tRNA Specific 2 (ADAT2), Cyclin L2 (CCNL2), Protein DBF4 homolog B (DBF4B), and Tripartite Motif Containing 45 (TRIM45) mRNAs display altered translation efficiencies with concomitant changes of their protein levels. In mammary tumors from patients mutated for BRCA1, expression of these proteins is also impacted as a consequence of BRCA1 loss of function. Our study uncovers translational control as a novel BRCA1 function with potential clinical impact.

2. Materials and Methods

2.1. Chemicals

Refer to Supplementary Material and Methods for details.

2.2. Case Selection

Patients with clinical history suggestive of BRCA1 or BRCA2 pathogenic variants (i.e., predictive risk of pathogenic variant according to the Eisinger score superior to 3 [19] were referred to a geneticist, in order to detect a germline BRCA1 or BRCA2 pathogenic variant. Case selection is detailed in Supplementary Material and Methods.

2.3. Tissue Microarray (TMA) and Immunohistochemistry (IHC)

Breast tumors from each of the 3 groups (exploratory mutant (MT) BRCA1, exploratory wild-type (WT) BRCA1, and validation MT BRCA1) were inserted into 3 different tissue microarray (TMA) blocks, each tumor specimen being divided and inserted in triplicate. Four microns thick unstained sections were cut from each TMA block and mounted on slides. The slides were processed for immunohistochemistry (IHC) and immunostained as described in Supplementary Material and Methods.

The proportion of positively stained tumor cells in each section was graded as follow: 0, $\leq 10\%$; 1, 20–40%; 2, 50–80%; and 3, $> 80\%$ positive cells. The staining intensity was recorded on a scale comprised between 0 and 3, corresponding to null to high intensity of the signal. The scores were

calculated as follows: staining intensity \times proportion of positively stained cells. A Chi-square statistical analysis was conducted to test for the significant difference between cohort scores.

2.4. Cell Culture

Refer to Supplementary Material and Methods for details.

2.5. Transfection with Plasmids

For transfection, MCF-7 and MDA-MB-231 cells were plated at 2×10^6 cells per 10 cm diameter dish 24 h before transfection with 4 μ g of plasmid and 20 μ L of ExGen 500 (Euromedex, Souffelweyersheim, France). Twenty four hours after transfection, cells were harvested. The pCDNA3 plasmid expressing full length BRCA1 protein (BRCA1) was previously described [20].

2.6. RNA Interference (RNAi)

The siRNA duplexes were purchased from Eurofins Genomics (Ebersberg, Germany) and provided as purified and annealed duplexes. The sequence of the siRNA against BRCA1 was si-BRCA1 5'-GGAACCUGUCUCCACAAAG-3' [17,20]. The siRNA used as control was designed against luciferase: si-Ctrl 5'-CGUACGCGGAAUACUUCGA-3'.

MCF-7, MDA-MB-231, and HMECs were plated at 2×10^6 cells per 10 cm diameter dish 24 h before transfection. Cells were transfected with 200 pmol/dish of siRNA and 16 μ L/dish of Lipofectamine RNAiMAX (Thermo Fisher Scientific, Bourgoin-Jallieu, France) using the protocol provided by the supplier. Cells were harvested 72 h after transfection.

2.7. Immunoblotting

Immunoblotting was performed as previously reported [18]. Refer to Supplementary Material and Methods for details.

2.8. Isolation of Polysomes and Total Cytoplasmic RNA

Isolation of polysomes and total cytoplasmic RNA was carried out as previously [18]. Refer to Supplementary Material and Methods for details.

2.9. RNA-Binding Protein Immunoprecipitation (RIP)

RNA-binding protein immunoprecipitation (RIP) experiments were performed mainly using the standard reference protocols including the relevant non-specific control [21]. Refer to Supplementary Material and Methods for details.

2.10. Microarray Analysis

RNA isolation following immunoprecipitation with anti-BRCA1 antibody and control antibody (NR) was performed in triplicate as described in the "RNA-binding protein immunoprecipitation (RIP)" section.

Microarray processing and data analysis was performed at the ProfileXpert core facility (Lyon, France). Details are provided in Supplementary Material and Method.

The complete set of raw and normalized data is available at the GEO database under accession number GSE119886 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE119886>).

For each replicate, the probe intensities issued from the BRCA1 and NR immunoprecipitations were obtained. The obtained data were normalized with Affymetrix Expression Console software using the RMA statistical algorithm. Control expression values (NR) were subtracted from their respective sample counterparts on a probe set basis; the 3 replicates were then averaged. Only genes showing an immunoprecipitation (IP) ratio IP BRCA1/IP NR (Fc) greater than 1.5 and a *p*-value lower than 0.05 were retained.

The retained genes of interest were listed and classified according to their biological process (BP), cellular compartment (CC), and molecular function (MF) using DAVID (Database for Annotation, Visualization and Integrated Discovery) on 10th of March 2017 [22] (<https://david.ncifcrf.gov/summary.jsp>).

2.11. Quantitative RT-PCR

2.11.1. RIP Analysis

Three additional replicates of RIP, including inputs, were dedicated to RT-qPCR. RNA extracted from RIP (IP BRCA1 and IP NR) and RNA extracted from the corresponding inputs, was reverse-transcribed and pre-amplified with the Nugen Ovation WTA System (TECAN, Redwood City, USA). Quantitative real-time PCR (qPCR) was carried out using the light cycler 480 II real-time PCR thermocycler (Roche, Meylan, France). The experimental protocol consisted in an initial polymerase activation for 10 min at 95 °C followed by an amplification program of 10 sec at 95 °C, 10 sec at 60 °C, and 10 sec at 72 °C, for 45 cycles. Expression of mRNA was quantified using the LightCycler 480 SYBR Green I Master Mix (Roche) and normalized using POP4 and Actin expression according to the $2^{-\Delta\Delta C_t}$ method. Immunoprecipitated RNA was expressed as fold enrichment over input compared to enrichment with non-relevant antibody (NR) used as a control; fold enrichment $Fe = (IP\ BRCA1 - IP\ NR)/Input$.

2.11.2. Polysome Profile Analysis

Total and polysomal RNA prepared as described above, from BRCA1-depleted and control cells, were purified and submitted to DNase digestion using NucleoSpin RNA XS Clean-up columns (Macherey-Nagel, Hoerd, France) following the manufacturer's instructions. Next, 250 ng of total RNA were reverse transcribed using the M-MLV RT kit and random primers (Thermo Fisher Scientific, Bourgoin-Jallieu, France), according to the manufacturer's instructions. For each RNA type (polysome RNA or total RNA prepared from the same cytoplasmic extract as described above), a ratio between si-BRCA1 and si-Ctrl sample was calculated. The ratios for polysomal RNA (polyRNA = si-BRCA1/si-Ctrl) and for total RNA (totRNA = si-BRCA1/si-Ctrl) were calculated. The polyRNA/totRNA ratio was then calculated and defined as translational efficiency (Te). Only genes showing a Te greater than 1.4 or lower than -1.4 in the three replicates were retained.

Primers were designed using the Primer-Blast software (National Centre for Biotechnology Information/NCBI, Bethesda, USA; <http://www.ncbi.nlm.nih.gov/tools/primer-blast>) and purchased from Eurofins Genomics (Ebersberg, Germany). For each gene, primers were designed within exons that displayed high exon probe intensities on arrays. Genes tested for polysomal profiling and RIP, and primers, are listed in the Table S1.

2.12. Quantification of A-to-I Editing

Quantification of tRNA editing was determined as previously described [23] with minor modifications. Refer to Supplementary Material and Methods for details. The RNAseq data generated for this study has been deposited at the EBI European Nucleotide Archive (Accession number PRJEB28622).

2.13. Statistical Analysis

Refer to Supplementary Material and Methods for details.

3. Results

3.1. BRCA1 Associates with mRNAs

Having previously shown that BRCA1 depletion affects protein synthesis and translational efficiency [17,18], we here investigated which mRNA were associated with BRCA1 protein. We used an RNA binding protein immunoprecipitation (RIP) assay followed by microarray analysis of the BRCA1-bound mRNA. Three total lysates of MCF-7 cells were independently treated. Each lysate was

split in two equal portions. One was incubated with anti-BRCA1 antibody (BRCA1) while the second was incubated with control antibody (NR). We first confirmed efficient immunoprecipitation of BRCA1 by Western blot analysis (Figure 1a).

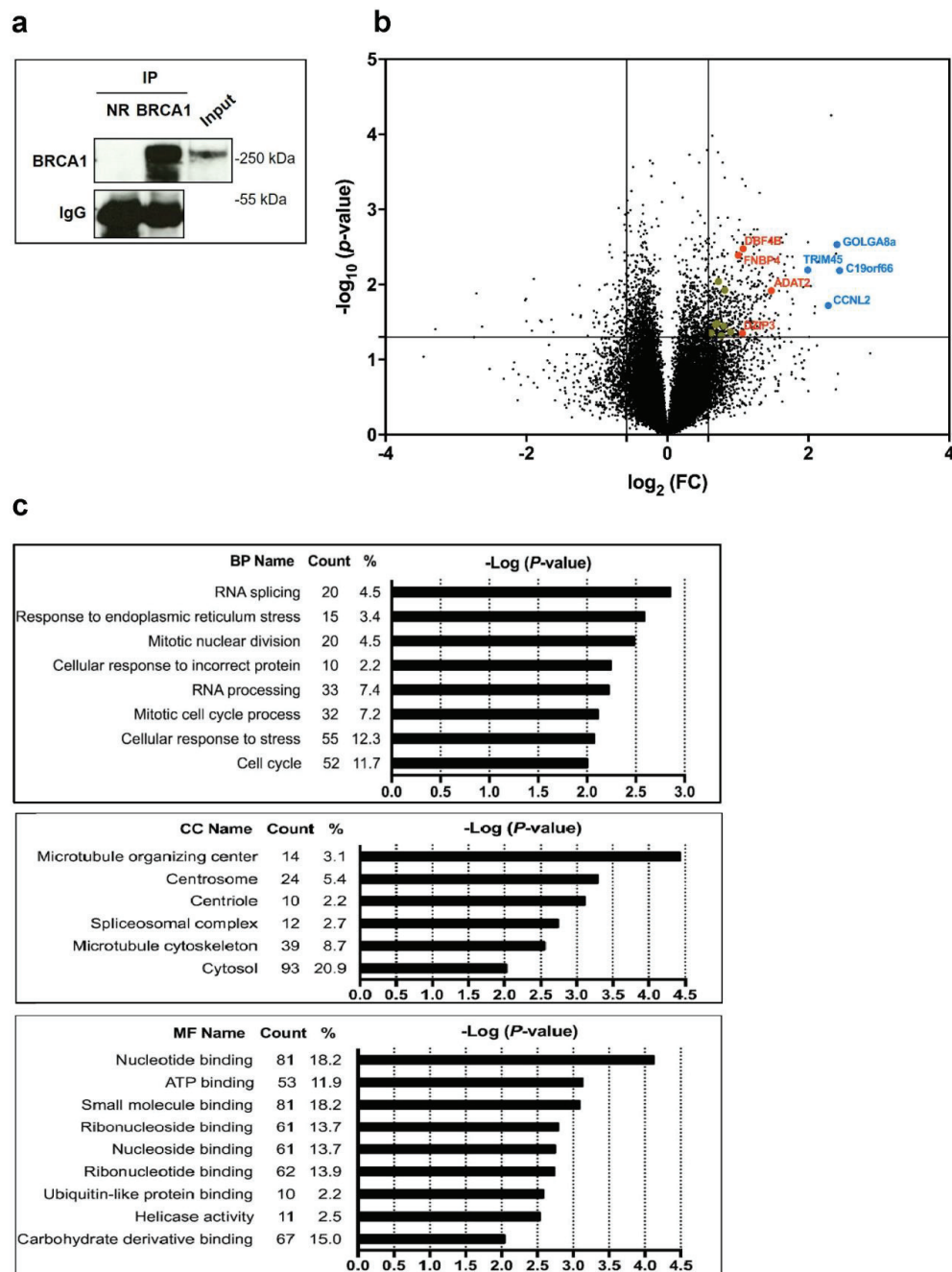


Figure 1. Identification of BRCA1-associated mRNAs using large-scale analysis. RNA-binding protein immunoprecipitation (RIP) performed in triplicate followed by microarray analysis. (a) Immunoblots showing that BRCA1 is immunoprecipitated in MCF-7 cells. Immunoprecipitation with anti-BRCA1 antibody (BRCA1) and with a control IgG (NR). (b) Microarray analysis of the BRCA1-associated mRNA isolated by RIP. Lines indicate $-\log_{10}(p\text{-value}) \geq 1.3$ and $abs(\log_2(\text{Fc})) \geq 0.6$ for $p \leq 0.05$ and $abs(\text{Fc}) \geq 1.5$ respectively. Among the mRNAs displaying a fold change (Fc) above 1.5, the 16 mRNA of interest (listed in Figure 2a) were colored in gray ($1.5 \leq \text{Fc} < 2.0$), orange ($2.0 \leq \text{Fc} < 3.0$), and blue ($3.0 \leq \text{Fc} \leq 5.4$). The name of the 8 genes with the most changes (colored in blue and orange) was included. (c) Gene ontology analysis of mRNA bound to BRCA1, using DAVID (Database for Annotation, Visualization and Integrated Discovery). BP: biological process; CC: cellular compartment; MF: molecular function.

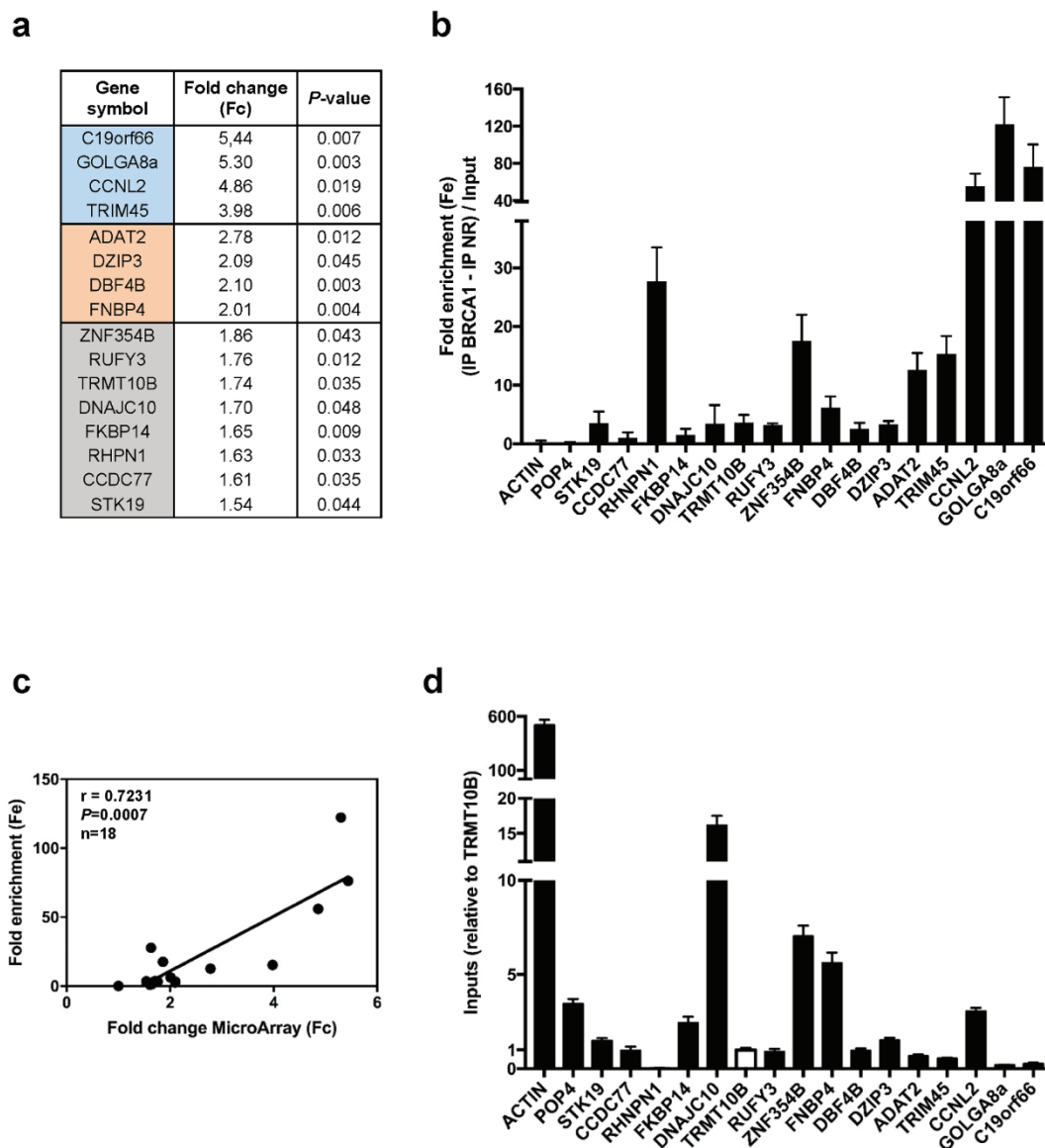


Figure 2. Analysis of a representative subset of BRCA1-associated mRNAs. RIP of three new independent replicates followed by RT-qPCR. (a) Subset of 16 BRCA1-associated mRNAs spanning the full range of Fc observed in Figure 1b. Colors indicate the range of Fc obtained by microarray. (b) Fold enrichment (Fe) of the 16 BRCA1-associated mRNAs determined by RT-qPCR. Data are expressed as means \pm SEM. $n = 3$. (c) Correlation between Fc calculated from the Affymetrix array and Fe measured by RT-qPCR. Mean RIP fold change obtained by microarray (Fc) and mean RIP fold enrichment obtained by RT-qPCR (Fe) were plotted and their correlation was assessed using the Spearman test. A significant correlation was observed. (d) Quantification by RT-qPCR of each BRCA1-associated mRNA in total extracts of MCF-7 cells (inputs). The TRMT10B value was arbitrarily set at 1 (white bar). Data are expressed as means \pm SEM. $n = 3$.

The microarray analysis was then performed for each replicate, and the probe intensities issued from the BRCA1 and NR immunoprecipitations were averaged. Thus, for each mRNA, the binding to control IgG and the binding to anti-BRCA1 antibody were measured in parallel. Only mRNAs showing an IP BRCA1/IP NR ratio (Fc) greater than 1.5 and a p -value lower than 0.05 were retained. Overall, among the 29591 mRNA detected on chip, 498 transcripts representing 2% of mRNAs expressed in MCF7 cells, were 1.5 fold more abundant in BRCA1 immunoprecipitates compared to control immunoprecipitates (Figure 1b). To provide a first view of the biological properties of the proteins

encoded by these mRNAs, we conducted a gene ontology analysis. The identification of the most enriched biological processes, molecular functions, and cellular compartments associated with these mRNAs, revealed that most of the enriched processes have previously been related to BRCA1 tumor suppressor activity such as stress response, RNA splicing, and cell cycle regulation. In addition, we noticed that the enriched cellular compartments and molecular function, namely cytosol and nucleotide binding, have been poorly associated with BRCA1 activity up to now (Figure 1c).

Next, we ascertained that the Fc observed for these mRNAs was correlated with their binding to BRCA1 and not with their cytoplasmic abundance. We selected a subset of 16 mRNAs spanning the full range of the Fc (Figure 2a).

We conducted three new independent RIP assays to measure the fold enrichment (Fe) of each mRNA normalized against its total abundance (input) in the whole cell (Figure 2b). A significant and positive correlation was observed between Fc obtained by microarray analyses and Fe (Figure 2c). The absence of correlation between RIP Fe and mRNA levels (Figure 2d) underlines that BRCA1 associates with these mRNA independently of their cytoplasmic quantity.

3.2. BRCA1 Controls Translation of a Subset of BRCA1-Associated mRNAs

To investigate whether BRCA1 regulates the translation of mRNA with which it associates, we silenced BRCA1 expression in MCF-7 cells using a previously described BRCA1-targeting siRNA [17,20], which achieved clear BRCA1 depletion (Figure 3a).

We assessed the translational efficiency (Te) of each of the 16 mRNAs by analyzing mRNA abundance in polysomal fractions (i.e., actively translated mRNA) related to the total cytoplasmic mRNA, and compared control and BRCA1-depleted cells, in three independent experiments (Figure 3b).

First we performed a polysome profiling following three sequential steps as follows. (1) A large fraction (80%) of the cytoplasmic extract was loaded on top of a 10%–40% sucrose gradient. After ultracentrifugation, untranslated mRNAs were present in the top fractions whereas 40S and 60S free subunits, 80S ribosomes, and polysome-associated mRNAs were distributed along the bottom part of the gradient. (2) The gradient was then collected using a flow cell coupled to a spectrophotometer to measure and register continuously the concentration of RNA in each fraction of the gradient represented by the OD at 254 nm. The gradient was collected into 14 fractions of equal volumes. The variations of RNA concentration were reported as a function of the gradient fractions and thus determine which fractions contain free ribosomal subunits, monosomal, or polysomal material. (3) Fractions containing polysomal material were recovered, pooled, and processed for RNA extraction. In parallel, 20% of the cytoplasmic extract was directly processed for RNA extraction.

By comparing profiles obtained with BRCA1-depleted cells with those of control cells, we observed that they were similar, notably regarding the fractions 7–14 containing actively translated mRNA, thus indicating that translation was not grossly altered by BRCA1.

Next, for each of the 16 mRNA of interest (see Figure 2), RT-qPCR were performed on total RNA from BRCA1-depleted and control cells and on polysomal RNA from BRCA1-depleted and control cells. Then the Te was calculated following a three steps procedure. (1) We first investigated the variations in mRNA content within the total cytoplasmic RNA fraction in the presence and absence of BRCA1, to determine whether BRCA1 modulates the amounts of cytoplasmic mRNAs reflecting their rates of synthesis, transport, and stability ($\text{totRNA} = \text{si-BRCA1}/\text{si-Ctrl}$). (2) Then, to evaluate how BRCA1 affects mRNA recruitment within polysomes, we analyzed the amount of polysomal RNA in the presence and absence of BRCA1 ($\text{polyRNA} = \text{si-BRCA1}/\text{si-Ctrl}$). (3) Finally, we determined the Te of each mRNA by calculating the following ratio: $(\text{change in abundance in polysomal mRNA})/(\text{change in abundance in total mRNA})$ ($\text{Te} = \text{polyRNA}/\text{totRNA}$). Therefore, for each mRNA expressed, the Te value reveals the change in its association with polysomes independently to any change in its total cytoplasmic amount.

