Université de Montréal

Approches de fractionnement biochimique couplé à la transcriptomique dans l'étude systématique de la localisation subcellulaire et extracellulaire des ARNs

par
Fabio Alexis Lefebvre

Département de Biochimie
Faculté de médecine

Thèse présentée à la Faculté de médecine en vue de
l'obtention du grade de doctorat en biochimie

Août 2018

Résumé

Divers transcrits acquièrent une asymétrie spatiale au sein de certaines cellules procaryotes et eucaryotes, phénomène qualifié de « localisation des ARNs ». Chez les types cellulaires dotés d'une polarisation marquée, notamment les neurones ou certains embryons, la localisation des ARNm constitue un mécanisme élégant permettant de restreindre l'expression et l'activité protéique associée à un contexte spatiotemporel précis. Bien que diverses techniques d'imagerie permettent d'apprécier la distribution spatiale des transcrits à haute résolution, elles se prêtent difficilement à l'étude systématique de l'asymétrie spatiale du transcriptome.

Cette thèse retrace d'abord le développement de méthodes biochimiques de fractionnement cellulaire et extracellulaire couplées au séquençage à haut débit des ARNs comme approche systématique dans l'étude de la localisation des ARNs. Ces méthodologies, qualifiées collectivement de CeFra-seq (« Cell Fractionation – RNA-seq »), se veulent complémentaires aux outils d'imagerie. Le chapitre 4 est consacré à une description technique détaillée de l'approche CeFra-seq chez les cellules humaines leucémiques K562, accompagnée d'étapes de validation et de transformation de données. Les chapitres subséquents traitent ensuite de l'application de ces méthodologies pour explorer quatre questions fonctionnelles chez divers systèmes biologiques.

Le chapitre 6 tire profit de l'approche CeFra-seq pour explorer les propriétés subcellulaires des ARNs ciblés aux vésicules extracellulaires (VEs). Les VEs correspondent à un groupe hétérogène de structures nanoscopiques constituées d'une bicouche lipidique qui contiennent un répertoire spécifique d'acides nucléiques et des protéines. Ubiquitaire au sein des liquides biologiques, les VEs ont été associées à la communication intercellulaire dans divers contextes, de la présentation des antigènes à la progression tumorale. Or, les mécanismes qui déterminent la localisation préférentielle de certains ARNs aux VEs demeurent nébuleux. En contrastant de manière systématique les populations d'ARNs contenus dans ces structures aux répertoires subcellulaires obtenus par l'approche de CeFra-seq, mon travail a permis de mettre en évidence certaines propriétés associées au ciblage extracellulaire, notamment l'accessibilité cytosolique, la taille des ARNs ainsi que des éléments de séquence en *cis*.

Le chapitre 7 propose une comparaison extensive des propriétés morphologiques et transcriptomiques des VEs issues d'une série de lignées cellulaires humaines et de lignées embryonnaires de Drosophile. Ce travail révèle que les VEs de Drosophile sont plus petites que

celles des cancers humains et que l'enrichissement d'ARNs courts transcrits par la polymérase III prévaut chez les VEs des deux espèces. Ensemble, ces résultats valident l'hypothèse d'une conservation élevée des phénomènes d'export de l'ARN.

Le chapitre 8 étend CeFra-seq à un nouveau contexte biologique : le développement embryonnaire de la Drosophile. Ici, cette méthodologie conduit à l'établissement de répertoires de transcrits dotés d'une forte asymétrie spatiale et temporelle au cours de l'embryogenèse. L'analyse des mutants SLBP, un facteur de maturation des ARNs d'histone, et de Chk1, régulateur des voies de dommage à l'ADN, montre ensuite que la déplétion de ces protéines compromet sélectivement l'expression des transcrits zygotiques, identifiés grâce à la méthode CeFra-seq.

Le chapitre 9 relate une étude des ARNs antisens issus du locus des histones pendant l'embryogenèse de la Drosophile. L'expression de ces transcrits non-polyadénylés fluctue pendant le développement et dépend de la protéine SLBP. Ici, l'approche CeFra-seq révèle que ces transcrits antisens, strictement zygotiques, co-ségréguent avec leurs ARNm complémentaires, un résultat qui évoque la formation d'ARN double-brin, puis de petits ARN interférents. Pour faire suite à cette hypothèse, j'ai démontré que de petits transcrits issus du locus des histones s'associent au facteur catalytique de la machinerie d'interférence aux ARNs, Argonaute-2. De plus, la déplétion d'Argonaute-2 mène à une dérepression des ARNm d'histones. Ensemble, ces résultats suggèrent un modèle de transcription antisens zygotique précoce menant à la formation de petits ARNs interférents qui contribuent à l'élimination des ARNm d'histones contribués maternellement.