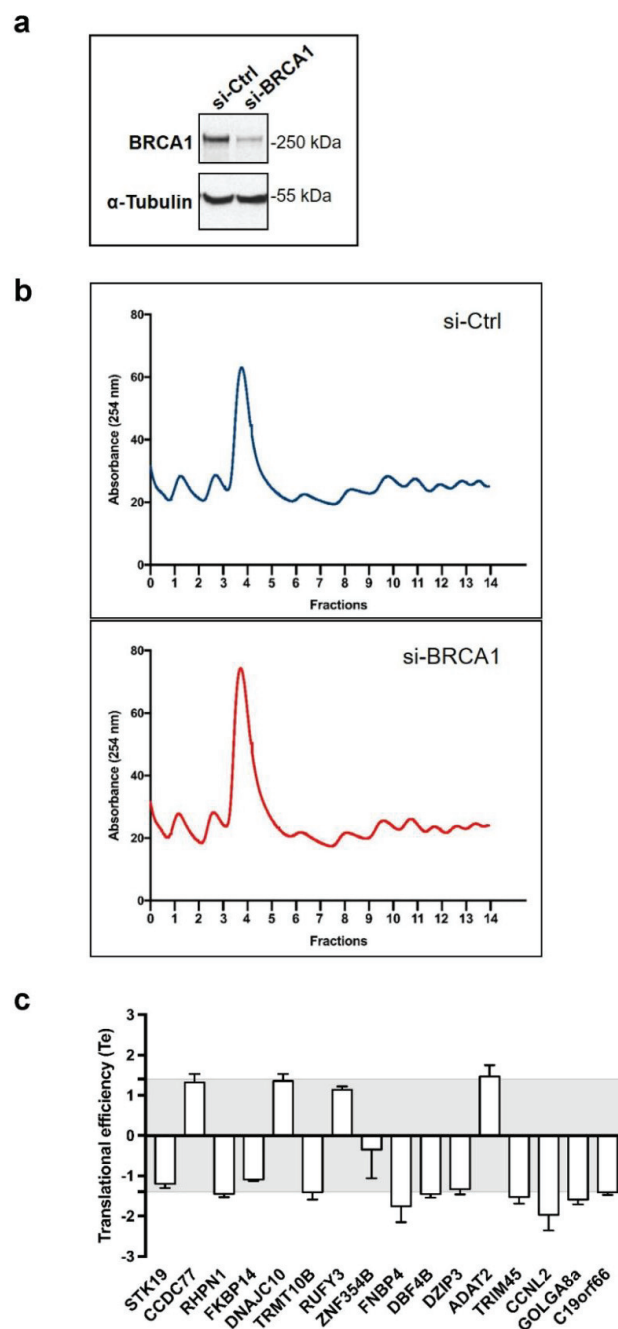


Figure 3. BRCA1 controls translation of a subset of BRCA1-associated mRNAs. (a) Immunoblots confirming siRNA inhibition of BRCA1 (si-BRCA1) when compared with control siRNA (si-Ctrl) in MCF-7 cells. α -Tubulin served as a loading control. (b) Polysomal profiles of MCF-7 cells in response to depletion of BRCA1. 40S and 60S ribosomal subunits, 80S ribosomes, and polysomes were separated by ultracentrifugation on sucrose gradients. One representative polysome profile of cells transfected with control siRNA (si-Ctrl) and with BRCA1-targeting siRNA (si-BRCA1) is shown. (c) Analysis of the translational efficiency (Te) of the 16 BRCA1-associated mRNAs identified in Figure 2. For BRCA1-depleted cells (si-BRCA1) and for control cells (si-Ctrl), fractions 7–14 containing polysomal material were recovered, pooled, and processed for RNA extraction. In parallel, a portion of the cytoplasmic extract was kept unprocessed to perform total RNA extraction from si-BRCA1 and from si-Ctrl cells. For each of the 16 mRNA of interest, RT-qPCR was performed on total RNA and polysomal RNA from si-BRCA1 and si-Ctrl cells. The Te was determined by calculating the following ratio: (change in abundance in polysomal mRNA in the absence and presence of BRCA1)/(change in abundance in total mRNA in absence and presence of BRCA1). Data are expressed as means \pm SEM. $n = 3$.

By applying cut-off values of 1.4, we found that loss of BRCA1 deregulated the translation of 7 out of the 16 mRNAs analyzed (Figure 3c) by impacting their association with polysomes independently of their abundance in the total cytoplasmic fraction (Figure S1). This cut-off value of 1.4 allowed selecting a number of mRNA, which was found to (i) further analyze by Western blot and (ii) more importantly further the investigation of their biological relevance using IHC on human breast cancer samples.

Taken together, these results established a set of seven mRNAs that were translationally controlled by the BRCA1 protein, with ADAT2 being upregulated, and CCNL2, DBF4B, FNBP4, GOLGA8a, RHPN1, and TRIM45 downregulated upon BRCA1 depletion.

3.3. BRCA1 Controls ADAT2, CCNL2, DBF4B, and TRIM45 Proteins in Mammary Epithelial Cell Lines

Next, we investigated whether modifications in translational efficiency of these mRNA upon BRCA1 depletion lead to changes in the corresponding protein levels. Due to antibody limitations, we could only perform immunoblots of four proteins, namely ADAT2, CCNL2, DBF4B, and TRIM45. We found that BRCA1 depletion induced an increase in the ADAT2 protein and conversely reduced CCNL2, DBF4B, and TRIM45 levels in MCF-7 cells, changes that correlated with the translational efficiency trend (Figure 4a).

These findings were corroborated in other tumoral (MDA-MB-231) and non-tumoral (HMECs) mammary cells (Figure 4b,c), indicating that BRCA1 deficiency alters the translation of selected mRNAs and further affects the quantity of their protein products. Conversely, overexpression of BRCA1 in MCF-7 cells decreased ADAT2 and increased CCNL2, DBF4B, and TRIM45 protein levels (Figure 4d).

To further ascertain the impact of BRCA1 overexpression on the function of the ADAT2 protein in MCF-7 cells, we examined changes in anticodon inosine abundance. Indeed, since ADAT2 generates inosine from adenosine at position 34 in the anticodon of tRNA [23], we quantified tRNA A34-to-I34 editing by comparing I34 levels in control MCF-7 cells and in those overexpressing BRCA1. We found that overexpression of BRCA1 (Figure 4e), while not affecting ADAT2 mRNA levels (Figure 4f), but reducing ADAT2 protein levels (Figure 4e), significantly reduced the levels of I34 in at least two substrates of ADAT2: tRNA^{Ser}AGA and tRNA^{Val}AAC (Figure 4g). Importantly, this reduction is equivalent to that reported in HEK293T cells upon ADAT2 depletion [23]. These results show that BRCA1 controlled the ADAT2 function by translationally modulating ADAT2 abundance.

3.4. Altered Expression of ADAT2, CCNL2, DBF4B, and TRIM45 Proteins in BRCA1 Deficient Human Breast Cancers

Given the functional implications of BRCA1 inactivation on the translation of ADAT2, CCNL2, DBF4B, and TRIM45, we wondered whether the levels of these proteins were correlated to BRCA1 status in breast tumors. We set up cohorts of patients with invasive breast cancers and documented germ-line BRCA1 and BRCA2 pathogenic variant statuses (Table 1).

We analyzed by immunohistochemistry (IHC) the expression of ADAT2, CCNL2, DBF4B, and TRIM45 in an exploratory cohort of 45 tumor samples, 26 from patients with no BRCA1 pathogenic variant (WT BRCA1), and 19 from patients with documented BRCA1 pathogenic variant (MT BRCA1) hampering BRCA1 protein production. Since patients in the MT BRCA1 and WT BRCA1 groups were matched according to other histoprognostic factors (Scarff Bloom and Richardson (SBR) grade, hormonal status, and histological subtype), BRCA1 expression was the only discriminant factor between the two groups, enabling the comparison of ADAT2, CCNL2, DBF4B, and TRIM45 expressions (Table 1).

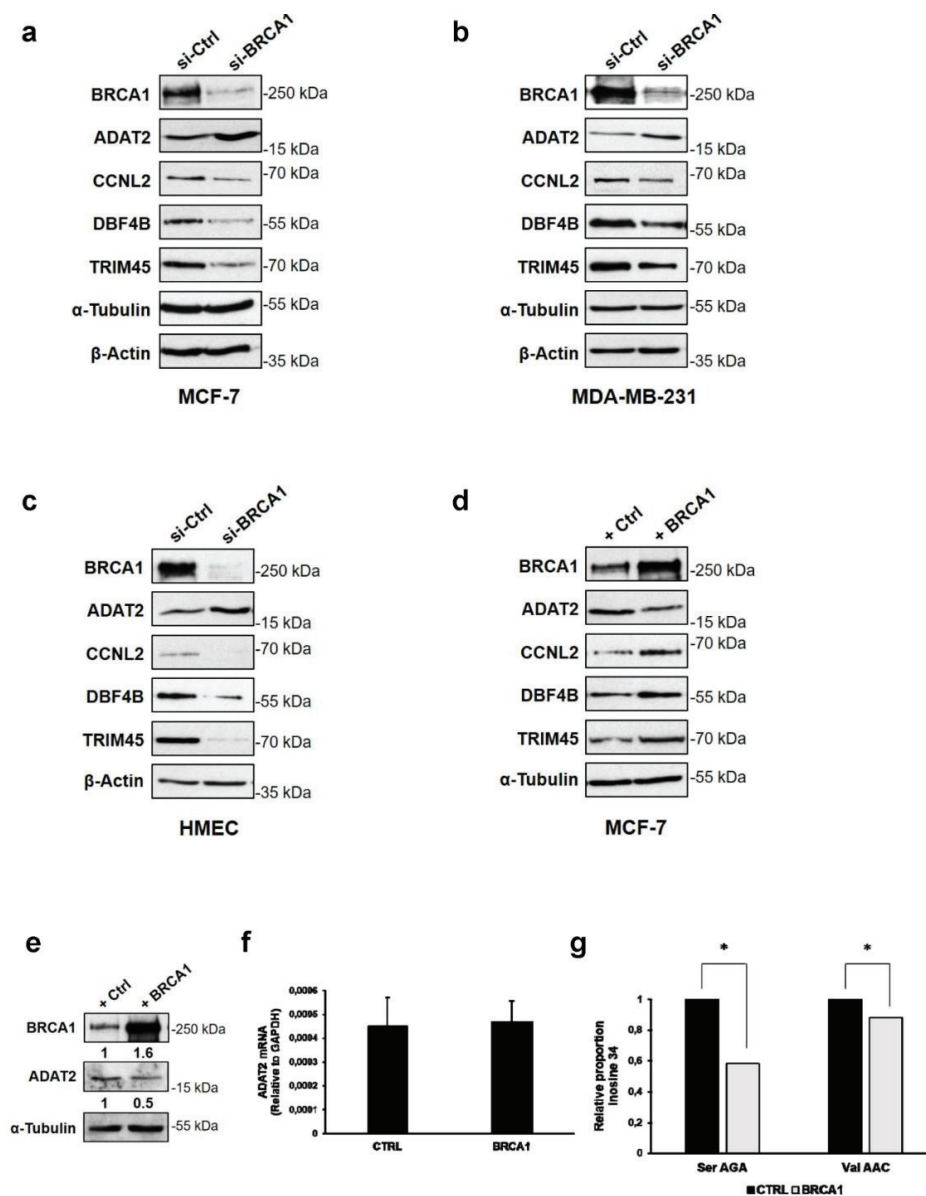


Figure 4. BRCA1 controls ADAT2, CCNL2, DBF4B, and TRIM45 proteins in mammary epithelial cell lines. (a–c) Immunoblot analyses of ADAT2, CCNL2, DBF4B, and TRIM45 proteins from MCF-7 cells (a), MDA-MB-231 cells (b), and HMECs (c) transfected with BRCA1-targeting siRNA (si-BRCA1) or control siRNA (si-Ctrl). (d) Immunoblot analysis of ADAT2, DBF4B, CCNL2, and TRIM45 proteins from MCF-7 cells transfected with BRCA1-expressing plasmid (+BRCA1) or empty plasmid as control (+Ctrl). β -actin and α -tubulin were used as loading controls. (e) Immunoblot showing increased levels of BRCA1 and decreased levels of ADAT2 in MCF-7 cells transfected with BRCA1-expressing plasmid (+BRCA1) compared with control (+Ctrl). Numbers indicate the quantity of BRCA1 and ADAT2 in BRCA1-enriched cells compared to control cells and normalized against α -Tubulin. (f) RT-qPCR analysis of ADAT2 mRNA in MCF-7 cells overexpressing BRCA1 (BRCA1) or a control vector (CTRL). Experiments were performed as previously described [23,24]. $n = 2$. (g) Relative proportion of inosine found at position 34 in tRNASerAGA and tRNAValAAC in BRCA1-enriched (BRCA1) compared to control (CTRL) MCF-7 cells. Data are expressed as means \pm SEM. $n = 2$. Results of the Fisher’s Exact Test are indicated by * $p < 0.05$.

Table 1. Clinical and tumor characteristics of breast carcinoma tissues.

Characteristics	Exploratory Cohort		Validation Cohort
	WT BRCA1	MT BRCA1	MT BRCA1
	<i>n</i> = 26	<i>n</i> = 19	<i>n</i> = 13
Tumor Type (%)			
Invasive ductal carcinoma	26 (100)	19 (100)	13 (100)
SBR Grade (%)			
1	0 (0)	0 (0)	0 (0)
2	8 (31)	0 (0)	2 (15)
3	18 (69)	19 (100)	11 (85)
Subtype (%)			
HR negative HER2 negative	17 (65)	13 (68)	10 (77)
HR positive HER2 negative	9 (35)	6 (32)	3 (23)
Age (year)			
Median (min-max)	43 (28–65)	38 (23–61)	43 (36–65)

Staining scores showed that the expression of ADAT2 was significantly increased in MT BRCA1 tumors compared to WT BRCA1 tumors (Figure 5a,e), consistent with ADAT2 upregulation in BRCA1-depleted mammary cell lines. Conversely, staining scores of CCNL2, DBF4B, and TRIM45 were significantly decreased in MT BRCA1 tumors, supporting the above findings of a downregulation of these proteins in BRCA1-depleted cells (Figure 5b–e).

Interestingly, ADAT2 appeared as the sole protein to be enriched in BRCA1-deficient tumors and was readily measured by IHC. This raised the possibility that ADAT2 could constitute a candidate biomarker of BRCA1 deficiency, since determining the BRCA1 status remains an issue. We analyzed ADAT2 protein levels in a validation cohort of 13 tumor samples from MT BRCA1 patients (Figure 5a and Table 1) and found that ADAT2 staining was statistically different compared to the control group (WT BRCA1), confirming that ADAT2 high expression was correlated with low expression of BRCA1 in breast tumor samples.

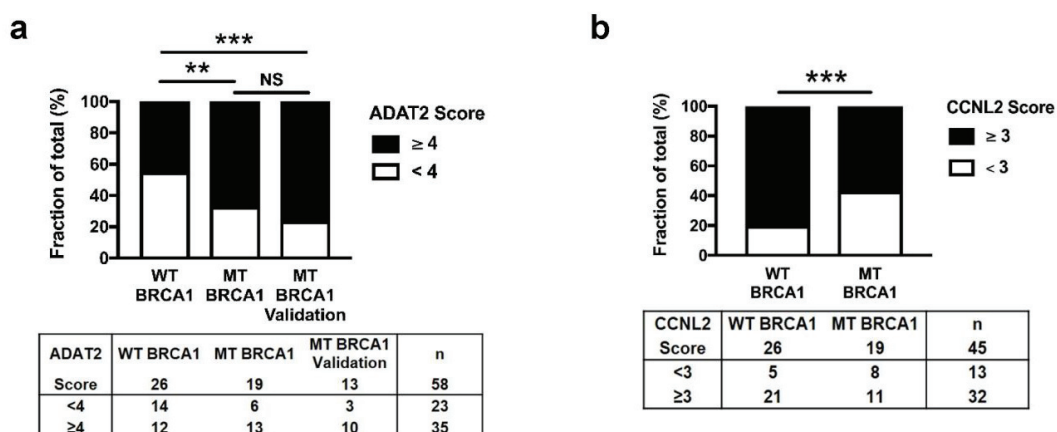


Figure 5. Cont.

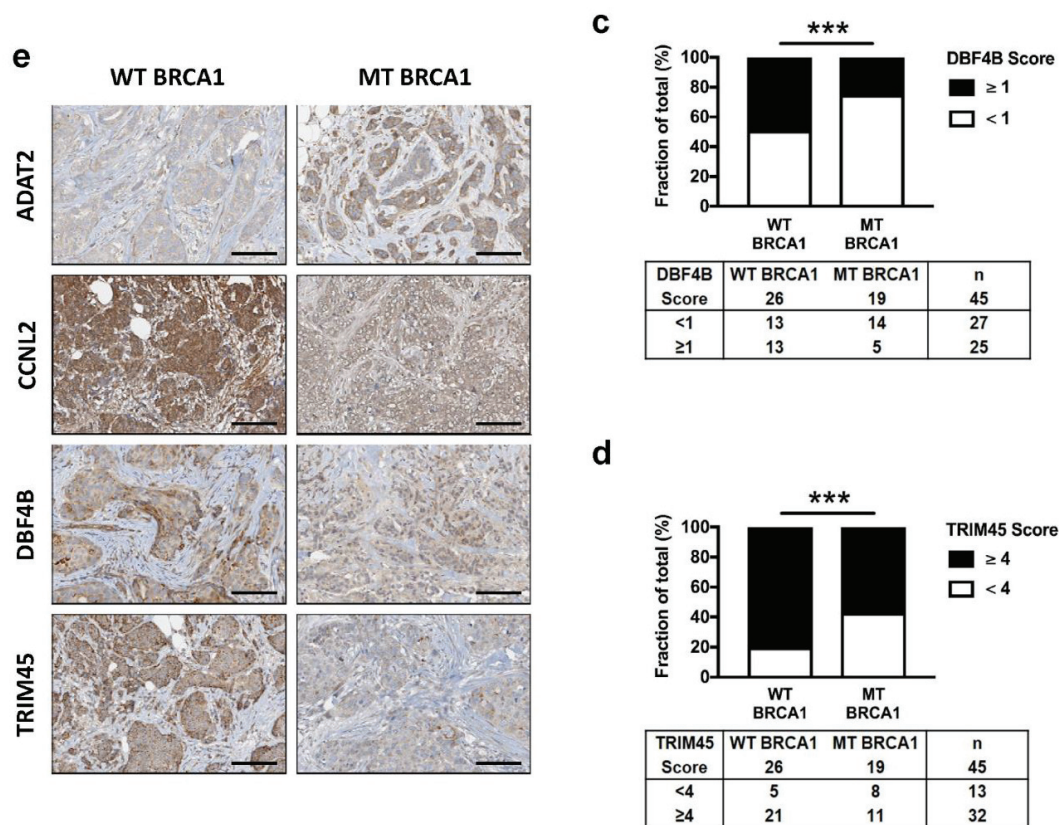


Figure 5. Altered expression of ADAT2, CCNL2, DBF4B, and TRIM45 proteins in BRCA1 deficient human breast cancers. (a–d) Score of ADAT2 (a), CCNL2 (b), DBF4B (c), and TRIM45 (d) in breast cancer samples of 45 patients (exploratory cohort): 26 patients non-mutated (WT BRCA1) and 19 patients mutated (MT BRCA1) for BRCA1. For ADAT2, a validation cohort of 13 BRCA1 mutated patients (MT BRCA1 validation) was additionally analyzed. Stacked bars represent the fraction of total tumor samples (expressed in %) from a tissue microarray (TMA) with low (white bar) and high (black bar) scores. Results of Chi-square test are indicated by NS = non-significant; ** $p < 0.01$; *** $p < 0.001$. Corresponding number of samples are presented under histograms. (e) Representative TMA immunohistochemical images demonstrating differential ADAT2, CCNL2, DBF4B, and TRIM45 staining between BRCA1 mutated (MT BRCA1) and BRCA1 non-mutated (WT BRCA1) tumors. Scale bars = 100 μm .

4. Discussion

We herein demonstrated that the BRCA1 protein was a translational regulator, which controlled the translational efficiency of a subset of mRNAs with which it associates.

To our knowledge, genome-wide association of mRNA with BRCA1 has not been described previously. In addition, few studies performing small scale analysis of BRCA1-dependent RIP identified some BRCA1-bound RNA [25–27]. However, no direct association with BRCA1 has been shown until now as two large scale analyses searching for RNA binding proteins by using UV-crosslink failed to identify BRCA1 among the proteins identified [28,29]. These data and our results underscored the importance of further deciphering at the molecular level the ribonucleoprotein complexes in which BRCA1 protein associates to mRNA. A search for particular RNA motif (Figure S2) within BRCA1-bound mRNAs identified 14 motifs among which only two (motif #2 and motif #7) happened to be recognized by the RBPs SRP14 and KHSRP respectively. Thus, it is tempting to speculate that BRCA1 regulates the translation of subsets of mRNA defined by either motif #2 or motif #7, through its interaction with specific translation initiation complex containing either SRP14 or KHSRP.

The 5′ untranslated regions (5′UTRs) and 3′ untranslated regions (3′UTRs) of mRNA contain cis-regulatory sequences that allow direct binding with trans-regulators and consequently contribute to modulate the translational efficiency of the corresponding mRNA [30]. Among them, some are secondary structures such as Internal Ribosome Entry Sites, stem-loops, and RNA G-quadruplex. Thus, it will be worth examining the features of the 5′UTR and the 3′UTR of the mRNA translationally controlled by and associated to the BRCA1 protein, to further decipher at the molecular level the mode of action of BRCA1. Especially ribonucleoprotein complexes containing both the BRCA1 protein and CCNL2, DBF4B, TRIM45, or ADAT2 mRNA will need deep scrutiny to shed light on the molecular mechanism through which BRCA1 regulates the translational efficiency of these mRNA.

Low BRCA1 expression in cells along with BRCA1 inactivation in breast tumors is associated with reduced expression of CCNL2, DBF4B, and TRIM45. These proteins are respectively implicated in mRNA splicing and apoptosis, DNA replication, and cell cycle regulation, transcription repression of pro-proliferative signaling, all biological functions known to contribute to the tumor suppressive activity of BRCA1.

Firstly, CCNL2 transcriptionally regulates pre-mRNA splicing [31,32]. Its overexpression induces apoptosis of human hepatocellular and gastric carcinoma cell lines [31,33], while its knockdown promotes growth of pulmonary artery smooth muscle cells [34]. A recent preclinical study conducted in breast cancer cell lines and tissue arrays reported that CCNL2 mRNA transcript expression was lower in breast cancer than in normal breast tissue [35]. Therefore, positively controlling CCNL2 translation may represent a novel mechanism through which BRCA1 exerts its tumor suppressive activity.

Secondly, DBF4B is required for the efficient progression of S and M phases [36], and for DNA replication [37]. Accordingly, DBF4B knockdown leads to an increase in DNA damage in HeLa cells and affects cell cycle in the *Xenopus laevis* model [38]. Therefore, BRCA1 may control DNA replication and S phase progression in response to DNA damage, two hallmarks of its tumor suppressor activity, through at least in part positive regulation of DBF4B translation.

Thirdly, TRIM45 belongs to the family of tripartite motif proteins that play important roles in cell proliferation, differentiation, and apoptosis. TRIM45 suppresses cell proliferation by negatively regulating the mitogen-activated protein kinase signaling pathway [39] and the transcriptional activity of NF-κB [40]. In a large-scale gene expression array analysis conducted in breast cancer samples from Taiwanese women, TRIM45 expression was diminished [41]. Consistently, TRIM45 was recently identified as a tumor suppressor in the brain [42]. Here, the translational control of TRIM45 by BRCA1 that leads to low TRIM45 levels in BRCA1-deficient breast tumors, may represent a novel mechanism contributing to the onco-suppressive role of BRCA1.

Our most intriguing finding was however the repression of ADAT2 expression and function by BRCA1. ADAT2 is the catalytic subunit of the heterodimer ADAT2/ADAT3 responsible for the post-transcriptional modification of tRNA at its first anticodon (wobble) position (position 34). The inosine, (I34), which is generated from adenosine, can translate not only codons ending in uracil, but also those ending in cytosine or adenine. This modification enlarges the codon recognition capacity during protein synthesis [43,44]. We previously suggested that ADAT might be controlling specific genetic programs [45]. Therefore, the BRCA1-driven negative regulation of ADAT2 may silence specific genetic programs the codon usage of which makes their translation dependent on tRNAs with I34 [46]. Emerging evidence indicates that tRNA expression and tRNA repertoire are modulated during tumorigenesis [47,48]. Perturbations of a number of tRNA modifications and altered expression of the corresponding modifying enzymes such as TRMT12, NSUN2, or DNMT2, have been linked to numerous human diseases, including cancer, neurological disorders, and mitochondrial-linked disorders [49,50]. Thus, ADAT2 up-regulation observed upon BRCA1 inactivation in breast tumors exposes A34-to-I34 editing as a novel tRNA modification that is altered in cancer.

These findings raise the possibility of using ADAT2 as a predictive biomarker of efficacy of anti-PARP-based therapy. Indeed, since PARP inhibitors are now available for the treatment of metastatic breast cancer [51], ADAT2 could be used as a predictor of efficacy of such treatments. Of

course, this would need further exploration, in particular in patients with BRCA1 pathogenic variant in the tumor (i.e., somatic BRCA1 pathogenic variant).

The anti-PARP-based therapy recently proposed to patients with BRCA1 deficiency makes the identification of BRCA1 pathogenic variant carriers essential. Today, BRCA1 gene sequencing remains the mainstay of BRCA1 pathogenic variants identification [52]. However, widespread BRCA1 genetic testing in unselected individuals is hardly feasible [53]. A simple IHC assay to monitor BRCA1 deficiency directly in tumor biopsies could be more convenient and affordable [54]. So far, none of the many attempts of setting up IHC BRCA1 staining to assign the tumor BRCA1 status has been successful [55–58]. Our analysis shows a correlation between the high level of ADAT2 and the BRCA1 deficiency among BRCA1 deficient breast tumor samples. Therefore, it could be worth examining further whether detection of ADAT2 by IHC could be proposed as a first screen for BRCA1 deficiency. Moreover, sporadic breast cancer cases, which harbor BRCA1 deficiency, may benefit from this IHC since they become more and more important in clinical practice, with respect to anti-PARP based therapy. Nevertheless, our analysis is based on a relatively small sample size and so further validation on a large cohort will be required to fully assess the potential of ADAT2 as a marker of BRCA1 deficiency.

BRCA1-mediated ADAT2 regulation may contribute to increase the diversity of the proteome. In addition we have shown that BRCA1 regulates protein synthesis [17] and it was reported that by its interaction with RNA Polymerase I and RNA Polymerase III machineries, BRCA1 could regulate synthesis of rRNA and tRNA [59,60]. This strongly supports the notion that BRCA1 is a translational regulator, a key function for its tumor suppressive activity as it has been shown for the P53 tumor suppressor [61,62].

5. Conclusions

In this study, we provided evidence that BRCA1 altered translational efficiency of subsets of mRNAs involved in major programs in human tumors. The BRCA1 protein associated with these mRNA in ribonucleoprotein complexes and modulated their translational efficiency. We focused on ADAT2, CCNL2, DBF4B, and TRIM45 mRNAs that played key roles in cancer biology to demonstrate that alteration of their translational efficiency occurred via a BRCA1-dependent mechanism in cultured cells and in patient tumors. This study raised the possibility of using ADAT2 as a predictive biomarker of efficacy of anti-PARP-based therapy. These findings are crucial not only for basic research but also for clinical outcome since most BRCA1 mutated tumors are defined as “triple negative breast tumors”, which are associated with a very poor prognosis.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4409/9/4/941/s1>, Supplementary Material and Methods, Figure S1: RT-qPCR analyses of differentially translated mRNA upon BRCA1 depletion, Figure S2: Search for RNA motifs present in the 488 RNA interacting with BRCA1, Table S1: Primers used for qPCR.

Author Contributions: Conceptualization, E.B., A.V., E.D., L.E., I.T., N.D.V. and J.-J.D.; methodology, E.B., A.V., E.D., L.E. and I.T.; validation, A.V., A.G.T., L.R.d.P., N.D.V. and V.M.; formal analysis, E.B., A.V., L.E., I.T., H.P.; A.G.T.; L.R.d.P., J.L. and N.D.V.; investigation, E.B., A.V., C.R., L.E., I.T. and A.G.T.; data curation, E.B., A.G.T., L.R.d.P., J.L. and C.R.; writing—original draft preparation, E.B., A.V., L.E., I.T., J.L. and A.G.T.; writing—review and editing, E.D., V.M., F.C., L.R.d.P., J.-J.D. and N.D.V.; supervision, J.-J.D. and N.D.V.; project administration, J.-J.D. and N.D.V.; funding acquisition, F.C., V.M., A.G.T., L.R.d.P., J.-J.D. and N.D.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the “Ligue Contre le Cancer, comité départemental du Rhône”, the Conseil de Radioprotection Electricité de France (EDF), the translational program CRCL-CLB, with AJORA (Association des Jeunes Oncologues de Rhône-Alpes), the Spanish Ministry of Economy and Competitiveness under Grant [BIO2015-64752] to L.R.P. J.J.D. laboratory is supported by Inserm, CNRS, Université Claude Bernard Lyon 1, Centre Léon Bérard, Agence Nationale pour la Recherche (RIBOMETH ANR-13-BSV8-0012-01), Programmes d’Actions Intégrées de Recherche (PAIR Sein, RiboTEM, ARC_INCa_LNCC_7625), SIRIC LYriCAN (INCa-DGOS-Inserm_12563), Ligue Nationale contre le Cancer, Ligue contre le Cancer comité départemental (Drôme, Rhône, Isère, Puy-de-Dôme, Allier) and Fondation ARC pour la Recherche sur le Cancer (2016120486). E.B. and E.D. were supported by a fellowship from the French Ministry of Research, H.P. was supported by a fellowship from the Ligue Nationale contre le Cancer.

Acknowledgments: We thank Brigitte Manship for editing the manuscript, Sanjeev Kumar for Gene Ontology analysis, Amélie Colombe Vermorel and Laetitia Odeyer for performing TMA and IHC.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca A Cancer J. Clin.* **2018**, *68*, 394–424. [[CrossRef](#)] [[PubMed](#)]
2. Malone, K.E.; Daling, J.R.; Doody, D.R.; Hsu, L.; Bernstein, L.; Coates, R.J.; Marchbanks, P.A.; Simon, M.S.; McDonald, J.A.; Norman, S.A.; et al. Prevalence and predictors of BRCA1 and BRCA2 mutations in a population-based study of breast cancer in white and black American women ages 35 to 64 years. *Cancer Res.* **2006**, *66*, 8297–8308. [[CrossRef](#)] [[PubMed](#)]
3. Kurian, A.W.; Gong, G.D.; John, E.M.; Miron, A.; Felberg, A.; Phipps, A.I.; West, D.W.; Whittemore, A.S. Performance of prediction models for BRCA mutation carriage in three racial/ethnic groups: Findings from the Northern California Breast Cancer Family Registry. *Cancer Epidemiol. Biomark. Prev.* **2009**, *18*, 1084–1091. [[CrossRef](#)]
4. Armstrong, N.; Ryder, S.; Forbes, C.; Ross, J.; Quek, R.G. A systematic review of the international prevalence of BRCA mutation in breast cancer. *Clin. Epidemiol.* **2019**, *11*, 543–561. [[CrossRef](#)] [[PubMed](#)]
5. Kuchenbaecker, K.B.; Hopper, J.L.; Barnes, D.R.; Phillips, K.A.; Mooij, T.M.; Roos-Blom, M.J.; Jervis, S.; van Leeuwen, F.E.; Milne, R.L.; Andrieu, N.; et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA* **2017**, *317*, 2402–2416. [[CrossRef](#)] [[PubMed](#)]
6. Engel, C.; Fischer, C.; Zachariae, S.; Bucksch, K.; Rhiem, K.; Giesecke, J.; Herold, N.; Wappenschmidt, B.; Hubbel, V.; Maringa, M.; et al. Breast cancer risk in BRCA1/2 mutation carriers and noncarriers under prospective intensified surveillance. *Int. J. Cancer* **2019**. [[CrossRef](#)]
7. Nik-Zainal, S.; Davies, H.; Staaf, J.; Ramakrishna, M.; Glodzik, D.; Zou, X.; Martincorena, I.; Alexandrov, L.B.; Martin, S.; Wedge, D.C.; et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **2016**, *534*, 47–54. [[CrossRef](#)]
8. Staff, S.; Isola, J.; Tanner, M. Haplo-insufficiency of BRCA1 in sporadic breast cancer. *Cancer Res.* **2003**, *63*, 4978–4983.
9. Atchley, D.P.; Albarracin, C.T.; Lopez, A.; Valero, V.; Amos, C.I.; Gonzalez-Angulo, A.M.; Hortobagyi, G.N.; Arun, B.K. Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. *J. Clin. Oncol.* **2008**, *26*, 4282–4288. [[CrossRef](#)]
10. Robson, M.; Im, S.A.; Senkus, E.; Xu, B.; Domchek, S.M.; Masuda, N.; Delaloge, S.; Li, W.; Tung, N.; Armstrong, A.; et al. Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. *N. Engl. J. Med.* **2017**, *377*, 523–533. [[CrossRef](#)]
11. Venkitaraman, A.R. Cancer suppression by the chromosome custodians, BRCA1 and BRCA2. *Science* **2014**, *343*, 1470–1475. [[CrossRef](#)] [[PubMed](#)]
12. Silver, D.P.; Livingston, D.M. Mechanisms of BRCA1 tumor suppression. *Cancer Discov.* **2012**, *2*, 679–684. [[CrossRef](#)]
13. Roy, R.; Chun, J.; Powell, S.N. BRCA1 and BRCA2: Different roles in a common pathway of genome protection. *Nat. Rev. Cancer* **2011**, *12*, 68–78. [[CrossRef](#)] [[PubMed](#)]
14. Pathania, S.; Bade, S.; Le Guillou, M.; Burke, K.; Reed, R.; Bowman-Colin, C.; Su, Y.; Ting, D.T.; Polyak, K.; Richardson, A.L.; et al. BRCA1 haploinsufficiency for replication stress suppression in primary cells. *Nat. Commun.* **2014**, *5*, 5496. [[CrossRef](#)]
15. Zhang, X.; Chiang, H.C.; Wang, Y.; Zhang, C.; Smith, S.; Zhao, X.; Nair, S.J.; Michalek, J.; Jatoi, I.; Lautner, M.; et al. Attenuation of RNA polymerase II pausing mitigates BRCA1-associated R-loop accumulation and tumorigenesis. *Nat. Commun.* **2017**, *8*, 15908. [[CrossRef](#)]
16. Savage, K.I.; Harkin, D.P. BRCA1, a ‘complex’ protein involved in the maintenance of genomic stability. *Febs J.* **2015**, *282*, 630–646. [[CrossRef](#)] [[PubMed](#)]
17. Dizin, E.; Gressier, C.; Magnard, C.; Ray, H.; Decimo, D.; Ohlmann, T.; Dalla Venezia, N. BRCA1 interacts with poly(A)-binding protein: Implication of BRCA1 in translation regulation. *J. Biol. Chem.* **2006**, *281*, 24236–24246. [[CrossRef](#)] [[PubMed](#)]

18. Dacheux, E.; Vincent, A.; Nazaret, N.; Combet, C.; Wierinckx, A.; Mazoyer, S.; Diaz, J.J.; Lachuer, J.; Venezia, N.D. BRCA1-Dependent Translational Regulation in Breast Cancer Cells. *PLoS ONE* **2013**, *8*, e67313. [[CrossRef](#)] [[PubMed](#)]
19. Eisinger, F.; Bressac, B.; Castaigne, D.; Cottu, P.H.; Lansac, J.; Lefranc, J.P.; Lesur, A.; Nogues, C.; Pierret, J.; Puy-Pernias, S.; et al. Identification and management of hereditary predisposition to cancer of the breast and the ovary (update 2004). *Bull. Du Cancer* **2004**, *91*, 219–237.
20. Vincent, A.; Berthel, E.; Dacheux, E.; Magnard, C.; Venezia, N.L. BRCA1 affects protein phosphatase 6 signalling through its interaction with ANKRD28. *Biochem. J.* **2016**, *473*, 949–960. [[CrossRef](#)]
21. Keene, J.D.; Komisarow, J.M.; Friedersdorf, M.B. RIP-Chip: The isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat. Protoc.* **2006**, *1*, 302–307. [[CrossRef](#)] [[PubMed](#)]
22. Huang da, W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57. [[CrossRef](#)] [[PubMed](#)]
23. Torres, A.G.; Pineyro, D.; Rodriguez-Escriba, M.; Camacho, N.; Reina, O.; Saint-Leger, A.; Filonava, L.; Batlle, E.; Ribas de Pouplana, L. Inosine modifications in human tRNAs are incorporated at the precursor tRNA level. *Nucleic Acids Res.* **2015**, *43*, 5145–5157. [[CrossRef](#)] [[PubMed](#)]
24. Wulff, T.F.; Arguello, R.J.; Molina Jordan, M.; Roura Frigole, H.; Hauquier, G.; Filonava, L.; Camacho, N.; Gatti, E.; Pierre, P.; Ribas de Pouplana, L.; et al. Detection of a Subset of Posttranscriptional Transfer RNA Modifications in Vivo with a Restriction Fragment Length Polymorphism-Based Method. *Biochemistry* **2017**, *56*, 4029–4038. [[CrossRef](#)] [[PubMed](#)]
25. Kawai, S.; Amano, A. BRCA1 regulates microRNA biogenesis via the DROSHA microprocessor complex. *J. Cell Biol.* **2012**, *197*, 201–208. [[CrossRef](#)] [[PubMed](#)]
26. Savage, K.I.; Gorski, J.J.; Barros, E.M.; Irwin, G.W.; Manti, L.; Powell, A.J.; Pellagatti, A.; Lukashchuk, N.; McCance, D.J.; McCluggage, W.G.; et al. Identification of a BRCA1-mRNA splicing complex required for efficient DNA repair and maintenance of genomic stability. *Mol. Cell* **2014**, *54*, 445–459. [[CrossRef](#)] [[PubMed](#)]
27. Sharma, V.; Khurana, S.; Kubben, N.; Abdelmohsen, K.; Oberdoerffer, P.; Gorospe, M.; Misteli, T. A BRCA1-interacting lncRNA regulates homologous recombination. *Embo Rep.* **2015**, *16*, 1520–1534. [[CrossRef](#)]
28. Baltz, A.G.; Munschauer, M.; Schwanhauser, B.; Vasile, A.; Murakawa, Y.; Schueler, M.; Youngs, N.; Penfold-Brown, D.; Drew, K.; Milek, M.; et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **2012**, *46*, 674–690. [[CrossRef](#)]
29. Castello, A.; Fischer, B.; Eichelbaum, K.; Horos, R.; Beckmann, B.M.; Strein, C.; Davey, N.E.; Humphreys, D.T.; Preiss, T.; Steinmetz, L.M.; et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **2012**, *149*, 1393–1406. [[CrossRef](#)]
30. Jackson, R.J.; Hellen, C.U.; Pestova, T.V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 113–127. [[CrossRef](#)]
31. Yang, L.; Li, N.; Wang, C.; Yu, Y.; Yuan, L.; Zhang, M.; Cao, X. Cyclin L2, a novel RNA polymerase II-associated cyclin, is involved in pre-mRNA splicing and induces apoptosis of human hepatocellular carcinoma cells. *J. Biol. Chem.* **2004**, *279*, 11639–11648. [[CrossRef](#)] [[PubMed](#)]
32. Loyer, P.; Trembley, J.H.; Grenet, J.A.; Busson, A.; Corlu, A.; Zhao, W.; Kocak, M.; Kidd, V.J.; Lahti, J.M. Characterization of cyclin L1 and L2 interactions with CDK11 and splicing factors: Influence of cyclin L isoforms on splice site selection. *J. Biol. Chem.* **2008**, *283*, 7721–7732. [[CrossRef](#)] [[PubMed](#)]
33. Li, H.L.; Huang, D.Z.; Deng, T.; Zhou, L.K.; Wang, X.; Bai, M.; Ba, Y. Overexpression of cyclin L2 inhibits growth and enhances chemosensitivity in human gastric cancer cells. *Asian Pac. J. Cancer Prev.* **2012**, *13*, 1425–1430. [[CrossRef](#)] [[PubMed](#)]
34. Liu, H.; Tao, Y.; Chen, M.; Yu, J.; Li, W.J.; Tao, L.; Li, Y.; Li, F. Upregulation of MicroRNA-214 Contributes to the Development of Vascular Remodeling in Hypoxia-induced Pulmonary Hypertension Via Targeting CCNL2. *Sci. Rep.* **2016**, *6*, 24661. [[CrossRef](#)] [[PubMed](#)]
35. Kren, B.T.; Unger, G.M.; Abedin, M.J.; Vogel, R.I.; Henzler, C.M.; Ahmed, K.; Trembley, J.H. Preclinical evaluation of cyclin dependent kinase 11 and casein kinase 2 survival kinases as RNA interference targets for triple negative breast cancer therapy. *Breast Cancer Res.* **2015**, *17*, 19. [[CrossRef](#)] [[PubMed](#)]
36. Yoshizawa-Sugata, N.; Ishii, A.; Taniyama, C.; Matsui, E.; Arai, K.; Masai, H. A second human Dbf4/ASK-related protein, Drf1/ASKL1, is required for efficient progression of S and M phases. *J. Biol. Chem.* **2005**, *280*, 13062–13070. [[CrossRef](#)] [[PubMed](#)]