Ainsi, cette thèse est échafaudée sur le développement d'une approche versatile de l'étude systématique de la localisation des ARN, CeFra-seq, décrite dans le chapitre 5. La mise au point de cette approche débouche ensuite sur des études fonctionnelles visant à mieux comprendre les propriétés des ARNs ciblés au VEs (Chapitre 6) et le degré de conservation de ce ciblage (Chapitre 7). La suite de la thèse exploite l'approche CeFra-seq pour explorer le phénotype transcriptomique de la déplétion de SLBP chez l'embryon de Drosophile (Chapitre 8), ainsi que les propriétés et fonctions d'ARN antisens issus du locus des histones chez l'embryon précoce (Chapitre 9).

Mots-clés : Localisation de l'ARN, Régulation post-transcriptionnelle, Biologie des systèmes, Transcriptomique, ARNs non-codants, Vésicules extracellulaires, Exosomes, Développement embryonnaire, Drosophile, Histones

Abstract

Several RNA transcripts acquire spatially-resolved patterns in diverse prokaryotic and eukaryotic cells, a phenomenon termed “RNA localization”. In highly polarized cells, such as neurons or certain embryos, mRNA localization provides an elegant mechanism to restrict protein expression and activity to a narrow spatiotemporal context. Diverse imaging approaches have been developed to study RNA localization, including RNA in situ hybridization and the MS2 system. While these techniques enable the visualization of RNA spatial distributions at a high resolution, they hardly allow for systematic, transcriptome-wide analyses of spatial asymmetry.

The first part of this thesis encompasses the development of biochemical, cell fractionation and extracellular milieu processing methods coupled to deep sequencing as a novel approach to study transcriptome-wide RNA localization. These methods, collectively termed CeFra-seq (“Cell Fractionation – RNA-seq”), are propose as a complementary tool with imaging approaches. Chapter 5 consists of a detailed technical description of the CeFra-seq methodology, along with a validation workflow and a relevant data transformation toolkit. The subsequent chapters discuss applications of these methods to investigate four outstanding questions in different biological systems.

Chapter 6 relies on CeFra-seq to explore the subcellular properties of RNAs targeted to extracellular vesicles (EVs). EVs form a group of heterogeneous nanoscopic structures delimited by a phospholipid bilayer that contain specific repertoires of protein and nucleic acids. Ubiquitous in biological fluids, EVs have been associated to intercellular communication in diverse biological contexts, notably antigen presentation and tumor progression. Yet, the mechanisms that account for the enrichment of specific RNAs in EVs remain unclear. By systematically contrasting RNA populations found in EVs with subcellular distributions, my work has revealed diverse properties linked to EV targeting, including cytosolic accessibility, RNA length and *cis*-acting elements.

Chapter 7 consists of an extensive comparison of the morphological and transcriptomic properties of EVs derived from several human and *Drosophila* cell lines. This work reveals that *Drosophila* EVs are smaller than their human counterparts and that they are both enriched in short, Polymerase III transcripts. Together, these results emphasize the high conservation of RNA export processes.

Chapter 8 extends subcellular fractionation approaches coupled to deep sequencing in an additional biological system: *Drosophila* embryogenesis. Here, the method leads to repertoires of transcripts displaying high spatial and temporal asymmetry during development. The analysis of SLBP mutants, a factor involved in histone mRNA processing, and Chk1 mutnts, a regulator of the DNA damage response, shows that the depletion of these proteins selectively hampers the expression of zygotic transcripts, identified through CeFra-seq.

Chapter 9 recounts a study of antisense transcripts derived from the histone gene locus during Drosophila embryogenesis. The expression of these non-polyadenylated RNAs fluctuates during development and depends on the protein SLBP. Here, CeFra-seq reveals that these antisense RNAs, which are strictly zygotic, co-segregate with their complementary mRNAs, hinting at the formation of double-stranded RNAs, precursors of small interfering RNAs. To follow-up on this hypothesis, I show that small RNAs derived from the histone gene locus bind to the catalytic factor of the RNA-induced silencing complex, Argonaute-2. In addition, depleting Argonaute-2 leads to a derepression of histone mRNAs. Together, these results suggest a model wherein precocious zygotic antisense transcription leads to the formation of small interfering RNAs, which contribute to the clearance of maternally deposited histone mRNAs.