37. Takahashi, T.S.; Walter, J.C. Cdc7-Drf1 is a developmentally regulated protein kinase required for the initiation of vertebrate DNA replication. *Genes Dev.* **2005**, *19*, 2295–2300. [[CrossRef](#)]
38. Collart, C.; Smith, J.C.; Zegerman, P. Chk1 Inhibition of the Replication Factor Drf1 Guarantees Cell-Cycle Elongation at the *Xenopus laevis* Mid-blastula Transition. *Dev. Cell* **2017**, *42*, 82–96.e83. [[CrossRef](#)]
39. Wang, Y.; Li, Y.; Qi, X.; Yuan, W.; Ai, J.; Zhu, C.; Cao, L.; Yang, H.; Liu, F.; Wu, X.; et al. TRIM45, a novel human RBCC/TRIM protein, inhibits transcriptional activities of Elk-1 and AP-1. *Biochem. Biophys. Res. Commun.* **2004**, *323*, 9–16. [[CrossRef](#)]
40. Shibata, M.; Sato, T.; Nukiwa, R.; Ariga, T.; Hatakeyama, S. TRIM45 negatively regulates NF- κ B-mediated transcription and suppresses cell proliferation. *Biochem. Biophys. Res. Commun.* **2012**, *423*, 104–109. [[CrossRef](#)]
41. Huang, C.C.; Tu, S.H.; Lien, H.H.; Jeng, J.Y.; Huang, C.S.; Huang, C.J.; Lai, L.C.; Chuang, E.Y. Concurrent gene signatures for han chinese breast cancers. *PLoS ONE* **2013**, *8*, e76421. [[CrossRef](#)] [[PubMed](#)]
42. Zhang, J.; Zhang, C.; Cui, J.; Ou, J.; Han, J.; Qin, Y.; Zhi, F.; Wang, R.F. TRIM45 functions as a tumor suppressor in the brain via its E3 ligase activity by stabilizing p53 through K63-linked ubiquitination. *Cell Death Dis.* **2017**, *8*, e2831. [[CrossRef](#)] [[PubMed](#)]
43. Gerber, A.P.; Keller, W. An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* **1999**, *286*, 1146–1149. [[CrossRef](#)] [[PubMed](#)]
44. Novoa, E.M.; Pavon-Eternod, M.; Pan, T.; Ribas de Pouplana, L. A role for tRNA modifications in genome structure and codon usage. *Cell* **2012**, *149*, 202–213. [[CrossRef](#)]
45. Torres, A.G.; Pineyro, D.; Filonava, L.; Stracker, T.H.; Batlle, E.; Ribas de Pouplana, L. A-to-I editing on tRNAs: Biochemical, biological and evolutionary implications. *Febs Lett.* **2014**, *588*, 4279–4286. [[CrossRef](#)]
46. Rafels-Ybern, A.; Torres, A.G.; Grau-Bove, X.; Ruiz-Trillo, I.; Ribas de Pouplana, L. Codon adaptation to tRNAs with Inosine modification at position 34 is widespread among Eukaryotes and present in two Bacterial phyla. *Rna Biol.* **2017**, 1–8. [[CrossRef](#)]
47. Pavon-Eternod, M.; Gomes, S.; Geslain, R.; Dai, Q.; Rosner, M.R.; Pan, T. tRNA over-expression in breast cancer and functional consequences. *Nucleic Acids Res.* **2009**, *37*, 7268–7280. [[CrossRef](#)]
48. Goodarzi, H.; Nguyen, H.C.B.; Zhang, S.; Dill, B.D.; Molina, H.; Tavazoie, S.F. Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. *Cell* **2016**, *165*, 1416–1427. [[CrossRef](#)]
49. Kirchner, S.; Ignatova, Z. Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nat. Rev. Genet.* **2015**, *16*, 98–112. [[CrossRef](#)]
50. Torres, A.G.; Batlle, E.; Ribas de Pouplana, L. Role of tRNA modifications in human diseases. *Trends Mol. Med.* **2014**, *20*, 306–314. [[CrossRef](#)]
51. Litton, J.K.; Rugo, H.S.; Ettl, J.; Hurvitz, S.A.; Goncalves, A.; Lee, K.H.; Fehrenbacher, L.; Yerushalmi, R.; Mina, L.A.; Martin, M.; et al. Talazoparib in Patients with Advanced Breast Cancer and a Germline BRCA Mutation. *N. Engl. J. Med.* **2018**, *379*, 753–763. [[CrossRef](#)] [[PubMed](#)]
52. Popova, T.; Manie, E.; Rieunier, G.; Caux-Moncoutier, V.; Tirapo, C.; Dubois, T.; Delattre, O.; Sigal-Zafrani, B.; Bollet, M.; Longy, M.; et al. Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res.* **2012**, *72*, 5454–5462. [[CrossRef](#)] [[PubMed](#)]
53. Katz, S.J.; Ward, K.C.; Hamilton, A.S.; McLeod, M.C.; Wallner, L.P.; Morrow, M.; Jagsi, R.; Hawley, S.T.; Kurian, A.W. Gaps in Receipt of Clinically Indicated Genetic Counseling After Diagnosis of Breast Cancer. *J. Clin. Oncol.* **2018**, *36*, 1218–1224. [[CrossRef](#)] [[PubMed](#)]
54. Bankhead, P.; Fernandez, J.A.; McArt, D.G.; Boyle, D.P.; Li, G.; Loughrey, M.B.; Irwin, G.W.; Harkin, D.P.; James, J.A.; McQuaid, S.; et al. Integrated tumor identification and automated scoring minimizes pathologist involvement and provides new insights to key biomarkers in breast cancer. *Lab. Invest.* **2018**, *98*, 15–26. [[CrossRef](#)]
55. Vorrius, T.R.; Snyder, K.; Pica-Mendez, A.; Tan, C.; Laterza, O.; Toniatti, C.; Carpenter, C.; Lee, H.; Tanaka, W.; Zhang, Z.Q. Immunohistochemical Detection of BRCA-1 and BRCA-2 Expression in Human Breast and Ovarian Tumors. *J. Histochemol.* **2009**, *32*, 202–203. [[CrossRef](#)]
56. Perez-Valles, A.; Martorell-Cebollada, M.; Nogueira-Vazquez, E.; Garcia-Garcia, J.A.; Fuster-Diana, E. The usefulness of antibodies to the BRCA1 protein in detecting the mutated BRCA1 gene. An immunohistochemical study. *J. Clin. Pathol.* **2001**, *54*, 476–480. [[CrossRef](#)]

57. Milner, R.; Wombwell, H.; Eckersley, S.; Barnes, D.; Warwicker, J.; Van Dorp, E.; Rowlinson, R.; Dearden, S.; Hughes, G.; Harbron, C.; et al. Validation of the BRCA1 antibody MS110 and the utility of BRCA1 as a patient selection biomarker in immunohistochemical analysis of breast and ovarian tumours. *Virchows Arch.* **2013**, *462*, 269–279. [[CrossRef](#)]
58. Mangia, A.; Chiriatti, A.; Tommasi, S.; Menolascina, F.; Petroni, S.; Zito, F.A.; Simone, G.; Schittulli, F.; Paradiso, A. BRCA1 expression and molecular alterations in familial breast cancer. *Histol. Histopathol.* **2009**, *24*, 69–76. [[CrossRef](#)]
59. Johnston, R.; D’Costa, Z.; Ray, S.; Gorski, J.; Harkin, D.P.; Mullan, P.; Panov, K.I. The identification of a novel role for BRCA1 in regulating RNA polymerase I transcription. *Oncotarget* **2016**, *7*, 68097–68110. [[CrossRef](#)]
60. Veras, I.; Rosen, E.M.; Schramm, L. Inhibition of RNA polymerase III transcription by BRCA1. *J. Mol. Biol.* **2009**, *387*, 523–531. [[CrossRef](#)]
61. Marcel, V.; Catez, F.; Diaz, J.J. p53, a translational regulator: Contribution to its tumour-suppressor activity. *Oncogene* **2015**, *34*, 5513–5523. [[CrossRef](#)] [[PubMed](#)]
62. Marcel, V.; Nguyen Van Long, F.; Diaz, J.J. 40 Years of Research Put p53 in Translation. *Cancers* **2018**, *10*, 152. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Supplemental Material and Methods

Chemicals

The primary antibodies used for immunoblotting were as follows: mouse monoclonal antibodies to BRCA1 (MS110, Calbiochem), β -Actin (AC-15, Sigma) and α -Tubulin (DM1A, Calbiochem), rabbit polyclonal antibodies for ADAT2 (ab135429, Abcam), CCNL2 (A301-677A, Bethyl), DBF4B (ab177911, Abcam), TRIM45 (ab84303, Abcam). Secondary antibodies used were peroxidase-conjugated anti-mouse or anti-rabbit immunoglobulins (Jackson ImmunoResearch).

Case selection

Among these patients with such a medical history, we set up a cohort of patients with documented germ-line BRCA1 and BRCA2 pathogenic variant statuses, as determined by Next Generation Sequencing (NGS) in blood samples. Variants were classified as pathogenic according to the ENIGMA BRCA1/2 Gene Variant Classification Criteria (<http://www.enigmaconsortium.org/>). Women with variants of uncertain significance were considered as non-carriers. All patients were treated at the Regional Cancer Hospital "Centre Léon Bérard" (CLB) for breast adenocarcinoma after 2006. All patients gave written consent for the use of their tumor tissue. All patients that are undergoing tissue sampling (including surgery) in our institution have to give their consent for the use of the samples for research purposes (Institutional Ethics Committee). All patient data were completely de-identified before access and analysis. This was a monocentric study and the institution provided consent to use biological samples that were collected as part of routine patient care. Under French law, the study is therefore exempt from prior authorization by the CNIL. This study has been approved by the Institutional Ethics Committee from the Centre Léon Bérard.

Patients were only included in our cohort if they had not received any neo-adjuvant treatment before breast surgery. Patients carrying a BRCA2 pathogenic variant were excluded (i.e. all patients in our cohort were BRCA2 wild-type). Patients for whom tumor samples were not suitable for TMA analysis were excluded as well.

We first set up an exploratory cohort including two groups of patients: patients with a BRCA1 pathogenic variant constituted the "MT BRCA1" group (n = 19), while those without made up the "WT BRCA1" group (n = 26). Whenever possible, both groups were matched according to their histopathological tumor SBR grade (Scarff Bloom and Richardson), hormone receptor (HR) status, and HER2 status. As expected, patients with BRCA1 pathogenic variant were younger and had a higher frequency of triple negative tumors, compared with WT BRCA1. The aim of this exploratory cohort was to detect "a tendency" between the two groups. ADAT2 was subsequently identified as the best antibody for further exploration (please, refer to "Results" section). To confirm the potential value of this antibody, we set up a validation cohort of 13 additional independent patients, selected following the same methodology as described above.

Tissue microarray (TMA) and immunohistochemistry (IHC)

The slides were incubated on a BenchMark ULTRA automate (Roche Diagnostics, Meylan, France) for deparaffinization and inhibition of endogenous peroxidases in H₂O₂ using the ULTRA view universal DAB kit (Roche, ref:760-500, Meylan, France). The fully automated workflow of the BenchMark ULTRA automate enabled processing of a maximum of 30 slides simultaneously. The IHC staining was therefore performed using batch of antibodies necessary for labelling 30 slides at most.

Heat-induced antigen retrieval was performed in CC1 buffer (pH 8.4) (950-224, Roche) for 64 min on the BenchMark ULTRA automate for rabbit polyclonal antibodies against ADAT2 (ab122280, Abcam), DBF4B (NBP2-13900, Novus Biologicals) and TRIM45 (NBP1-82266, Novus Biologicals). Rabbit polyclonal antibody against CCNL2 (NBP2-47497, Novus Biologicals) did not require antigen retrieval. Anti-ADAT2 was diluted at 1/100, anti-CCNL2 at 1/25, anti-DBF4B at 1/50 and anti-TRIM45 at 1/100.

The slides were incubated with the primary antibody on the BenchMark ULTRA automate for 60 min at 37°C, except for the anti-DBF4B antibody which was incubated for 60 min at room temperature.

Antibody binding was detected using the Ultra view Universal DAB Detection kit (760-500, Roche) for all antibodies: secondary antibodies were incubated for 8 min at 37°C followed by the DAB substrate for another 8 min incubation at 37°C. Regarding the anti-CCNL2 and anti-TRIM45 antibodies, an additional step using the Amplification kit (760-080, Roche) was performed before antibody binding detection using the protocol provided by the supplier.

Sections were counterstained with Hematoxylin (760-2021, Roche), and bluing (760-2037, Roche). Then, the slides were analyzed by a senior pathologist experienced in breast pathology who assessed in the invasive carcinoma, the percentage and intensity of cytoplasmic staining for anti-ADAT2, anti-DBF4B, anti-TRIM45 and the percentage and intensity of both cytoplasmic and nuclear staining for anti-CCNL2.

Cell culture

The human breast carcinoma MDA-MB-231 (basal, ER-, PR-) and MCF-7 (luminal, ER+, PR+) cell lines (available from the ATCC) were maintained in Dulbecco's Modified Eagle's Medium (DMEM) containing 4.5 g/l of glucose supplemented with 10 % fetal calf serum, 100 µg/ml streptomycin and 100 units/ml of penicillin. For MCF-7 cell line, 0.1 mM non-essential amino acids, 1 mM sodium-pyruvate and 0.01 mg/ml recombinant human insulin were added to the medium.

Normal human mammary epithelial cells (HMECs) (available from the ATCC) were maintained in DMEM-F12 supplemented with 0.5 µg/ml hydrocortisone, 10 ng/ml EGF, 0.3 U/ml insulin, 10% fetal calf serum, 100 µg/ml streptomycin and 100 units/ml of penicillin.

Immunoblotting

The MCF-7, MDA-MB-231 or HMEC cells were lysed in lysis buffer A (20 mM Hepes-KOH pH 7.2, 100 mM KCl, 1 mM DTT, 0.5 mM EDTA, 0.5 % NP40, 10 % Glycerol) supplemented with protease inhibitor (Complete EDTA free, Roche). Proteins were subjected to SDS-PAGE, and blotted onto poly-vinylidene difluoride (PVDF) membranes (Immobilon-P, Millipore). Membranes were blocked in Tris-buffered saline solution containing 0.05 % Tween 20 and 5 % non-fat milk and incubated with primary antibodies. Horseradish peroxidase conjugated secondary antibodies (Jackson ImmunoResearch) were used for the detection of immunoreactive proteins by chemiluminescence (ECL, GE Healthcare). We performed this experiment first in MCF-7 cells with which we identified mRNA that were specifically deregulated under BRCA1 deficiency (please, refer to "Results" section). Then, we enlarged our investigation to two additional cell lines: (i) the MDA-MB-231 mammary cell line because it presents opposite characteristics to MCF-7 cell line; (ii) the HMEC cell line because it is representative of normal human mammary epithelial cells.

Isolation of polysomes and total cytoplasmic RNA

Extracts from MCF-7 cells were prepared by lysis at 4°C in extraction buffer (50 mM Tris-HCl, pH 7.4, 25 mM KCl, 5 mM MgCl₂, 250 mM sucrose, 0.7% Nonidet-P40, and 100 µg/ml cycloheximide). After elimination of nuclei and mitochondria by centrifugation, the supernatant (cytoplasmic fraction) was divided in two. One fifth was used to extract total cytoplasmic RNA and the remaining 80 % were layered onto a linear 10-40 % sucrose gradient. A 2 h centrifugation at 250 000 g and at 4°C in a SW41Ti rotor (Beckman) was then performed. The gradients were collected in 14 fractions and absorbance profiles were generated at 254 nm using an ISCO UA-6 detector. Polysomal fractions (fractions 7 to 14) were pooled for further RNA extraction using Trizol LS Reagent as described by the manufacturer (Life technologies).

RNA-binding protein immunoprecipitation (RIP)

The MCF-7 cells were washed with ice-cold PBS and lysed in RIP buffer (25 mM Tris-HCl at pH 7.4, 150 mM KCl, 2 mM EDTA, 0.5 % NP40, 0.5 mM DTT, 200 U/ml RNaseOut, protease and phosphatase inhibitors). Where indicated, a small fraction of the lysate (5 %) was used to prepare total RNA extracts

and this was considered the input. The remaining lysates were then split into equal parts: half was incubated overnight at 4°C with mouse monoclonal anti-BRCA1 antibody (BRCA1) (MS110, Calbiochem) coupled to Dynabeads (Life Technologies) and half was incubated with ChIP Grade mouse control IgG (NR) (ab18413, Abcam) used as control and coupled to Dynabeads. Immunoprecipitates were washed seven times with RIPA buffer (50 mM Tris-HCl at pH 7.4, 150 mM KCl, 0.5 % NP40). The RNA associated with each type of beads, and the total RNA (input) were then eluted with elution buffer (50 mM Tris-HCl at pH 7.4, 5 mM EDTA, 10 mM DTT, 1 % SDS, 200 U/ml RNase Out) and digested with proteinase K. RNA was finally extracted using the RNeasy micro kit (Qiagen).

Microarray analysis

Microarray analysis was performed using a high-density oligonucleotide array (Human Transcriptome Array GeneChip HG-ST 2.0, Affymetrix, Santa Clara, CA, USA). RNA (900 pg) was reverse-transcribed and pre-amplified with the Ovation WTA System (Nugen). cDNA quantification was performed with a nanodrop and quality checked with the Agilent 2100 Bioanalyzer (Agilent technologies, Inc, Palto Alto, CA, USA). Fragmentation and labeling of cDNA (3.6 µg) using Encore Biotin Module kit (Nugen) were performed. Before amplification, spikes of synthetic mRNA at different concentrations were added to all samples; these positive controls were used to ascertain the quality of the process. After final purification, hybridization was performed following Affymetrix protocols (<http://www.affymetrix.com>). The arrays were read with a confocal laser (GeneChip Scanner 3000 7G, Affymetrix). Then CEL files were generated using the Affymetrix GeneChip Command Console (AGCC) software 3.0.

Quantification of A-to-I editing

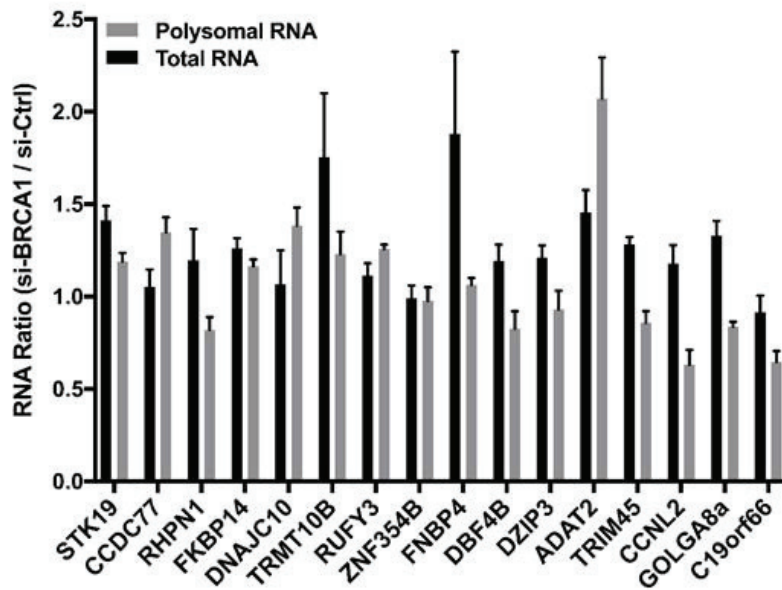
A 50 nucleotide single-end sequencing was done. Analyses were performed in duplicate. 4 samples (2 Control and 2 BRCA1-overexpression samples) were run in one single sequencing lane obtaining approximately 50-70 million reads per sample. However, we were only able to obtain sufficient reads mapping to the anticodon of tRNA substrates of ADAT for tRNA^{Ser}AGA and tRNA^{Val}AAC (at least 100 reads having I34 in the Control samples); therefore, A-to-I editing was exclusively evaluated for those two substrates.

Statistical analysis

Statistical analyses were performed using GraphPad Prism 5.0a (GraphPad Software, Inc). Correlation between fold enrichment obtained by RT-qPCR (Fe) and the fold change obtained by microarray (Fc) or total RNA levels (Inputs) were tested using the non-parametric Spearman r test. For the A-to-I editing experiments, we used a Fisher's Exact test with Benjamini-Hochberg adjusted p-values as previously described [23]. For IHC analyses, enrichment in low and high staining scores was investigated using the Chi-square test. P-values < 0.05 were considered statistically significant. Asterisks denote statistical significance (* p < 0.05; ** p < 0.01; *** p < 0.001).

Standard errors of the means (SEM) shown in figures are the standard deviations for the samples divided by the square root of the sample size. For results shown in each figure, except figures 4f and 4g, at least three independent experiments were performed, ensuring statistical independence of the values in the sample.

a



b

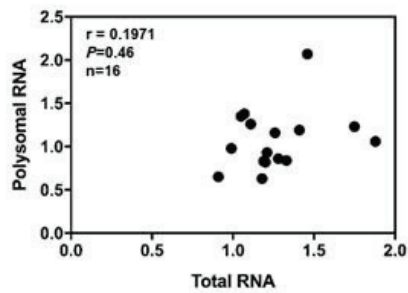


Figure S1: RT-qPCR analyses of differentially translated mRNA upon BRCA1 depletion.

a For each gene, the ratios for polysomal RNA (polyRNA = si-BRCA1/si-Ctrl) (grey) and for total RNA (totRNA = si-BRCA1/si-Ctrl) (black) were calculated. Data are expressed as means \pm SEM. N = 3. **b** Comparison of polysomal and total RNA levels. Mean ratios for polysomal RNA and mean ratios for total RNA were plotted and their correlation was assessed using the Spearman test. No significant correlation was observed.

a

Motif ID	Motif Sequence	Number of mRNA	E-value
1	CGCSCGGC	404	1.8e-098
2	CGCCTGTA	400	1.6e-086
3	CCGCKCG	381	4.6e-077
4	ACCCGTC	394	3.8e-071
5	CGGCCRC	333	5.7e-071
6	CGAGTAGC	364	4.6e-070
7	CCGCKCC	464	7.9e-066
8	CGAGACCA	387	1.3e-060
9	CACRCGC	419	2.4e-059
10	ATCTCGGC	374	2.3e-058
11	CGCCGCS	339	7.9e-053
12	TAAAAATA	440	1.6e-052
13	GAGACGGA	362	7.2e-052
14	AGAATCGC	347	3.4e-051

Legend:

K "Keto" = GT , R "Purine" = AG , S "Strong" = CG

b



Motif	RBP
#2 	SRP14 Signal Recognition Particle 14
#7 	KHSRP KH type-Splicing Regulatory Protein

Figure S2: Search for RNA motifs present in the 488 RNA interacting with BRCA1.

a From Gene Symbol of the 488 genes, the Gene ID (ENSG) were retrieved in Ensembl database, using `ensemldb` R package (Rainer, J et al, 2019, Bioinformatics) and their sequences were downloaded via Ensembl REST API. Using DREME, 14 motifs were identified to be common to at least 333 genes, with statistically sound values (Bailey, TL, 2011, Bioinformatics).

b Using the ATtRACT database (<https://attract.cnic.es/searchmotif>) to determine whether one or several of these motifs were known to interact with some protein, two motifs were associated with known RNA Binding Proteins (RBP).

Gene	Primer F (5' > 3')	Primer R (5' > 3')
ACTIN	CTCTTCCAGCCTTCCTTCCT	TGTTGGCGTACAGGTCTTTG
ADAT2	TGTCGTCAAAGTGGCAAGAG	GATTTTCATCAGGCGGAGAG
C19orf66	GGCTCTGATGAGGAAATTCG	TGGGCATCCAGATCGTTACT
CCDC77	AAAAGTGTTGCCCGTTATGC	GGAAATGCACTGCTCTTGGT
CCNL2	TGCCTGCATTTATCTTGCTG	AAGAAACCAATGGGGACGA
DBF4B	GACTCGGTGCCTCTAAGCAG	ACGTGCAGAATCCTCACTCC
DNAJC10	TGCAATGCAGCATGTTAGAAG	CAATACCAGCAGCAAAGCA
DZIP3	TGCGGTAGACAGTTGGAATG	TGGAGGCAAGGTACTIONCAAGG
FKBP14	TCGCTGCTTTTGTAGGGACT	GCCCCAATCAAAGAAGTGAC
FNBP4	GGAAGCAGCAACATCTACCC	CCGTGTTCTCATCCCAGACT
GOLGA8a	GATACGCATCGGGTAGAGGA	CACCTCAGAGGACACTGCTG
POP4	GCCAAGCTCTTAAAGGCAGA	CTTTCAGGCGGTCTTCTTTG
RHPN1	GCCCAGGAATGTGTGTTTG	CCGGTGCCTAGCCTGTACT
RUFY3	GGTGATGGAGCACTGTCTGA	GCTGTTATCTCTGCGGCTTC
STK19	CTTGGCCTCACCTACCATGT	GCTGAGCAATGATGAGCAGA
TRIM45	GTTGCCGTTGTCCCTAAAGA	TTGATGCAGACCCACACAGT
TRMT10B	TGTGTGAGGATGAATGATGGA	GGTACACAAGGGTTTCCAAGG
ZNF354B	TGGGAGGTGGAGAAAGACAG	TGAAGTTCAGGGTTCATCC

Table S1. Primers used for qPCR.

Bibliographie

- ALAWI, Faizan et al. (déc. 2010). « Correlation of dyskerin expression with active proliferation independent of telomerase ». In : *Head & Neck* 33.7, p. 1041-1051. DOI : [10.1002/hed.21579](https://doi.org/10.1002/hed.21579). URL : <https://doi.org/10.1002/hed.21579>.
- AMSTERDAM, Adam et al. (mai 2004). « Many Ribosomal Protein Genes Are Cancer Genes in Zebrafish ». In : *PLoS Biology* 2.5. Sous la dir. de Derek STEMPLE, e139. DOI : [10.1371/journal.pbio.0020139](https://doi.org/10.1371/journal.pbio.0020139). URL : <https://doi.org/10.1371/journal.pbio.0020139>.
- ANDERS, S., P. T. PYL et W. HUBER (sept. 2014). « HTSeq—a Python framework to work with high-throughput sequencing data ». In : *Bioinformatics* 31.2, p. 166-169. DOI : [10.1093/bioinformatics/btu638](https://doi.org/10.1093/bioinformatics/btu638). URL : <https://doi.org/10.1093/bioinformatics/btu638>.
- ANDREWS, Simon et al. (jan. 2012). *FastQC*. Babraham Institute. Babraham, UK.
- ANGER, Andreas M. et al. (mai 2013). « Structures of the human and *Drosophila* 80S ribosome ». In : *Nature* 497.7447, p. 80-85. DOI : [10.1038/nature12104](https://doi.org/10.1038/nature12104). URL : <https://doi.org/10.1038/nature12104>.
- BABAIAN, Artem et al. (mai 2020). « Loss of m1acp3Ψ ribosomal RNA modification is a major feature of cancer ». In : *Cell Rep.* 31.5, p. 107611.
- BALIGA, Nitin S. et al. (nov. 2004). « Genome sequence of *Halobacterium salinarum* R1 ». In : *Genome Research* 14.11, p. 2221-2234. DOI : [10.1101/gr.2700304](https://doi.org/10.1101/gr.2700304). URL : <https://doi.org/10.1101/gr.2700304>.
- BAN, Nenad et al. (fév. 2014). « A new system for naming ribosomal proteins ». In : *Current Opinion in Structural Biology* 24, p. 165-169. DOI : [10.1016/j.sbi.2014.01.002](https://doi.org/10.1016/j.sbi.2014.01.002). URL : <https://doi.org/10.1016/j.sbi.2014.01.002>.
- BARAKAT, Abdelali et al. (oct. 2001). « The Organization of Cytoplasmic Ribosomal Protein Genes in the Arabidopsis Genome ». In : *Plant Physiology*

- 127.2, p. 398-415. DOI : [10.1104/pp.010265](https://doi.org/10.1104/pp.010265). URL : <https://doi.org/10.1104/pp.010265>.
- BARNA, Maria et al. (nov. 2008). « Suppression of Myc oncogenic activity by ribosomal protein haploinsufficiency ». In : *Nature* 456.7224, p. 971-975. DOI : [10.1038/nature07449](https://doi.org/10.1038/nature07449). URL : <https://doi.org/10.1038/nature07449>.
- BARTA, A et al. (juin 1984). « Identification of a site on 23S ribosomal RNA located at the peptidyl transferase center. » In : *Proceedings of the National Academy of Sciences* 81.12, p. 3607-3611. DOI : [10.1073/pnas.81.12.3607](https://doi.org/10.1073/pnas.81.12.3607). URL : <https://doi.org/10.1073/pnas.81.12.3607>.
- BASHAN, Anat et al. (août 2003). « Ribosomal crystallography : Peptide bond formation and its inhibition ». In : *Biopolymers* 70.1, p. 19-41. DOI : [10.1002/bip.10412](https://doi.org/10.1002/bip.10412). URL : <https://doi.org/10.1002/bip.10412>.
- BEGIK, Oguzhan et al. (mai 2021). « Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing ». In : *Nature Biotechnology* 39.10, p. 1278-1291. DOI : [10.1038/s41587-021-00915-6](https://doi.org/10.1038/s41587-021-00915-6). URL : <https://doi.org/10.1038/s41587-021-00915-6>.
- BELIN, Stéphane et al. (sept. 2009). « Dysregulation of Ribosome Biogenesis and Translational Capacity Is Associated with Tumor Progression of Human Breast Cancer Cells ». In : *PLoS ONE* 4.9. Sous la dir. de Thomas PREISS, e7147. DOI : [10.1371/journal.pone.0007147](https://doi.org/10.1371/journal.pone.0007147). URL : <https://doi.org/10.1371/journal.pone.0007147>.
- BELLODI, Cristian et al. (juill. 2010). « Loss of Function of the Tumor Suppressor DKC1 Perturbs p27 Translation Control and Contributes to Pituitary Tumorigenesis ». In : *Cancer Research* 70.14, p. 6026-6035. DOI : [10.1158/0008-5472.can-09-4730](https://doi.org/10.1158/0008-5472.can-09-4730). URL : <https://doi.org/10.1158/0008-5472.can-09-4730>.
- BEN-SHEM, Adam et al. (déc. 2011). « The Structure of the Eukaryotic Ribosome at 3.0 angstrom resolution ». In : *Science* 334.6062, p. 1524-1529. DOI : [10.1126/science.1212642](https://doi.org/10.1126/science.1212642). URL : <https://doi.org/10.1126/science.1212642>.
- BENJAMINI, Yoav et Yosef HOCHBERG (jan. 1995). « Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing ». In : *Journal of the Royal Statistical Society : Series B (Methodological)* 57.1, p. 289-300. DOI : [10.1111/j.2517-6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x). URL : <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.

- BI, Ningrui et al. (juin 2020). « Identification of 40S ribosomal protein S8 as a novel biomarker for alcohol-associated hepatocellular carcinoma using weighted gene co-expression network analysis ». In : *Oncology Reports* 44.2, p. 611-627. DOI : [10.3892/or.2020.7634](https://doi.org/10.3892/or.2020.7634). URL : <https://doi.org/10.3892/or.2020.7634>.
- BIRKEDAL, Ulf, Mikkel CHRISTENSEN-DALSGAARD et al. (nov. 2014). « Profiling of Ribose Methylations in RNA by High-Throughput Sequencing ». In : *Angewandte Chemie* 127.2, p. 461-465. DOI : [10.1002/ange.201408362](https://doi.org/10.1002/ange.201408362). URL : <https://doi.org/10.1002/ange.201408362>.
- BIRKEDAL, Ulf, Nicolai KROGH et al. (2016). « Methods for Determination of 2'-O-Me in RNA ». In : *Modified Nucleic Acids in Biology and Medicine*. Springer International Publishing, p. 187-205. DOI : [10.1007/978-3-319-34175-0_8](https://doi.org/10.1007/978-3-319-34175-0_8). URL : https://doi.org/10.1007/978-3-319-34175-0_8.
- BOIVIN, Vincent et al. (avr. 2018). « Simultaneous sequencing of coding and noncoding RNA reveals a human transcriptome dominated by a small number of highly expressed noncoding genes ». In : *RNA* 24.7, p. 950-965. DOI : [10.1261/rna.064493.117](https://doi.org/10.1261/rna.064493.117). URL : <https://doi.org/10.1261/rna.064493.117>.
- BOLGER, Anthony M., Marc LOHSE et Bjoern USADEL (avr. 2014). « Trimmomatic : a flexible trimmer for Illumina sequence data ». In : *Bioinformatics* 30.15, p. 2114-2120. DOI : [10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170). URL : <https://doi.org/10.1093/bioinformatics/btu170>.
- BORTOLUZZI, S. et al. (déc. 2001). « Differential expression of genes coding for ribosomal proteins in different human tissues ». In : *Bioinformatics* 17.12, p. 1152-1157. DOI : [10.1093/bioinformatics/17.12.1152](https://doi.org/10.1093/bioinformatics/17.12.1152). URL : <https://doi.org/10.1093/bioinformatics/17.12.1152>.
- BOUCHARD-BOURELLE, Philia et al. (oct. 2019). « snoDB : an interactive database of human snoRNA sequences, abundance and interactions ». In : *Nucleic Acids Research* 48.D1, p. D220-D225. DOI : [10.1093/nar/gkz884](https://doi.org/10.1093/nar/gkz884). URL : <https://doi.org/10.1093/nar/gkz884>.
- BRACHET, Jean (1941). « La localisation des acides pentosenucléiques dans les tissus animaux et les oeufs d'amphibiens en voie de développement ». In.
- BRENNER, S., F. JACOB et M. MESELSON (mai 1961). « An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis ». In : *Nature*.

- BRUNO, Peter M et al. (fév. 2017). « A subset of platinum-containing chemotherapeutic agents kills cells by inducing ribosome biogenesis stress ». In : *Nature Medicine* 23.4, p. 461-471. DOI : [10.1038/nm.4291](https://doi.org/10.1038/nm.4291). URL : <https://doi.org/10.1038/nm.4291>.
- BUCHHAUPT, Markus et al. (fév. 2014). « Partial Methylation at Am100 in 18S rRNA of Baker's Yeast Reveals Ribosome Heterogeneity on the Level of Eukaryotic rRNA Modification ». In : *PLoS ONE* 9.2. Sous la dir. de Thomas PREISS, e89640. DOI : [10.1371/journal.pone.0089640](https://doi.org/10.1371/journal.pone.0089640). URL : <https://doi.org/10.1371/journal.pone.0089640>.
- BYWATER, Megan J. et al. (juill. 2012). « Inhibition of RNA Polymerase I as a Therapeutic Strategy to Promote Cancer-Specific Activation of p53 ». In : *Cancer Cell* 22.1, p. 51-65. DOI : [10.1016/j.ccr.2012.05.019](https://doi.org/10.1016/j.ccr.2012.05.019). URL : <https://doi.org/10.1016/j.ccr.2012.05.019>.
- CAIRNS, C. A. (juin 1998). « p53 is a general repressor of RNA polymerase III transcription ». In : *The EMBO Journal* 17.11, p. 3112-3123. DOI : [10.1093/emboj/17.11.3112](https://doi.org/10.1093/emboj/17.11.3112). URL : <https://doi.org/10.1093/emboj/17.11.3112>.
- CAMPAGNOLI, Maria Francesca et al. (juill. 2008). « RPS19 mutations in patients with Diamond-Blackfan anemia ». In : *Human Mutation* 29.7, p. 911-920. DOI : [10.1002/humu.20752](https://doi.org/10.1002/humu.20752). URL : <https://doi.org/10.1002/humu.20752>.
- CARLILE, Thomas M., Maria F. ROJAS-DURAN et Wendy V. GILBERT (2015). « Pseudo-Seq ». In : *Methods in Enzymology*. Elsevier, p. 219-245. DOI : [10.1016/bs.mie.2015.03.011](https://doi.org/10.1016/bs.mie.2015.03.011). URL : <https://doi.org/10.1016/bs.mie.2015.03.011>.
- CASPERSSON, T (avr. 1939). « Pentose Nucleotides in the Cytoplasm of Growing Tissues ». In : *Nature* 143.3623, p. 602-603.
- (jan. 1941). « Studien über den Eiweißumsatz der Zelle ». In : *Naturwissenschaften* 29.3, p. 33-43.
- CATEZ, Frédéric et al. (jan. 2019). « Ribosome biogenesis : An emerging drug-gable pathway for cancer therapeutics ». In : *Biochemical Pharmacology* 159, p. 74-81. DOI : [10.1016/j.bcp.2018.11.014](https://doi.org/10.1016/j.bcp.2018.11.014). URL : <https://doi.org/10.1016/j.bcp.2018.11.014>.
- CELLERINO, Alessandro et Michele SANGUANINI (2018). *Transcriptome Analysis*. Scuola Normale Superiore. DOI : [10.1007/978-88-7642-642-1](https://doi.org/10.1007/978-88-7642-642-1). URL : <https://doi.org/10.1007/978-88-7642-642-1>.

- CLAUDE, A (jan. 1941). « Particulate components of cytoplasm ». In : *Cold Spring Harb. Symp. Quant. Biol.* 9.0, p. 263-271.
- (mai 1943). « The constitution of protoplasm ». en. In : *Science* 97.2525, p. 451-456.
- CONESA, Ana et al. (jan. 2016). « A survey of best practices for RNA-seq data analysis ». In : *Genome Biology* 17.1. DOI : [10.1186/s13059-016-0881-8](https://doi.org/10.1186/s13059-016-0881-8). URL : <https://doi.org/10.1186/s13059-016-0881-8>.
- COSTA, Lydie Da, Thierry LEBLANC et Narla MOHANDAS (sept. 2020). « Diamond-Blackfan anemia ». In : *Blood* 136.11, p. 1262-1273. DOI : [10.1182/blood.2019000947](https://doi.org/10.1182/blood.2019000947). URL : <https://doi.org/10.1182/blood.2019000947>.
- CRÉANCIER, Laurent et al. (juill. 2000). « Fibroblast Growth Factor 2 Internal Ribosome Entry Site (Ires) Activity Ex Vivo and in Transgenic Mice Reveals a Stringent Tissue-Specific Regulation ». In : *Journal of Cell Biology* 150.1, p. 275-281. DOI : [10.1083/jcb.150.1.275](https://doi.org/10.1083/jcb.150.1.275). URL : <https://doi.org/10.1083/jcb.150.1.275>.
- CRICK, F H (1958). « On protein synthesis ». en. In : *Symp. Soc. Exp. Biol.* 12, p. 138-163.
- CRUZ, Jesús de la, Katrin KARBSTEIN et John L. WOOLFORD (juin 2015). « Functions of Ribosomal Proteins in Assembly of Eukaryotic Ribosomes In Vivo ». In : *Annual Review of Biochemistry* 84.1, p. 93-129. DOI : [10.1146/annurev-biochem-060614-033917](https://doi.org/10.1146/annurev-biochem-060614-033917). URL : <https://doi.org/10.1146/annurev-biochem-060614-033917>.
- DAGOGO-JACK, Ibiayi et Alice T. SHAW (nov. 2017). « Tumour heterogeneity and resistance to cancer therapies ». In : *Nature Reviews Clinical Oncology* 15.2, p. 81-94. DOI : [10.1038/nrclinonc.2017.166](https://doi.org/10.1038/nrclinonc.2017.166). URL : <https://doi.org/10.1038/nrclinonc.2017.166>.
- DAI, Qing et al. (mai 2017). « Nm-seq maps 2'-O-methylation sites in human mRNA with base precision ». In : *Nature Methods* 14.7, p. 695-698. DOI : [10.1038/nmeth.4294](https://doi.org/10.1038/nmeth.4294). URL : <https://doi.org/10.1038/nmeth.4294>.
- DANG, Chi V. (mars 2012). « MYC on the Path to Cancer ». In : *Cell* 149.1, p. 22-35. DOI : [10.1016/j.cell.2012.03.003](https://doi.org/10.1016/j.cell.2012.03.003). URL : <https://doi.org/10.1016/j.cell.2012.03.003>.
- DEGENHARDT, Rory F. et Peta C. BONHAM-SMITH (mars 2008). « Arabidopsis Ribosomal Proteins RPL23aA and RPL23aB Are Differentially Targeted to the Nucleolus and Are Disparately Required for Normal Development

- ». In : *Plant Physiology* 147.1, p. 128-142. DOI : 10.1104/pp.107.111799. URL : <https://doi.org/10.1104/pp.107.111799>.
- DELLOYE-BOURGEOIS, Céline et al. (août 2012). « Nucleolar Localization of a Netrin-1 Isoform Enhances Tumor Cell Proliferation ». In : *Science Signaling* 5.236. DOI : 10.1126/scisignal.2002456. URL : <https://doi.org/10.1126/scisignal.2002456>.
- DERENZINI, M (avr. 2000). « The AgNORs ». In : *Micron* 31.2, p. 117-120. DOI : 10.1016/S0968-4328(99)00067-0. URL : [https://doi.org/10.1016/S0968-4328\(99\)00067-0](https://doi.org/10.1016/S0968-4328(99)00067-0).
- DERENZINI, Massimo, Lorenzo MONTANARO et Davide TRERÉ (mai 2009). « What the nucleolus says to a tumour pathologist ». In : *Histopathology* 54.6, p. 753-762. DOI : 10.1111/j.1365-2559.2008.03168.x. URL : <https://doi.org/10.1111/j.1365-2559.2008.03168.x>.
- DESCHAMPS-FRANCOEUR, Gabrielle et al. (mai 2019). « CoCo : RNA-seq read assignment correction for nested genes and multimapped reads ». In : *Bioinformatics* 35.23. Sous la dir. de Bonnie BERGER, p. 5039-5047. DOI : 10.1093/bioinformatics/btz433. URL : <https://doi.org/10.1093/bioinformatics/btz433>.
- DEUSSER, ELLEN et HEINZ-GÜNTER WITTMANN (août 1972). « Biological Sciences : Ribosomal Proteins : Variation of the Protein Composition in Escherichia coli Ribosomes as Function of Growth Rate ». In : *Nature* 238.5362, p. 269-270. DOI : 10.1038/238269a0. URL : <https://doi.org/10.1038/238269a0>.
- DINMAN, Jonathan D. (mai 2016). « Pathways to Specialized Ribosomes : The Brussels Lecture ». In : *Journal of Molecular Biology* 428.10, p. 2186-2194. DOI : 10.1016/j.jmb.2015.12.021. URL : <https://doi.org/10.1016/j.jmb.2015.12.021>.
- DOBIN, Alexander et al. (oct. 2012). « STAR : ultrafast universal RNA-seq aligner ». In : *Bioinformatics* 29.1, p. 15-21. DOI : 10.1093/bioinformatics/bts635. URL : <https://doi.org/10.1093/bioinformatics/bts635>.
- DRAPTCHINSKAIA, Natalia et al. (fév. 1999). « The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia ». In : *Nature Genetics* 21.2, p. 169-175. DOI : 10.1038/5951. URL : <https://doi.org/10.1038/5951>.
- EBRIGHT, Richard Y. et al. (mars 2020). « Deregulation of ribosomal protein expression and translation promotes breast cancer metastasis ». In : *Science*

- 367.6485, p. 1468-1473. DOI : [10.1126/science.aay0939](https://doi.org/10.1126/science.aay0939). URL : <https://doi.org/10.1126/science.aay0939>.
- EIFERT, Cheryl et R. Scott POWERS (juin 2012). « From cancer genomes to oncogenic drivers, tumour dependencies and therapeutic targets ». In : *Nature Reviews Cancer* 12.8, p. 572-578. DOI : [10.1038/nrc3299](https://doi.org/10.1038/nrc3299). URL : <https://doi.org/10.1038/nrc3299>.
- ERALES, Jenny et al. (nov. 2017). « Evidence for rRNA 2'-O-methylation plasticity : Control of intrinsic translational capabilities of human ribosomes ». In : *Proceedings of the National Academy of Sciences* 114.49, p. 12934-12939. DOI : [10.1073/pnas.1707674114](https://doi.org/10.1073/pnas.1707674114). URL : <https://doi.org/10.1073/pnas.1707674114>.
- EWELS, Philip et al. (juin 2016). « MultiQC : summarize analysis results for multiple tools and samples in a single report ». In : *Bioinformatics* 32.19, p. 3047-3048. DOI : [10.1093/bioinformatics/btw354](https://doi.org/10.1093/bioinformatics/btw354). URL : <https://doi.org/10.1093/bioinformatics/btw354>.
- FAFARD-COUTURE, Étienne et al. (juin 2021). « Annotation of snoRNA abundance across human tissues reveals complex snoRNA-host gene relationships ». In : *Genome Biology* 22.1. DOI : [10.1186/s13059-021-02391-2](https://doi.org/10.1186/s13059-021-02391-2). URL : <https://doi.org/10.1186/s13059-021-02391-2>.
- FANCELLO, Laura et al. (jan. 2017). « The ribosomal protein gene RPL5 is a haploinsufficient tumor suppressor in multiple cancer types ». In : *Oncotarget* 8.9, p. 14462-14478. DOI : [10.18632/oncotarget.14895](https://doi.org/10.18632/oncotarget.14895). URL : <https://doi.org/10.18632/oncotarget.14895>.
- FARRAR, Jason E. et al. (sept. 2008). « Abnormalities of the large ribosomal subunit protein, Rpl35a, in Diamond-Blackfan anemia ». In : *Blood* 112.5, p. 1582-1592. DOI : [10.1182/blood-2008-02-140012](https://doi.org/10.1182/blood-2008-02-140012). URL : <https://doi.org/10.1182/blood-2008-02-140012>.
- FATIMA, Grace, Ganeshan MATHAN et Vijay KUMAR (avr. 2012). « The HBx protein of hepatitis B virus regulates the expression, intracellular distribution and functions of ribosomal protein S27a ». In : *Journal of General Virology* 93.4, p. 706-715. DOI : [10.1099/vir.0.035691-0](https://doi.org/10.1099/vir.0.035691-0). URL : <https://doi.org/10.1099/vir.0.035691-0>.
- FAUCHER-GIGUÈRE, Laurence et al. (jan. 2022). « High-grade ovarian cancer associated H/ACA snoRNAs promote cancer cell proliferation and survival ». In : *NAR Cancer* 4.1. DOI : [10.1093/narcan/zcab050](https://doi.org/10.1093/narcan/zcab050). URL : <https://doi.org/10.1093/narcan/zcab050>.