Hence, this thesis reflects the development and application of a versatile approach to study RNA localization, termed CeFra-seq and described in chapter 5. The use of this method leads to functional studies aiming to investigate the properties of EV-targeted RNAs (Chapter 6) and the extent of evolutionary conservation of this targeting process (Chapter 7). The rest of the thesis exploits the CeFra-seq approach to explore the transcriptomic phenotype of SLBP depletion in *Drosophila* embryos (Chapter 8), as well as the properties and functions of antisense RNAs produced by the histone gene locus in early embryos (Chapter 9).

Keywords : RNA localization, Posttranscriptional regulation, Systems biology, Transcriptomics, Non-coding RNAs, Extracellular vesicles, Exosomes, Embryonic development, *Drosophila*, Histones

Table des matières

Liste des tableaux

Chapitre 2

Chapitre 5

Chapitre 6

Chapitre 7

Chapitre 8

Chapitre 9

Liste des figures

Chapitre 1

Chapitre 2

Chapitre 3

Chapitre 4

Chapitre 5

Chapitre 6

Chapitre 7

Chapitre 8

Chapitre 9

Chapitre 10

Liste des abréviations

3'UTR : 3' Untranslated region

Abd-B : Abdominal-B

ADN : acide désoxyribonucléique

ag : attogramme (10^{-18} g)

Ago : Argonaute

Ago1 : Argonaute-1

Ago2 : Argonaute-2

Alix : Apoptosis-linked gene 2-interacting protein X

ANOVA : Analysis of variance

ARN : acide ribonucléique

ARNm : ARN messager

ARNr : ARN ribosomal

ARNt : ARN de transfert

Ash1 : Asymmetric synthesis of HO

ATM : Ataxia telangiectasia mutated

ATR : Ataxia telangiectasia and Rad3-related protein

BAM : Binary alignment map

BDNF : Brain-derived neurotrophic factor

Bp : Base pair

BRAT : Brain tumor

C3PO : Component 3 promoter of RISC

Ca : Calcium

$CaCl_2$: Chlorure de calcium

CCR4-NOT complex : Carbon Catabolite Repressor 4-Negative on TATA

CD43 : Cluster of differentiation 43

Cdk1 : Cyclin-dependent kinase 1

CeFra-seq : Cell fractionation, RNA-sequencing

CeFra-seq : Cell fractionation, RNA-sequencing

ChIP-seq : Chromatin immunoprecipitation, DNA-sequencing

Chk1 : Checkpoint kinase 1

Chk2 : Checkpoint kinase 2

CIHR : Canadian Institute of Health Research

Cos : Cosinus

CPSF73 : Cleavage and polyadenylation specificity factor 73

CRM : cis-regulatory motif

CycB : Cyclin B

CYP51 : Cytochrome P51

CYP6AE14 : Cytochrome P6AE14

DDR : DNA damage response

Df : Deficiency

DGCR8 : DiGeorge syndrome chromosomal [or critical] region 8

diRNA : Double stranded breaks (DSB)-induced RNA

DNAse : Desoxyribonuclease

DRE : DNA replication-related element

DSB : Double stranded breaks

dsRNA : double-stranded RNA

DTT : Dithiothréitol

EDTA : Ethylène diamine tétraacétique

EGTA : Ethylène glycol tétraacétique

En : Engrailed

Endo-siRNA : Endogenous siRNA

ER : Endoplasmic reticulum

ERK : Extracellular signal-regulated kinase

eRNA : enhancer RNA

ESCRT : Endosomal sorting complexes required for transport

ET : Cryo-electron tomography

EV : Extracellular vesicle

exRNA : Extracellular RNA

FISH : Fluorescence in situ hybridization

fm : femtogramme (10^{-15} g)