- FERLAY, Jacques et al. (avr. 2021). « Cancer statistics for the year 2020 : An overview ». In : *International Journal of Cancer* 149.4, p. 778-789. DOI : [10.1002/ijc.33588](https://doi.org/10.1002/ijc.33588). URL : <https://doi.org/10.1002/ijc.33588>.
- FERRETTI, Max B et al. (juill. 2017). « Rps26 directs mRNA-specific translation by recognition of Kozak sequence elements ». In : *Nature Structural & Molecular Biology* 24.9, p. 700-707. DOI : [10.1038/nsmb.3442](https://doi.org/10.1038/nsmb.3442). URL : <https://doi.org/10.1038/nsmb.3442>.
- FIELD, Katharine G. et al. (fév. 1988). « Molecular Phylogeny of the Animal Kingdom ». In : *Science* 239.4841, p. 748-753. DOI : [10.1126/science.3277277](https://doi.org/10.1126/science.3277277). URL : <https://doi.org/10.1126/science.3277277>.
- GALVANIN, Adeline et al. (déc. 2018). « Mapping and Quantification of tRNA 2'-O-Methylation by RiboMethSeq ». In : *Epitranscriptomics*. Springer New York, p. 273-295. DOI : [10.1007/978-1-4939-8808-2_21](https://doi.org/10.1007/978-1-4939-8808-2_21). URL : https://doi.org/10.1007/978-1-4939-8808-2_21.
- GAZDA, Hanna T. et al. (déc. 2008). « Ribosomal Protein L5 and L11 Mutations Are Associated with Cleft Palate and Abnormal Thumbs in Diamond-Blackfan Anemia Patients ». In : *The American Journal of Human Genetics* 83.6, p. 769-780. DOI : [10.1016/j.ajhg.2008.11.004](https://doi.org/10.1016/j.ajhg.2008.11.004). URL : <https://doi.org/10.1016/j.ajhg.2008.11.004>.
- GE, Junhui et Yi-Tao YU (avr. 2013). « RNA pseudouridylation : new insights into an old modification ». In : *Trends in Biochemical Sciences* 38.4, p. 210-218. DOI : [10.1016/j.tibs.2013.01.002](https://doi.org/10.1016/j.tibs.2013.01.002). URL : <https://doi.org/10.1016/j.tibs.2013.01.002>.
- GENUTH, Naomi R. et Maria BARNA (mai 2018). « Heterogeneity and specialized functions of translation machinery : from genes to organisms ». In : *Nature Reviews Genetics* 19.7, p. 431-452. DOI : [10.1038/s41576-018-0008-z](https://doi.org/10.1038/s41576-018-0008-z). URL : <https://doi.org/10.1038/s41576-018-0008-z>.
- GIBBONS, John G. et al. (jan. 2015). « Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes ». In : *Proceedings of the National Academy of Sciences* 112.8, p. 2485-2490. DOI : [10.1073/pnas.1416878112](https://doi.org/10.1073/pnas.1416878112). URL : <https://doi.org/10.1073/pnas.1416878112>.
- GILBERT, Wendy V. (mars 2011). « Functional specialization of ribosomes? ». In : *Trends in Biochemical Sciences* 36.3, p. 127-132. DOI : [10.1016/j.tibs.2010.12.002](https://doi.org/10.1016/j.tibs.2010.12.002). URL : <https://doi.org/10.1016/j.tibs.2010.12.002>.
- GIRAUD, Stéphane et al. (fév. 2001). « Translation Initiation of the Insulin-like Growth Factor I Receptor mRNA Is Mediated by an Internal Ribosome

- Entry Site ». In : *Journal of Biological Chemistry* 276.8, p. 5668-5675. DOI : 10.1074/jbc.m005928200. URL : <https://doi.org/10.1074/jbc.m005928200>.
- GOH, Wilson Wen Bin, Wei WANG et Limsoon WONG (juin 2017). « Why Batch Effects Matter in Omics Data, and How to Avoid Them ». In : *Trends in Biotechnology* 35.6, p. 498-507. DOI : 10.1016/j.tibtech.2017.02.012. URL : <https://doi.org/10.1016/j.tibtech.2017.02.012>.
- GOUDARZI, KAVEH M. et MIKAEL S. LINDSTRÖM (fév. 2016). « Role of ribosomal protein mutations in tumor development (Review) ». In : *International Journal of Oncology* 48.4, p. 1313-1324. DOI : 10.3892/ijo.2016.3387. URL : <https://doi.org/10.3892/ijo.2016.3387>.
- GRESSNER, Axel M. et Ira G. WOOL (nov. 1974). « The Phosphorylation of Liver Ribosomal Proteins in Vivo ». In : *Journal of Biological Chemistry* 249.21, p. 6917-6925. DOI : 10.1016/s0021-9258(19)42145-5. URL : [https://doi.org/10.1016/s0021-9258\(19\)42145-5](https://doi.org/10.1016/s0021-9258(19)42145-5).
- GROS, FRANCOIS et al. (mai 1961). « Unstable Ribonucleic Acid Revealed by Pulse Labelling of Escherichia Coli ». In : *Nature* 190.4776, p. 581-585. DOI : 10.1038/190581a0. URL : <https://doi.org/10.1038/190581a0>.
- GUERRIER-TAKADA, Cecilia et al. (déc. 1983). « The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme ». In : *Cell* 35.3, p. 849-857. DOI : 10.1016/0092-8674(83)90117-4. URL : [https://doi.org/10.1016/0092-8674\(83\)90117-4](https://doi.org/10.1016/0092-8674(83)90117-4).
- HANAHAN, Douglas (jan. 2022). « Hallmarks of Cancer : New Dimensions ». In : *Cancer Discovery* 12.1, p. 31-46. DOI : 10.1158/2159-8290.cd-21-1059. URL : <https://doi.org/10.1158/2159-8290.cd-21-1059>.
- HANAHAN, Douglas et Robert A WEINBERG (jan. 2000). « The Hallmarks of Cancer ». In : *Cell* 100.1, p. 57-70. DOI : 10.1016/s0092-8674(00)81683-9. URL : [https://doi.org/10.1016/s0092-8674\(00\)81683-9](https://doi.org/10.1016/s0092-8674(00)81683-9).
- (mars 2011). « Hallmarks of Cancer : The Next Generation ». In : *Cell* 144.5, p. 646-674. DOI : 10.1016/j.cell.2011.02.013. URL : <https://doi.org/10.1016/j.cell.2011.02.013>.
- HART, Steven N. et al. (déc. 2013). « Calculating Sample Size Estimates for RNA Sequencing Data ». In : *Journal of Computational Biology* 20.12, p. 970-978. DOI : 10.1089/cmb.2012.0283. URL : <https://doi.org/10.1089/cmb.2012.0283>.
- HEBRAS, Jade et al. (sept. 2019). « Developmental changes of rRNA ribose methylations in the mouse ». In : *RNA Biology* 17.1, p. 150-164. DOI : 10.

- 1080 / 15476286 . 2019 . 1670598. URL : <https://doi.org/10.1080/15476286.2019.1670598>.
- HEISS, Nina S. et al. (mai 1998). « X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions ». In : *Nature Genetics* 19.1, p. 32-38. DOI : [10.1038/ng0598-32](https://doi.org/10.1038/ng0598-32). URL : <https://doi.org/10.1038/ng0598-32>.
- HEISSENBERGER, Clemens et al. (nov. 2019). « Loss of the ribosomal RNA methyltransferase NSUN5 impairs global protein synthesis and normal growth ». In : *Nucleic Acids Research* 47.22, p. 11807-11825. DOI : [10.1093/nar/gkz1043](https://doi.org/10.1093/nar/gkz1043). URL : <https://doi.org/10.1093/nar/gkz1043>.
- HERSHEY, J. W. B., N. SONENBERG et M. B. MATHEWS (déc. 2012). « Principles of Translational Control : An Overview ». In : *Cold Spring Harbor Perspectives in Biology* 4.12, a011528-a011528. DOI : [10.1101/cshperspect.a011528](https://doi.org/10.1101/cshperspect.a011528). URL : <https://doi.org/10.1101/cshperspect.a011528>.
- HUANG, Chi-Jung et al. (juin 2013). « Ribosomal Protein S27-Like in Colorectal Cancer : A Candidate for Predicting Prognoses ». In : *PLoS ONE* 8.6. Sous la dir. de Syed A. AZIZ, e67043. DOI : [10.1371/journal.pone.0067043](https://doi.org/10.1371/journal.pone.0067043). URL : <https://doi.org/10.1371/journal.pone.0067043>.
- HUEZ, Isabelle et al. (nov. 1998). « Two Independent Internal Ribosome Entry Sites Are Involved in Translation Initiation of Vascular Endothelial Growth Factor mRNA ». In : *Molecular and Cellular Biology* 18.11, p. 6178-6190. DOI : [10.1128/mcb.18.11.6178](https://doi.org/10.1128/mcb.18.11.6178). URL : <https://doi.org/10.1128/mcb.18.11.6178>.
- INCARNATO, Danny et al. (sept. 2016). « High-throughput single-base resolution mapping of RNA 2'-O-methylated residues ». In : *Nucleic Acids Research* 45.3, p. 1433-1441. DOI : [10.1093/nar/gkw810](https://doi.org/10.1093/nar/gkw810). URL : <https://doi.org/10.1093/nar/gkw810>.
- JAAFAR, Mariam et al. (juill. 2021). « 2'-O-Ribose Methylation of Ribosomal RNAs : Natural Diversity in Living Organisms, Biological Processes, and Diseases ». In : *Cells* 10.8, p. 1948. DOI : [10.3390/cells10081948](https://doi.org/10.3390/cells10081948). URL : <https://doi.org/10.3390/cells10081948>.
- JACK, Karen et al. (nov. 2011). « rRNA Pseudouridylation Defects Affect Ribosomal Ligand Binding and Translational Fidelity from Yeast to Human Cells ». In : *Molecular Cell* 44.4, p. 660-666. DOI : [10.1016/j.molcel.2011.09.017](https://doi.org/10.1016/j.molcel.2011.09.017). URL : <https://doi.org/10.1016/j.molcel.2011.09.017>.

- JACKSON, Richard J., Christopher U.T. HELLEN et Tatyana V. PESTOVA (2012). « Termination and post-termination events in eukaryotic translation ». In : *Fidelity and Quality Control in Gene Expression*. Elsevier, p. 45-93. DOI : [10.1016/b978-0-12-386497-0.00002-5](https://doi.org/10.1016/b978-0-12-386497-0.00002-5). URL : <https://doi.org/10.1016/b978-0-12-386497-0.00002-5>.
- JACOB, François et Jacques MONOD (juin 1961). « Genetic regulatory mechanisms in the synthesis of proteins ». In : *Journal of Molecular Biology* 3.3, p. 318-356. DOI : [10.1016/s0022-2836\(61\)80072-7](https://doi.org/10.1016/s0022-2836(61)80072-7). URL : [https://doi.org/10.1016/s0022-2836\(61\)80072-7](https://doi.org/10.1016/s0022-2836(61)80072-7).
- JANG, S K et al. (août 1988). « A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation ». In : *Journal of Virology* 62.8, p. 2636-2643. DOI : [10.1128/jvi.62.8.2636-2643.1988](https://doi.org/10.1128/jvi.62.8.2636-2643.1988). URL : <https://doi.org/10.1128/jvi.62.8.2636-2643.1988>.
- JANIN, Maxime et al. (août 2019). « Epigenetic loss of RNA-methyltransferase NSUN5 in glioma targets ribosomes to drive a stress adaptive translational program ». In : *Acta Neuropathologica* 138.6, p. 1053-1074. DOI : [10.1007/s00401-019-02062-4](https://doi.org/10.1007/s00401-019-02062-4). URL : <https://doi.org/10.1007/s00401-019-02062-4>.
- JANSSON, Martin D. et al. (nov. 2021). « Regulation of translation by site-specific ribosomal RNA methylation ». In : *Nature Structural & Molecular Biology* 28.11, p. 889-899. DOI : [10.1038/s41594-021-00669-4](https://doi.org/10.1038/s41594-021-00669-4). URL : <https://doi.org/10.1038/s41594-021-00669-4>.
- JENUWEIN, Thomas et C. David ALLIS (août 2001). « Translating the Histone Code ». In : *Science* 293.5532, p. 1074-1080. DOI : [10.1126/science.1063127](https://doi.org/10.1126/science.1063127). URL : <https://doi.org/10.1126/science.1063127>.
- JHA, Prerana et al. (juin 2015). « Genome-wide small noncoding RNA profiling of pediatric high-grade gliomas reveals deregulation of several miRNAs, identifies downregulation of snoRNA cluster HBII-52 and delineates H3F3A and TP53 mutant-specific miRNAs and snoRNAs ». In : *International Journal of Cancer* 137.10, p. 2343-2353. DOI : [10.1002/ijc.29610](https://doi.org/10.1002/ijc.29610). URL : <https://doi.org/10.1002/ijc.29610>.
- JIANG, Z. et al. (2020). « High expression of NSUN5 promotes cell proliferation via cell cycle regulation in colorectal cancer ». In : *Am J Transl Res* 12.7, p. 3858-3870.

- JOHNSON, W. Evan, Cheng LI et Ariel RABINOVIC (avr. 2006). « Adjusting batch effects in microarray expression data using empirical Bayes methods ». In : *Biostatistics* 8.1, p. 118-127. DOI : [10.1093/biostatistics/kxj037](https://doi.org/10.1093/biostatistics/kxj037). URL : <https://doi.org/10.1093/biostatistics/kxj037>.
- JORJANI, Hadi et al. (mai 2016). « An updated human snoRNAome ». In : *Nucleic Acids Research* 44.11, p. 5068-5082. DOI : [10.1093/nar/gkw386](https://doi.org/10.1093/nar/gkw386). URL : <https://doi.org/10.1093/nar/gkw386>.
- KAMPEN, Kim R. et al. (juin 2018). « The ribosomal RPL10 R98S mutation drives IRES-dependent BCL-2 translation in T-ALL ». In : *Leukemia* 33.2, p. 319-332. DOI : [10.1038/s41375-018-0176-z](https://doi.org/10.1038/s41375-018-0176-z). URL : <https://doi.org/10.1038/s41375-018-0176-z>.
- KAWAHATA, Takuto et al. (jan. 2020). « Involvement of ribosomal protein L11 expression in sensitivity of gastric cancer against 5-FU ». In : *Oncology Letters*. DOI : [10.3892/ol.2020.11352](https://doi.org/10.3892/ol.2020.11352). URL : <https://doi.org/10.3892/ol.2020.11352>.
- KEARSE, M. G., A. S. CHEN et V. C. WARE (déc. 2010). « Expression of ribosomal protein L22e family members in *Drosophila melanogaster* : rpL22-like is differentially expressed and alternatively spliced ». In : *Nucleic Acids Research* 39.7, p. 2701-2716. DOI : [10.1093/nar/gkq1218](https://doi.org/10.1093/nar/gkq1218). URL : <https://doi.org/10.1093/nar/gkq1218>.
- KEERSMAECKER, Kim De et al. (déc. 2012). « Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia ». In : *Nature Genetics* 45.2, p. 186-190. DOI : [10.1038/ng.2508](https://doi.org/10.1038/ng.2508). URL : <https://doi.org/10.1038/ng.2508>.
- KELLER, E B et J W LITTLEFIELD (jan. 1957). « Incorporation of C14-amino acids into ribonucleoprotein particles from the Ehrlich mouse ascites tumor ». en. In : *J Biol Chem* 224.1, p. 13-30.
- KELLIS, Manolis, Bruce W. BIRREN et Eric S. LANDER (mars 2004). « Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae* ». In : *Nature* 428.6983, p. 617-624. DOI : [10.1038/nature02424](https://doi.org/10.1038/nature02424). URL : <https://doi.org/10.1038/nature02424>.
- KHATTER, H. et al. (juin 2015). *STRUCTURE OF THE HUMAN 80S RIBOSOME*. DOI : [10.2210/pdb4ug0/pdb](https://doi.org/10.2210/pdb4ug0/pdb). URL : <https://doi.org/10.2210/pdb4ug0/pdb>.

- KHATTER, Heena et al. (avr. 2015). « Structure of the human 80S ribosome ». In : *Nature* 520.7549, p. 640-645. DOI : [10.1038/nature14427](https://doi.org/10.1038/nature14427). URL : <https://doi.org/10.1038/nature14427>.
- KHOURY, Wendy El et Zeina NASR (déc. 2021). « Deregulation of ribosomal proteins in human cancers ». In : *Bioscience Reports* 41.12. DOI : [10.1042/bsr20211577](https://doi.org/10.1042/bsr20211577). URL : <https://doi.org/10.1042/bsr20211577>.
- KIM, Mi-Young et al. (juill. 2001). « Expression profile of nine novel genes differentially expressed in hepatitis B virus-associated hepatocellular carcinomas ». In : *Oncogene* 20.33, p. 4568-4575. DOI : [10.1038/sj.onc.1204626](https://doi.org/10.1038/sj.onc.1204626). URL : <https://doi.org/10.1038/sj.onc.1204626>.
- KLINGE, Sebastian et John L. WOOLFORD (nov. 2018). « Ribosome assembly coming into focus ». In : *Nature Reviews Molecular Cell Biology* 20.2, p. 116-131. DOI : [10.1038/s41580-018-0078-y](https://doi.org/10.1038/s41580-018-0078-y). URL : <https://doi.org/10.1038/s41580-018-0078-y>.
- KNIGHT, Zachary A. et al. (nov. 2012). « Molecular Profiling of Activated Neurons by Phosphorylated Ribosome Capture ». In : *Cell* 151.5, p. 1126-1137. DOI : [10.1016/j.cell.2012.10.039](https://doi.org/10.1016/j.cell.2012.10.039). URL : <https://doi.org/10.1016/j.cell.2012.10.039>.
- KOCH, Clarissa M. et al. (août 2018). « A Beginner's Guide to Analysis of RNA Sequencing Data ». In : *American Journal of Respiratory Cell and Molecular Biology* 59.2, p. 145-157. DOI : [10.1165/rcmb.2017-0430tr](https://doi.org/10.1165/rcmb.2017-0430tr). URL : <https://doi.org/10.1165/rcmb.2017-0430tr>.
- KOMILI, Suzanne et al. (nov. 2007). « Functional Specificity among Ribosomal Proteins Regulates Gene Expression ». In : *Cell* 131.3, p. 557-571. DOI : [10.1016/j.cell.2007.08.037](https://doi.org/10.1016/j.cell.2007.08.037). URL : <https://doi.org/10.1016/j.cell.2007.08.037>.
- KONDRASHOV, Nadya et al. (avr. 2011). « Ribosome-Mediated Specificity in Hox mRNA Translation and Vertebrate Tissue Patterning ». In : *Cell* 145.3, p. 383-397. DOI : [10.1016/j.cell.2011.03.028](https://doi.org/10.1016/j.cell.2011.03.028). URL : <https://doi.org/10.1016/j.cell.2011.03.028>.
- KROGH, Nicolai, Fazila ASMAR et al. (déc. 2020). « Profiling of ribose methylations in ribosomal RNA from diffuse large B-cell lymphoma patients for evaluation of ribosomes as drug targets ». In : *NAR Cancer* 2.4. DOI : [10.1093/narcan/zcaa035](https://doi.org/10.1093/narcan/zcaa035). URL : <https://doi.org/10.1093/narcan/zcaa035>.
- KROGH, Nicolai, Martin D. JANSSON et al. (juin 2016). « Profiling of 2*i*-iO/i-Me in human rRNA reveals a subset of fractionally modified positions and

- provides evidence for ribosome heterogeneity ». In : *Nucleic Acids Research* 44.16, p. 7884-7895. DOI : [10.1093/nar/gkw482](https://doi.org/10.1093/nar/gkw482). URL : <https://doi.org/10.1093/nar/gkw482>.
- KROGH, Nicolai et Henrik NIELSEN (mars 2019). « Sequencing-based methods for detection and quantitation of ribose methylations in RNA ». In : *Methods* 156, p. 5-15. DOI : [10.1016/j.ymeth.2018.11.017](https://doi.org/10.1016/j.ymeth.2018.11.017). URL : <https://doi.org/10.1016/j.ymeth.2018.11.017>.
- KRUGER, Kelly et al. (nov. 1982). « Self-splicing RNA : Autoexcision and auto-cyclization of the ribosomal RNA intervening sequence of tetrahymena ». In : *Cell* 31.1, p. 147-157. DOI : [10.1016/0092-8674\(82\)90414-7](https://doi.org/10.1016/0092-8674(82)90414-7). URL : [https://doi.org/10.1016/0092-8674\(82\)90414-7](https://doi.org/10.1016/0092-8674(82)90414-7).
- KWAN, Thaddaeus et Sunnie R. THOMPSON (juin 2018). « Noncanonical Translation Initiation in Eukaryotes ». In : *Cold Spring Harbor Perspectives in Biology* 11.4, a032672. DOI : [10.1101/cshperspect.a032672](https://doi.org/10.1101/cshperspect.a032672). URL : <https://doi.org/10.1101/cshperspect.a032672>.
- LAHENS, Nicholas F. et al. (août 2017). « A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression ». In : *BMC Genomics* 18.1. DOI : [10.1186/s12864-017-4011-0](https://doi.org/10.1186/s12864-017-4011-0). URL : <https://doi.org/10.1186/s12864-017-4011-0>.
- LANE, B.G. et T. TAMAOKI (juill. 1967). « Studies of the chain termini and alkali-stable dinucleotide sequences in 16 s and 28 s ribosomal RNA from L Cells ». In : *Journal of Molecular Biology* 27.2, p. 335-348. DOI : [10.1016/0022-2836\(67\)90024-1](https://doi.org/10.1016/0022-2836(67)90024-1). URL : [https://doi.org/10.1016/0022-2836\(67\)90024-1](https://doi.org/10.1016/0022-2836(67)90024-1).
- LANE, D J et al. (oct. 1985). « Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. » In : *Proceedings of the National Academy of Sciences* 82.20, p. 6955-6959. DOI : [10.1073/pnas.82.20.6955](https://doi.org/10.1073/pnas.82.20.6955). URL : <https://doi.org/10.1073/pnas.82.20.6955>.
- LANGMEAD, Ben et al. (2009). « Ultrafast and memory-efficient alignment of short DNA sequences to the human genome ». In : *Genome Biology* 10.3, R25. DOI : [10.1186/gb-2009-10-3-r25](https://doi.org/10.1186/gb-2009-10-3-r25). URL : <https://doi.org/10.1186/gb-2009-10-3-r25>.
- LEEK, Jeffrey T. (oct. 2014). « svaseq : removing batch effects and other unwanted noise from sequencing data ». In : *Nucleic Acids Research* 42.21, e161-e161. DOI : [10.1093/nar/gku864](https://doi.org/10.1093/nar/gku864). URL : <https://doi.org/10.1093/nar/gku864>.

- LESTRADE, L. (jan. 2006). « snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs ». In : *Nucleic Acids Research* 34.90001, p. D158-D162. DOI : [10.1093/nar/gkj002](https://doi.org/10.1093/nar/gkj002). URL : <https://doi.org/10.1093/nar/gkj002>.
- LI, H. et R. DURBIN (mai 2009). « Fast and accurate short read alignment with Burrows-Wheeler transform ». In : *Bioinformatics* 25.14, p. 1754-1760. DOI : [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324). URL : <https://doi.org/10.1093/bioinformatics/btp324>.
- LIAO, Y., G. K. SMYTH et W. SHI (nov. 2013). « featureCounts : an efficient general purpose program for assigning sequence reads to genomic features ». In : *Bioinformatics* 30.7, p. 923-930. DOI : [10.1093/bioinformatics/btt656](https://doi.org/10.1093/bioinformatics/btt656). URL : <https://doi.org/10.1093/bioinformatics/btt656>.
- LIM, Keo-Heun et al. (août 2011). « RPS3a Over-Expressed in HBV-Associated Hepatocellular Carcinoma Enhances the HBx-Induced NF-B Signaling via Its Novel Chaperoning Function ». In : *PLoS ONE* 6.8. Sous la dir. de Dong-Yan JIN, e22258. DOI : [10.1371/journal.pone.0022258](https://doi.org/10.1371/journal.pone.0022258). URL : <https://doi.org/10.1371/journal.pone.0022258>.
- LIN, Xinyao et al. (mai 2019). « RNA m6A methylation regulates the epithelial mesenchymal transition of cancer cells and translation of Snail ». In : *Nature Communications* 10.1. DOI : [10.1038/s41467-019-09865-9](https://doi.org/10.1038/s41467-019-09865-9). URL : <https://doi.org/10.1038/s41467-019-09865-9>.
- LITTLEFIELD, John W et al. (nov. 1955). « STUDIES ON CYTOPLASMIC RIBONUCLEOPROTEIN PARTICLES FROM THE LIVER OF THE RAT ». In : *Journal of Biological Chemistry* 217.1, p. 111-124.
- LIU, Bei et al. (août 2012). « Dyskerin Overexpression in Human Hepatocellular Carcinoma Is Associated with Advanced Clinical Stage and Poor Patient Prognosis ». In : *PLoS ONE* 7.8. Sous la dir. de Paul J. GALARDY, e43147. DOI : [10.1371/journal.pone.0043147](https://doi.org/10.1371/journal.pone.0043147). URL : <https://doi.org/10.1371/journal.pone.0043147>.
- LOCATI, Mauro D., Johanna F.B. PAGANO, Wim A. ENSINK et al. (déc. 2016). « Linking maternal and somatic 5S rRNA types with different sequence-specific non-LTR retrotransposons ». In : *RNA* 23.4, p. 446-456. DOI : [10.1261/rna.059642.116](https://doi.org/10.1261/rna.059642.116). URL : <https://doi.org/10.1261/rna.059642.116>.
- LOCATI, Mauro D., Johanna F.B. PAGANO, Geneviève GIRARD et al. (mai 2017). « Expression of distinct maternal and somatic 5.8S, 18S, and 28S rRNA types during zebrafish development ». In : *RNA* 23.8, p. 1188-1199.

- DOI : [10.1261/rna.061515.117](https://doi.org/10.1261/rna.061515.117). URL : <https://doi.org/10.1261/rna.061515.117>.
- LONG, Flora Nguyen Van et al. (mai 2022). « Low level of Fibrillarin, a ribosome biogenesis factor, is a new independent marker of poor outcome in breast cancer ». In : *BMC Cancer* 22.1. DOI : [10.1186/s12885-022-09552-x](https://doi.org/10.1186/s12885-022-09552-x). URL : <https://doi.org/10.1186/s12885-022-09552-x>.
- LOPES, Alexandra M et al. (2010). « The human RPS4 paralogue on Yq11.223 encodes a structurally conserved ribosomal protein and is preferentially expressed during spermatogenesis ». In : *BMC Molecular Biology* 11.1, p. 33. DOI : [10.1186/1471-2199-11-33](https://doi.org/10.1186/1471-2199-11-33). URL : <https://doi.org/10.1186/1471-2199-11-33>.
- MACHNICKA, Magdalena A. et al. (oct. 2012). « MODOMICS : a database of RNA modification pathways—2013 update ». In : *Nucleic Acids Research* 41.D1, p. D262-D267. DOI : [10.1093/nar/gks1007](https://doi.org/10.1093/nar/gks1007). URL : <https://doi.org/10.1093/nar/gks1007>.
- MADEN, B.Edward H. (nov. 2001). « Mapping 2'-O-Methyl Groups in Ribosomal RNA ». In : *Methods* 25.3, p. 374-382. DOI : [10.1006/meth.2001.1250](https://doi.org/10.1006/meth.2001.1250). URL : <https://doi.org/10.1006/meth.2001.1250>.
- MAGEENEY, Catherine M. et Vassie C. WARE (août 2019). « Specialized eRpL22 paralogue-specific ribosomes regulate specific mRNA translation in spermatogenesis in *Drosophila melanogaster* ». In : *Molecular Biology of the Cell* 30.17. Sous la dir. d'A. Gregory MATERA, p. 2240-2253. DOI : [10.1091/mbc.e19-02-0086](https://doi.org/10.1091/mbc.e19-02-0086). URL : <https://doi.org/10.1091/mbc.e19-02-0086>.
- MARCEL, V, F CATEZ et J-J DIAZ (mars 2015). « p53, a translational regulator : contribution to its tumour-suppressor activity ». In : *Oncogene* 34.44, p. 5513-5523. DOI : [10.1038/onc.2015.25](https://doi.org/10.1038/onc.2015.25). URL : <https://doi.org/10.1038/onc.2015.25>.
- MARCEL, Virginie, Sandra E. GHAYAD et al. (sept. 2013). « p53 Acts as a Safeguard of Translational Control by Regulating Fibrillarin and rRNA Methylation in Cancer ». In : *Cancer Cell* 24.3, p. 318-330. DOI : [10.1016/j.ccr.2013.08.013](https://doi.org/10.1016/j.ccr.2013.08.013). URL : <https://doi.org/10.1016/j.ccr.2013.08.013>.
- MARCEL, Virginie, Janice KIELBASSA et al. (déc. 2020). « Ribosomal RNA 2'-O-methylation as a novel layer of inter-tumour heterogeneity in breast cancer ». In : *NAR Cancer* 2.4. DOI : [10.1093/narcan/zcaa036](https://doi.org/10.1093/narcan/zcaa036). URL : <https://doi.org/10.1093/narcan/zcaa036>.

- MARCHAND, Virginie, Florence BLANLOEIL-OILLO et al. (juin 2016). « Illumina-based RibomethSeq approach for mapping of 2'-O-Me residues in RNA ». In : *Nucleic Acids Research* 44.16, e135-e135. DOI : [10.1093/nar/gkw547](https://doi.org/10.1093/nar/gkw547). URL : <https://doi.org/10.1093/nar/gkw547>.
- MARCHAND, Virginie, Florian PICHOT et al. (sept. 2020). « HydraPsiSeq : a method for systematic and quantitative mapping of pseudouridines in RNA ». In : *Nucleic Acids Research* 48.19, e110-e110. DOI : [10.1093/nar/gkaa769](https://doi.org/10.1093/nar/gkaa769). URL : <https://doi.org/10.1093/nar/gkaa769>.
- MARTIN, Marcel (mai 2011). « Cutadapt removes adapter sequences from high-throughput sequencing reads ». In : *EMBNET.JOURNAL* 17.1, p. 10. DOI : [10.14806/ej.17.1.200](https://doi.org/10.14806/ej.17.1.200). URL : <https://doi.org/10.14806/ej.17.1.200>.
- MARTINEAU, Yvan et al. (sept. 2004). « Internal Ribosome Entry Site Structural Motifs Conserved among Mammalian Fibroblast Growth Factor 1 Alternatively Spliced mRNAs ». In : *Molecular and Cellular Biology* 24.17, p. 7622-7635. DOI : [10.1128/mcb.24.17.7622-7635.2004](https://doi.org/10.1128/mcb.24.17.7622-7635.2004). URL : <https://doi.org/10.1128/mcb.24.17.7622-7635.2004>.
- MARYGOLD, Steven J et al. (2007). « The ribosomal protein genes and Minute loci of *Drosophila melanogaster* ». In : *Genome Biology* 8.10, R216. DOI : [10.1186/gb-2007-8-10-r216](https://doi.org/10.1186/gb-2007-8-10-r216). URL : <https://doi.org/10.1186/gb-2007-8-10-r216>.
- MAURO, Vincent P. et Gerald M. EDELMAN (sept. 2002). « The ribosome filter hypothesis ». In : *Proceedings of the National Academy of Sciences* 99.19, p. 12031-12036. DOI : [10.1073/pnas.192442499](https://doi.org/10.1073/pnas.192442499). URL : <https://doi.org/10.1073/pnas.192442499>.
- MCMAHON, Mary et al. (sept. 2019). « A single H/ACA small nucleolar RNA mediates tumor suppression downstream of oncogenic RAS ». In : *eLife* 8. DOI : [10.7554/elife.48847](https://doi.org/10.7554/elife.48847). URL : <https://doi.org/10.7554/elife.48847>.
- MEYUHAS, Oded (2015). « Ribosomal Protein S6 Phosphorylation ». In : *International Review of Cell and Molecular Biology*. Elsevier, p. 41-73. DOI : [10.1016/bs.ircmb.2015.07.006](https://doi.org/10.1016/bs.ircmb.2015.07.006). URL : <https://doi.org/10.1016/bs.ircmb.2015.07.006>.
- MICHAEL LOVE, Simon Anders (2017). *DESeq2*. DOI : [10.18129/B9.BIOC.DESEQ2](https://doi.org/10.18129/B9.BIOC.DESEQ2). URL : <https://bioconductor.org/packages/DESeq2>.
- MONTANARO, L et al. (2006). « Dyskerin expression influences the level of ribosomal RNA pseudo-uridylation and telomerase RNA component in

- human breast cancer ». In : *The Journal of Pathology* 210.1, p. 10-18. DOI : [10.1002/path.2023](https://doi.org/10.1002/path.2023). URL : <https://doi.org/10.1002/path.2023>.
- MONTANARO, Lorenzo, Maria CALIENNI et al. (juin 2010). « Novel Dyskerin-Mediated Mechanism of p53 Inactivation through Defective mRNA Translation ». In : *Cancer Research* 70.11, p. 4767-4777. DOI : [10.1158/0008-5472.can-09-4024](https://doi.org/10.1158/0008-5472.can-09-4024). URL : <https://doi.org/10.1158/0008-5472.can-09-4024>.
- MONTANARO, Lorenzo, Davide TRERÉ et Massimo DERENZINI (août 2008). « Nucleolus, Ribosomes, and Cancer ». In : *The American Journal of Pathology* 173.2, p. 301-310. DOI : [10.2353/ajpath.2008.070752](https://doi.org/10.2353/ajpath.2008.070752). URL : <https://doi.org/10.2353/ajpath.2008.070752>.
- MORCELLE, Carmen et al. (sept. 2019). « Oncogenic MYC Induces the Impaired Ribosome Biogenesis Checkpoint and Stabilizes p53 Independent of Increased Ribosome Content ». In : *Cancer Research* 79.17, p. 4348-4359. DOI : [10.1158/0008-5472.can-18-2718](https://doi.org/10.1158/0008-5472.can-18-2718). URL : <https://doi.org/10.1158/0008-5472.can-18-2718>.
- MORGADO-PALACIN, Lucia et al. (oct. 2015). « Partial Loss of Rpl11 in Adult Mice Recapitulates Diamond-Blackfan Anemia and Promotes Lymphomagenesis ». In : *Cell Reports* 13.4, p. 712-722. DOI : [10.1016/j.celrep.2015.09.038](https://doi.org/10.1016/j.celrep.2015.09.038). URL : <https://doi.org/10.1016/j.celrep.2015.09.038>.
- MORIN, Chloé et al. (mai 2022). « Heterogeneity and dynamic of EMT through the plasticity of ribosome and mRNA translation ». In : *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1877.3, p. 188718. DOI : [10.1016/j.bbcan.2022.188718](https://doi.org/10.1016/j.bbcan.2022.188718). URL : <https://doi.org/10.1016/j.bbcan.2022.188718>.
- MOSS, T. et al. (déc. 2006). « A housekeeper with power of attorney : the rRNA genes in ribosome biogenesis ». In : *Cellular and Molecular Life Sciences* 64.1, p. 29-49. DOI : [10.1007/s00018-006-6278-1](https://doi.org/10.1007/s00018-006-6278-1). URL : <https://doi.org/10.1007/s00018-006-6278-1>.
- MOTORIN, Yuri et Virginie MARCHAND (déc. 2018). « Detection and Analysis of RNA Ribose 2'-O-Methylations : Challenges and Solutions ». In : *Genes* 9.12, p. 642. DOI : [10.3390/genes9120642](https://doi.org/10.3390/genes9120642). URL : <https://doi.org/10.3390/genes9120642>.
- NORRIS, Karl, Tayah HOPES et Julie Louise ASPDEN (2021). « Ribosome heterogeneity and specialization in development ». In : *WIREs RNA* 12.4, e1644. DOI : <https://doi.org/10.1002/wrna.1644>. eprint : <https://wires>.

- onlinelibrary.wiley.com/doi/pdf/10.1002/wrna.1644. URL : <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wrna.1644>.
- NOTTINGHAM, Ryan M. et al. (jan. 2016). « RNA-seq of human reference RNA samples using a thermostable group II intron reverse transcriptase ». In : *RNA* 22.4, p. 597-613. DOI : [10.1261/rna.055558.115](https://doi.org/10.1261/rna.055558.115). URL : <https://doi.org/10.1261/rna.055558.115>.
- ODINTSOVA, Tatyana I. et al. (2003). In : *Journal of Protein Chemistry* 22.3, p. 249-258. DOI : [10.1023/a:1025068419698](https://doi.org/10.1023/a:1025068419698). URL : <https://doi.org/10.1023/a:1025068419698>.
- PALADE, George E (jan. 1955). « A small particulate component of the cytoplasm ». en. In : *J. Cell Biol.* 1.1, p. 59-68.
- PARKER, Michael S. et al. (mars 2018). « The Expansion Segments of 28S Ribosomal RNA Extensively Match Human Messenger RNAs ». In : *Frontiers in Genetics* 9. DOI : [10.3389/fgene.2018.00066](https://doi.org/10.3389/fgene.2018.00066). URL : <https://doi.org/10.3389/fgene.2018.00066>.
- PARKS, Matthew M. et al. (fév. 2018). « Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression ». In : *Science Advances* 4.2. DOI : [10.1126/sciadv.aao0665](https://doi.org/10.1126/sciadv.aao0665). URL : <https://doi.org/10.1126/sciadv.aao0665>.
- PECORARO, Annalisa et al. (mai 2021). « Ribosome Biogenesis and Cancer : Overview on Ribosomal Proteins ». In : *International Journal of Molecular Sciences* 22.11, p. 5496. DOI : [10.3390/ijms22115496](https://doi.org/10.3390/ijms22115496). URL : <https://doi.org/10.3390/ijms22115496>.
- PELLETIER, Jerry et Nahum SONENBERG (juill. 1988). « Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA ». In : *Nature* 334.6180, p. 320-325. DOI : [10.1038/334320a0](https://doi.org/10.1038/334320a0). URL : <https://doi.org/10.1038/334320a0>.
- PELTONEN, Karita et al. (jan. 2014). « A Targeting Modality for Destruction of RNA Polymerase I that Possesses Anticancer Activity ». In : *Cancer Cell* 25.1, p. 77-90. DOI : [10.1016/j.ccr.2013.12.009](https://doi.org/10.1016/j.ccr.2013.12.009). URL : <https://doi.org/10.1016/j.ccr.2013.12.009>.
- PENZO, Marianna, Vienna LUDOVINI et al. (juill. 2015). « Dyskerin and TERC expression may condition survival in lung cancer patients ». In : *Oncotarget* 6.25, p. 21755-21760. DOI : [10.18632/oncotarget.4580](https://doi.org/10.18632/oncotarget.4580). URL : <https://doi.org/10.18632/oncotarget.4580>.

- PENZO, Marianna et Lorenzo MONTANARO (juin 2018). « Turning Uridines around : Role of rRNA Pseudouridylation in Ribosome Biogenesis and Ribosomal Function ». In : *Biomolecules* 8.2, p. 38. DOI : [10.3390/biom8020038](https://doi.org/10.3390/biom8020038). URL : <https://doi.org/10.3390/biom8020038>.
- PENZO, Marianna, Lorenzo MONTANARO et al. (jan. 2019). « The Ribosome Biogenesis—Cancer Connection ». In : *Cells* 8.1, p. 55. DOI : [10.3390/cells8010055](https://doi.org/10.3390/cells8010055). URL : <https://doi.org/10.3390/cells8010055>.
- PEREIRA, Michele Araújo, Eddie Luidy IMADA et Rafael Lucas Muniz GUEDES (sept. 2017). « RNA-seq : Applications and Best Practices ». In : *Applications of RNA-Seq and Omics Strategies - From Microorganisms to Human Health*. In-Tech. DOI : [10.5772/intechopen.69250](https://doi.org/10.5772/intechopen.69250). URL : <https://doi.org/10.5772/intechopen.69250>.
- PRUSINER, P., N. YATHINDRA et M. SUNDARALINGAM (oct. 1974). « Effect of ribose O(2')-methylation on the conformation of nucleosides and nucleotides ». In : *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* 366.2, p. 115-123. DOI : [10.1016/0005-2787\(74\)90325-6](https://doi.org/10.1016/0005-2787(74)90325-6). URL : [https://doi.org/10.1016/0005-2787\(74\)90325-6](https://doi.org/10.1016/0005-2787(74)90325-6).
- QIN, Xiaoyu, Bin JIANG et Yanjie ZHANG (fév. 2016). « 4E-BP1, a multifactor regulated multifunctional protein ». In : *Cell Cycle* 15.6, p. 781-786. DOI : [10.1080/15384101.2016.1151581](https://doi.org/10.1080/15384101.2016.1151581). URL : <https://doi.org/10.1080/15384101.2016.1151581>.
- QUARELLO, P. et al. (sept. 2009). « Diamond-Blackfan anemia : genotype-phenotype correlations in Italian patients with RPL5 and RPL11 mutations ». In : *Haematologica* 95.2, p. 206-213. DOI : [10.3324/haematol.2009.011783](https://doi.org/10.3324/haematol.2009.011783). URL : <https://doi.org/10.3324/haematol.2009.011783>.
- RAGHAVAN, A. (déc. 2002). « Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes ». In : *Nucleic Acids Research* 30.24, p. 5529-5538. DOI : [10.1093/nar/gkf682](https://doi.org/10.1093/nar/gkf682). URL : <https://doi.org/10.1093/nar/gkf682>.
- RAJAN, K. Shanmugha et al. (avr. 2020). « The large repertoire of 2'-O-methylation guided by C/D snoRNAs on Trypanosoma brucei rRNA ». In : *RNA Biology* 17.7, p. 1018-1039. DOI : [10.1080/15476286.2020.1750842](https://doi.org/10.1080/15476286.2020.1750842). URL : <https://doi.org/10.1080/15476286.2020.1750842>.
- RAMAGOPAL, Subbanaidu et Herbert L. ENNIS (mai 1981). « Regulation of synthesis of cell-specific ribosomal proteins during differentiation of iDictyostelium discoideum/i ». In : *Proceedings of the National Academy of Sciences* 78.5,

- p. 3083-3087. DOI : [10.1073/pnas.78.5.3083](https://doi.org/10.1073/pnas.78.5.3083). URL : <https://doi.org/10.1073/pnas.78.5.3083>.
- RAO, Shuyun et al. (nov. 2012). « Inactivation of ribosomal protein L22 promotes transformation by induction of the stemness factor, Lin28B ». In : *Blood* 120.18, p. 3764-3773. DOI : [10.1182/blood-2012-03-415349](https://doi.org/10.1182/blood-2012-03-415349). URL : <https://doi.org/10.1182/blood-2012-03-415349>.
- RAY, Partho Sarothi, Richa GROVER et Saumitra DAS (jan. 2006). « Two internal ribosome entry sites mediate the translation of p53 isoforms ». In : *EMBO reports* 7.4, p. 404-410. DOI : [10.1038/sj.embor.7400623](https://doi.org/10.1038/sj.embor.7400623). URL : <https://doi.org/10.1038/sj.embor.7400623>.
- RISSEO, Davide et al. (août 2014). « Normalization of RNA-seq data using factor analysis of control genes or samples ». In : *Nature Biotechnology* 32.9, p. 896-902. DOI : [10.1038/nbt.2931](https://doi.org/10.1038/nbt.2931). URL : <https://doi.org/10.1038/nbt.2931>.
- RIVAS, Mario et George E. FOX (sept. 2020). « Further Characterization of the Pseudo-Symmetrical Ribosomal Region ». In : *Life* 10.9, p. 201. DOI : [10.3390/life10090201](https://doi.org/10.3390/life10090201). URL : <https://doi.org/10.3390/life10090201>.
- ROBINSON, M. D., D. J. MCCARTHY et G. K. SMYTH (nov. 2009). « edgeR : a Bioconductor package for differential expression analysis of digital gene expression data ». In : *Bioinformatics* 26.1, p. 139-140. DOI : [10.1093/bioinformatics/btp616](https://doi.org/10.1093/bioinformatics/btp616). URL : <https://doi.org/10.1093/bioinformatics/btp616>.
- ROCCHI, Laura et al. (juill. 2013). « Dyskerin depletion increases VEGF mRNA internal ribosome entry site-mediated translation ». In : *Nucleic Acids Research* 41.17, p. 8308-8318. DOI : [10.1093/nar/gkt587](https://doi.org/10.1093/nar/gkt587). URL : <https://doi.org/10.1093/nar/gkt587>.
- RONCHETTI, Domenica et al. (sept. 2013). « Small nucleolar RNAs as new biomarkers in chronic lymphocytic leukemia ». In : *BMC Medical Genomics* 6.1. DOI : [10.1186/1755-8794-6-27](https://doi.org/10.1186/1755-8794-6-27). URL : <https://doi.org/10.1186/1755-8794-6-27>.
- RONG, Bowen et al. (déc. 2020). « Ribosome 18S m6A Methyltransferase METTL5 Promotes Translation Initiation and Breast Cancer Cell Growth ». In : *Cell Reports* 33.12, p. 108544. DOI : [10.1016/j.celrep.2020.108544](https://doi.org/10.1016/j.celrep.2020.108544). URL : <https://doi.org/10.1016/j.celrep.2020.108544>.
- RUGGERO, Davide et al. (jan. 2003). « Dyskeratosis Congenita and Cancer in Mice Deficient in Ribosomal RNA Modification ». In : *Science* 299.5604,

- p. 259-262. DOI : [10.1126/science.1079447](https://doi.org/10.1126/science.1079447). URL : <https://doi.org/10.1126/science.1079447>.
- SANCHEZ, M. Isabel G. Lopez et al. (juill. 2020). « Methylation of Ribosomal RNA : A Mitochondrial Perspective ». In : *Frontiers in Genetics* 11. DOI : [10.3389/fgene.2020.00761](https://doi.org/10.3389/fgene.2020.00761). URL : <https://doi.org/10.3389/fgene.2020.00761>.
- SAS-CHEN, Aldema et al. (juin 2020). « Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping ». In : *Nature* 583.7817, p. 638-643. DOI : [10.1038/s41586-020-2418-2](https://doi.org/10.1038/s41586-020-2418-2). URL : <https://doi.org/10.1038/s41586-020-2418-2>.
- SCHACHMAN, Howard K, Arthur B PARDEE et Roger Y STANIER (juill. 1952). « Studies on the macromolecular organization of microbial cells ». In : *Archives of Biochemistry and Biophysics* 38.1, p. 245-260.
- SCHLOSSER, I. (nov. 2003). « A role for c-Myc in the regulation of ribosomal RNA processing ». In : *Nucleic Acids Research* 31.21, p. 6148-6156. DOI : [10.1093/nar/gkg794](https://doi.org/10.1093/nar/gkg794). URL : <https://doi.org/10.1093/nar/gkg794>.
- SCHWANHÄUSSER, Björn et al. (mai 2011). « Global quantification of mammalian gene expression control ». In : *Nature* 473.7347, p. 337-342. DOI : [10.1038/nature10098](https://doi.org/10.1038/nature10098). URL : <https://doi.org/10.1038/nature10098>.
- SHARMA, Sunny et al. (sept. 2017). « Identification of sites of 2'-O-methylation vulnerability in human ribosomal RNAs by systematic mapping ». In : *Scientific Reports* 7.1. DOI : [10.1038/s41598-017-09734-9](https://doi.org/10.1038/s41598-017-09734-9). URL : <https://doi.org/10.1038/s41598-017-09734-9>.
- SHI, Yongquan et al. (juin 2004). « Ribosomal proteins S13 and L23 promote multidrug resistance in gastric cancer cells by suppressing drug-induced apoptosis ». In : *Experimental Cell Research* 296.2, p. 337-346. DOI : [10.1016/j.yexcr.2004.02.009](https://doi.org/10.1016/j.yexcr.2004.02.009). URL : <https://doi.org/10.1016/j.yexcr.2004.02.009>.
- SHI, Zhen et Maria BARNA (nov. 2015). « Translating the Genome in Time and Space : Specialized Ribosomes, RNA Regulons, and RNA-Binding Proteins ». In : *Annual Review of Cell and Developmental Biology* 31.1, p. 31-54. DOI : [10.1146/annurev-cellbio-100814-125346](https://doi.org/10.1146/annurev-cellbio-100814-125346). URL : <https://doi.org/10.1146/annurev-cellbio-100814-125346>.

- SHI, Zhen, Kotaro FUJII et al. (juill. 2017). « Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide ». In : *Molecular Cell* 67.1, 71-83.e7. DOI : [10.1016/j.molcel.2017.05.021](https://doi.org/10.1016/j.molcel.2017.05.021). URL : <https://doi.org/10.1016/j.molcel.2017.05.021>.
- SHILIN ZHAO AND CHUNG-I LI AND YAN GUO AND QUANHU SHENG AND YU SHYR (2020). *RnaSeqSampleSize*. DOI : [10.18129/B9.BIOC.RNASEQSAMPLESIZE](https://doi.org/10.18129/B9.BIOC.RNASEQSAMPLESIZE). URL : <https://bioconductor.org/packages/RnaSeqSampleSize>.
- SLAVOV, Nikolai et al. (nov. 2015). « Differential Stoichiometry among Core Ribosomal Proteins ». In : *Cell Reports* 13.5, p. 865-873. DOI : [10.1016/j.celrep.2015.09.056](https://doi.org/10.1016/j.celrep.2015.09.056). URL : <https://doi.org/10.1016/j.celrep.2015.09.056>.
- SLIMANE, Sophie Nait et al. (oct. 2020). « Ribosome Biogenesis Alterations in Colorectal Cancer ». In : *Cells* 9.11, p. 2361. DOI : [10.3390/cells9112361](https://doi.org/10.3390/cells9112361). URL : <https://doi.org/10.3390/cells9112361>.
- SMITH, Andrew M. et al. (mai 2019). « Reading canonical and modified nucleobases in 16S ribosomal RNA using nanopore native RNA sequencing ». In : *PLOS ONE* 14.5. Sous la dir. d'Hans-Joachim WIEDEN, e0216709. DOI : [10.1371/journal.pone.0216709](https://doi.org/10.1371/journal.pone.0216709). URL : <https://doi.org/10.1371/journal.pone.0216709>.
- SOCIETY. , BIOPHYSICAL et Richard B ROBERTS (nov. 1958). *Microsomal particles and protein synthesis; papers presented at the First Symposium of the Biophysical Society, at the Massachusetts Institute of Technology, Cambridge, February 5, 6, and 8, 1958*. Microsomal particles and protein synthesis; papers presented at the First Symposium of the Biophysical Society, at the Massachusetts Institute of Technology, Cambridge, February 5, 6, and 8, 1958. New York : Published on behalf of the Washington Academy of Sciences, Washington, D.C., by Pergamon Press.
- SONG, Min-Jung et al. (déc. 2009). « A novel initiation codon mutation in the ribosomal protein S17 gene (iRPS17/i) in a patient with Diamond-Blackfan anemia ». In : *Pediatric Blood & Cancer* 54.4, p. 629-631. DOI : [10.1002/pbc.22316](https://doi.org/10.1002/pbc.22316). URL : <https://doi.org/10.1002/pbc.22316>.
- SONG, Myeong Jun et al. (sept. 2011). « RPL36 as a prognostic marker in hepatocellular carcinoma ». In : *Pathology International* 61.11, p. 638-644. DOI : [10.1111/j.1440-1827.2011.02716.x](https://doi.org/10.1111/j.1440-1827.2011.02716.x). URL : <https://doi.org/10.1111/j.1440-1827.2011.02716.x>.

- SPAENDONK, Rosalina M.L. van et al. (juin 2001). « Functional Equivalence of Structurally Distinct Ribosomes in the Malaria Parasite, *Plasmodium berghei* ». In : *Journal of Biological Chemistry* 276.25, p. 22638-22647. DOI : [10.1074/jbc.m101234200](https://doi.org/10.1074/jbc.m101234200). URL : <https://doi.org/10.1074/jbc.m101234200>.
- SPAHN, C M et al. (nov. 2001). « Structure of the 80S ribosome from *Saccharomyces cerevisiae*—tRNA-ribosome and subunit-subunit interactions ». en. In : *Cell* 107.3, p. 373-386.
- STATELLO, Luisa et al. (déc. 2020). « Gene regulation by long non-coding RNAs and its biological functions ». In : *Nature Reviews Molecular Cell Biology* 22.2, p. 96-118. DOI : [10.1038/s41580-020-00315-9](https://doi.org/10.1038/s41580-020-00315-9). URL : <https://doi.org/10.1038/s41580-020-00315-9>.
- STEPHENSON, William et al. (fév. 2022). « Direct detection of RNA modifications and structure using single-molecule nanopore sequencing ». In : *Cell Genomics* 2.2, p. 100097. DOI : [10.1016/j.xgen.2022.100097](https://doi.org/10.1016/j.xgen.2022.100097). URL : <https://doi.org/10.1016/j.xgen.2022.100097>.
- STONELEY, Mark et al. (jan. 1998). « C-Myc 5' untranslated region contains an internal ribosome entry segment ». In : *Oncogene* 16.3, p. 423-428. DOI : [10.1038/sj.onc.1201763](https://doi.org/10.1038/sj.onc.1201763). URL : <https://doi.org/10.1038/sj.onc.1201763>.
- STULTS, Dawn M. et al. (nov. 2007). « Genomic architecture and inheritance of human ribosomal RNA gene clusters ». In : *Genome Research* 18.1, p. 13-18. DOI : [10.1101/gr.6858507](https://doi.org/10.1101/gr.6858507). URL : <https://doi.org/10.1101/gr.6858507>.
- SU, H et al. (avr. 2013). « Elevated snoRNA biogenesis is essential in breast cancer ». In : *Oncogene* 33.11, p. 1348-1358. DOI : [10.1038/onc.2013.89](https://doi.org/10.1038/onc.2013.89). URL : <https://doi.org/10.1038/onc.2013.89>.
- SUGIHARA, Yoshihiko et al. (fév. 2010). « Proteomic Analysis of Rodent Ribosomes Revealed Heterogeneity Including Ribosomal Proteins L10-like, L22-like 1, and L39-like ». In : *Journal of Proteome Research* 9.3, p. 1351-1366. DOI : [10.1021/pr9008964](https://doi.org/10.1021/pr9008964). URL : <https://doi.org/10.1021/pr9008964>.
- SULIMA, Sergey, Kim KAMPEN et Kim De KEERSMAECKER (mars 2019). « Cancer Biogenesis in Ribosomopathies ». In : *Cells* 8.3, p. 229. DOI : [10.3390/cells8030229](https://doi.org/10.3390/cells8030229). URL : <https://doi.org/10.3390/cells8030229>.
- SULIMA, Sergey O. et al. (oct. 2017). « How Ribosomes Translate Cancer ». In : *Cancer Discovery* 7.10, p. 1069-1087. DOI : [10.1158/2159-8290.cd-17-0550](https://doi.org/10.1158/2159-8290.cd-17-0550). URL : <https://doi.org/10.1158/2159-8290.cd-17-0550>.

- SUNG, Hyuna et al. (fév. 2021). « Global Cancer Statistics 2020 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries ». In : *CA : A Cancer Journal for Clinicians* 71.3, p. 209-249. DOI : [10.3322/caac.21660](https://doi.org/10.3322/caac.21660). URL : <https://doi.org/10.3322/caac.21660>.
- TAOKA, Masato, Yuko NOBE, Yuka YAMAKI, Ko SATO et al. (sept. 2018). « Landscape of the complete RNA chemical modifications in the human 80S ribosome ». In : *Nucleic Acids Research* 46.18, p. 9289-9298. DOI : [10.1093/nar/gky811](https://doi.org/10.1093/nar/gky811). URL : <https://doi.org/10.1093/nar/gky811>.
- TAOKA, Masato, Yuko NOBE, Yuka YAMAKI, Yoshio YAMAUCHI et al. (juin 2016). « The complete chemical structure of *Saccharomyces cerevisiae* rRNA : partial pseudouridylation of U2345 in 25S rRNA by snoRNA snR9 ». In : *Nucleic Acids Research* 44.18, p. 8951-8961. DOI : [10.1093/nar/gkw564](https://doi.org/10.1093/nar/gkw564). URL : <https://doi.org/10.1093/nar/gkw564>.
- THIND, Amarinder Singh et al. (juill. 2021). « Demystifying emerging bulk RNA-Seq applications : the application and utility of bioinformatic methodology ». In : *Briefings in Bioinformatics* 22.6. DOI : [10.1093/bib/bbab259](https://doi.org/10.1093/bib/bbab259). URL : <https://doi.org/10.1093/bib/bbab259>.
- THOMSON, Emma, Sébastien FERREIRA-CERCA et Ed HURT (nov. 2013). « Eukaryotic ribosome biogenesis at a glance ». In : *Journal of Cell Science* 126.21, p. 4815-4821. DOI : [10.1242/jcs.111948](https://doi.org/10.1242/jcs.111948). URL : <https://doi.org/10.1242/jcs.111948>.
- TUROWSKI, Tomasz W. et David TOLLERVEY (août 2014). « Cotranscriptional events in eukaryotic ribosome synthesis ». In : *Wiley Interdisciplinary Reviews : RNA* 6.1, p. 129-139. DOI : [10.1002/wrna.1263](https://doi.org/10.1002/wrna.1263). URL : <https://doi.org/10.1002/wrna.1263>.
- VELÁZQUEZ-FLORES, M. Á. et al. (juill. 2020). « Noncoding RNAs as potential biomarkers for DIPG diagnosis and prognosis : XIST and XIST-210 involvement ». In : *Clinical and Translational Oncology* 23.3, p. 501-513. DOI : [10.1007/s12094-020-02443-2](https://doi.org/10.1007/s12094-020-02443-2). URL : <https://doi.org/10.1007/s12094-020-02443-2>.
- VENEZIA, Nicole Dalla et al. (mars 2019). « Emerging Role of Eukaryote Ribosomes in Translational Control ». In : *International Journal of Molecular Sciences* 20.5, p. 1226. DOI : [10.3390/ijms20051226](https://doi.org/10.3390/ijms20051226). URL : <https://doi.org/10.3390/ijms20051226>.

- WAGNER, Günter P., Koryu KIN et Vincent J. LYNCH (août 2012). « Measurement of mRNA abundance using RNA-seq data : RPKM measure is inconsistent among samples ». In : *Theory in Biosciences* 131.4, p. 281-285. DOI : 10.1007/s12064-012-0162-3. URL : <https://doi.org/10.1007/s12064-012-0162-3>.
- WALSH, D., M. B. MATHEWS et I. MOHR (déc. 2012). « Tinkering with Translation : Protein Synthesis in Virus-Infected Cells ». In : *Cold Spring Harbor Perspectives in Biology* 5.1, a012351-a012351. DOI : 10.1101/cshperspect.a012351. URL : <https://doi.org/10.1101/cshperspect.a012351>.
- WANG, Hui et al. (oct. 2022). « SARS-CoV-2 N protein potentiates host NPM1-snoRNA translation machinery to enhance viral replication ». In : *Signal Transduction and Targeted Therapy* 7.1. DOI : 10.1038/s41392-022-01210-9. URL : <https://doi.org/10.1038/s41392-022-01210-9>.
- WARNER, Jonathan R (nov. 1999). « The economics of ribosome biosynthesis in yeast ». In : *Trends in Biochemical Sciences* 24.11, p. 437-440. DOI : 10.1016/s0968-0004(99)01460-7. URL : [https://doi.org/10.1016/s0968-0004\(99\)01460-7](https://doi.org/10.1016/s0968-0004(99)01460-7).
- WARNER, Jonathan R. et Kerri B. MCINTOSH (avr. 2009). « How Common Are Extraribosomal Functions of Ribosomal Proteins ? » In : *Molecular Cell* 34.1, p. 3-11. DOI : 10.1016/j.molcel.2009.03.006. URL : <https://doi.org/10.1016/j.molcel.2009.03.006>.
- WEBER, H. J. (sept. 1972). « Stoichiometric measurements of 30S and 50S ribosomal proteins from Escherichia coli ». In : *Molecular and General Genetics* MGG 119.3, p. 233-248. DOI : 10.1007/bf00333861. URL : <https://doi.org/10.1007/bf00333861>.
- WEIJERS, Dolf et al. (nov. 2001). « An iArabidopsis/i Minute-like phenotype caused by a semi-dominant mutation in a iRIBOSOMAL PROTEIN S5/i gene ». In : *Development* 128.21, p. 4289-4299. DOI : 10.1242/dev.128.21.4289. URL : <https://doi.org/10.1242/dev.128.21.4289>.
- WU, Qiong et al. (oct. 2011). « Downregulation of RPL6 by siRNA Inhibits Proliferation and Cell Cycle Progression of Human Gastric Cancer Cell Lines ». In : *PLoS ONE* 6.10. Sous la dir. de Xin-yuan GUAN, e26401. DOI : 10.1371/journal.pone.0026401. URL : <https://doi.org/10.1371/journal.pone.0026401>.

- XIONG, Xiufang et al. (nov. 2018). « Ribosomal protein S27-like regulates autophagy via the -TrCP-DEPTOR-mTORC1 axis ». In : *Cell Death & Disease* 9.11. DOI : [10.1038/s41419-018-1168-7](https://doi.org/10.1038/s41419-018-1168-7). URL : <https://doi.org/10.1038/s41419-018-1168-7>.
- XUE, Shifeng et Maria BARNA (mai 2012). « Specialized ribosomes : a new frontier in gene regulation and organismal biology ». In : *Nature Reviews Molecular Cell Biology* 13.6, p. 355-369. DOI : [10.1038/nrm3359](https://doi.org/10.1038/nrm3359). URL : <https://doi.org/10.1038/nrm3359>.
- XUE, Shifeng, Siqi TIAN et al. (nov. 2014). « RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation ». In : *Nature* 517.7532, p. 33-38. DOI : [10.1038/nature14010](https://doi.org/10.1038/nature14010). URL : <https://doi.org/10.1038/nature14010>.
- YAMAUCHI, Yoshio et al. (déc. 2015). « A mass spectrometry-based method for direct determination of pseudouridine in RNA ». In : *Nucleic Acids Research* 44.6, e59-e59. DOI : [10.1093/nar/gkv1462](https://doi.org/10.1093/nar/gkv1462). URL : <https://doi.org/10.1093/nar/gkv1462>.
- YANG, Jun et al. (mars 2017). « Correction : Mapping of Complete Set of Ribose and Base Modifications of Yeast rRNA by RP-HPLC and Mung Bean Nuclease Assay ». In : *PLOS ONE* 12.3, e0173940. DOI : [10.1371/journal.pone.0173940](https://doi.org/10.1371/journal.pone.0173940). URL : <https://doi.org/10.1371/journal.pone.0173940>.
- YI, Yang et al. (avr. 2021). « A PRC2-independent function for EZH2 in regulating rRNA 2'-O methylation and IRES-dependent translation ». In : *Nature Cell Biology* 23.4, p. 341-354. DOI : [10.1038/s41556-021-00653-6](https://doi.org/10.1038/s41556-021-00653-6). URL : <https://doi.org/10.1038/s41556-021-00653-6>.
- YOON, Andrew et al. (mai 2006). « Impaired Control of IRES-Mediated Translation in X-Linked Dyskeratosis Congenita ». In : *Science* 312.5775, p. 902-906. DOI : [10.1126/science.1123835](https://doi.org/10.1126/science.1123835). URL : <https://doi.org/10.1126/science.1123835>.
- YOSHIHAMA, Maki, Akihiro NAKAO et Naoya KENMOCHI (oct. 2013). « snOPY : a small nucleolar RNA orthological gene database ». In : *BMC Research Notes* 6.1. DOI : [10.1186/1756-0500-6-426](https://doi.org/10.1186/1756-0500-6-426). URL : <https://doi.org/10.1186/1756-0500-6-426>.
- YU, Yonghao et al. (jan. 2009). « Mass spectrometric analysis of the human 40S ribosomal subunit : Native and HCV IRES-bound complexes ». In : *Protein Science* 14.6, p. 1438-1446. DOI : [10.1110/ps.041293005](https://doi.org/10.1110/ps.041293005). URL : <https://doi.org/10.1110/ps.041293005>.

- YUSUPOVA, Gulnara Zh. et al. (juill. 2001). « The Path of Messenger RNA through the Ribosome ». In : *Cell* 106.2, p. 233-241. DOI : [10.1016/s0092-8674\(01\)00435-4](https://doi.org/10.1016/s0092-8674(01)00435-4). URL : [https://doi.org/10.1016/s0092-8674\(01\)00435-4](https://doi.org/10.1016/s0092-8674(01)00435-4).
- ZHAI, Weiguo et Lucio COMAI (août 2000). « Repression of RNA Polymerase I Transcription by the Tumor Suppressor p53 ». In : *Molecular and Cellular Biology* 20.16, p. 5930-5938. DOI : [10.1128/mcb.20.16.5930-5938.2000](https://doi.org/10.1128/mcb.20.16.5930-5938.2000). URL : <https://doi.org/10.1128/mcb.20.16.5930-5938.2000>.
- ZHANG, Jing et al. (déc. 2020). « Increased fibrillarin expression is associated with tumor progression and an unfavorable prognosis in hepatocellular carcinoma ». In : *Oncology Letters* 21.2. DOI : [10.3892/ol.2020.12353](https://doi.org/10.3892/ol.2020.12353). URL : <https://doi.org/10.3892/ol.2020.12353>.
- ZHAO, Jian et al. (avr. 2020). « IRESbase : A Comprehensive Database of Experimentally Validated Internal Ribosome Entry Sites ». In : *Genomics, Proteomics & Bioinformatics* 18.2, p. 129-139. DOI : [10.1016/j.gpb.2020.03.001](https://doi.org/10.1016/j.gpb.2020.03.001). URL : <https://doi.org/10.1016/j.gpb.2020.03.001>.
- ZHOU, Fengbiao et al. (oct. 2022). « A dynamic rRNA ribomethylome drives stemness in acute myeloid leukemia ». In : *Cancer Discovery*. CD-22-0210. ISSN : 2159-8274. DOI : [10.1158/2159-8290.CD-22-0210](https://doi.org/10.1158/2159-8290.CD-22-0210). eprint : <https://aacrjournals.org/cancerdiscovery/article-pdf/doi/10.1158/2159-8290.CD-22-0210/3215082/cd-22-0210.pdf>. URL : <https://doi.org/10.1158/2159-8290.CD-22-0210>.
- ZHU, Yinzhou, Stephan P. PIRNIE et Gordon G. CARMICHAEL (mai 2017). « High-throughput and site-specific identification of 2'-iO/i-methylation sites using ribose oxidation sequencing (RibOxi-seq) ». In : *RNA* 23.8, p. 1303-1314. DOI : [10.1261/rna.061549.117](https://doi.org/10.1261/rna.061549.117). URL : <https://doi.org/10.1261/rna.061549.117>.