FPKM : Fragments per kilobase of exon per million fragments mapped

FRQS : Fonds de Recherche Québec Santé

GFP : Green fluorescent protein

GLUE : GRAM-Like Ubiquitin-binding in EAP45

Grp : Grapes

H1 : Histone H1

H2A : Histone H2A

H2B : Histone H2B

H3 : Histone H3

H3K18ac: Mono-acetylation of Lysine 18 of Histone 3

H3K27me3: Tri-methylation of Lysine 27 of Histone 3

H3K36me3: Tri-methylation of Lysine 36 of Histone 3

H3K4me1: Mono-methylation of Lysine 4 of Histone 3

H3K4me3: Tri-methylation of Lysine 4 of Histone 3

H3K9ac: Mono-acetylation of Lysine 9 of Histone 3

H4 : Histone H4

H4K8ac: Mono-acetylation of Lysine 8 of Histone 4

HCl : Acide chlorhydrique

HDE : Histone downstream element

HGS-hpRNAi : Host gene silencing-hairpin RNAi

His1 : Histone H1

His2A : Histone H2A

His2B : Histone H2B

His3 : Histone H3

His 4 : Histone H4

HIV : Human immunodeficiency virus

hnRNP : Heterogeneous ribonucleoprotein particle

hnRNPA2B1 : Heterogeneous ribonucleoprotein particle A2B1

HOTAIR : HOX transcript antisense RNA

IFN-γ : Interferon-gamma

IL-13 : Interleukin-13

IL-4 : Interleukin-4

IL-5 : Interleukin-5

IRCM : Institut de recherches cliniques de Montréal

kb : Kilobase

KCl : Chlorure de potassium

KLF4 : Kruppel-like factor 4

KLF5 : Kruppel-like factor 5

Kr : Krüppel

KRAS : V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog

lncRNA : long non-coding RNA

LT-β : Lymphotoxin beta

LTR : Long terminal repeat

M-MLV RT : Moloney Murine Leukemia Virus Reverse Transcriptase

MBT: Midblastula transition

ME31B: Maternal expression at 31B

Mei-41 : meiotic 41

MEK : MAPK/ERK kinase

mESC : Mouse Embryonic Stem Cell

$MgCl_2$: Chlorure de magnésium

MHC-II : Major histocompatibility complex class II

min AEL : Minutes after egg laying

miRNA : Micro ARN

MPK1 : Mitogen-activated protein kinase 1

MPK2 : Mitogen-activated protein kinase 2

MPKKK4 : Mitogen-activated protein kinase kinase kinase 4

mRNA : Messenger RNA

MSP-70 : Mitochondrial Stress-70 protein precursor

MST2 : Monosaccharide Transporter 2

MVE : Multivesicular endosome

MZT : Mother-to-zygotic transition

N/C : Nucleocytoplasmic ratio

NC : Nuclear cycle

ncRNA : Non-coding RNA

NEAT1 : Nuclear Enriched Abundant Transcript 1

NFR : Nucleosome-free region

NP-40 : Nonyl-phénoxypolyéthoxyléthanol

NPM1 : Nucleophosmin-1

nSMase : Neutral sphingomyelinase

nt : nucléotide

NTA : Nanoparticle tracking analysis

Nup160 : Nucleoporin 160

P : P-value

P-bodies : Processing bodies

PACT : Protein activator of the interferon-induced protein kinase

PAZ : Piwi Argonaute Zwille

pDNA : plasmidic DNA

PGCs: Primordial germ cells

PI(3)K : Phosphoinositide 3-kinase

piRNA : Piwi-interacting RNA

PMSF : Fluorure d'α-toluènesulfonyle

pre-miRNA : precursor miRNA

pri-miRNA : primary miRNA

PRXIIF : Peroxiredoxin

PTEN : Phosphatase and TENsin homolog

PUM : Pumilio

qPCR : Quantitative PCR

RBP : RNA-binding protein

Relm-α : Resistin-like molecule alpha

Relm-β : Resistin-like molecule beta

RELMβ : Resistin-like molecule beta

Rho : Rhomboid

RISC : RNA-induced silencing complex

RNA : Ribonucleic acid

RNAi : RNA interference

RNAse : Ribonuclease

RNAse A : Riobonucléase A

RNAse P : Ribonuclease P

RNP : Ribonucleoprotein

RPA70: Replication protein A 70 kDa

SD_{NTA} : Écart-type des diamètres déterminés par Nanoparticle tracking analysis (NTA)

SDS-PAGE : Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SD_{TEM} : Écart-type des diamètres déterminés par microscopie électronique (TEM)

Ser : Serine

SID-1 : Systemic RNA interference defective protein 1

SID-2 : Systemic RNA interference defective protein 2

Sin : Sinus

siRNA : Small interfering RNA

SL : Stem loop

SLBP : Stem loop binding protein

SMG : Smaug

snoARN : Petit ARN nucléolaire (snoRNA)

snoRNA : Small nucleolar RNA

snARN : Petit ARN nucléaire (snRNAs)

snRNA : Small nuclear RNA

snRNP : small nuclear ribonucleoprotein

Sog : Short gastrulation

sRNA : Small RNA

ssDNA : single-stranded DNA

SUMO : Small ubiquitin-like modifier

TADs : Topologically associated domains

TEM : Transmission electron microscopy

TF : Transcription factor

TGF-β : Trasnforming growth factor beta

Thr : Threonine

Ti plasmid : Tumor-inducing plasmid

TP: Tyrosine phosphatase

TRBP : TAR RNA binding protein

Trl : Trithorax-like

tRNA : Transfer RNA

TSLP : Thymic stromal lymphopoietin

UCSC : University of California, Santa Cruz

VEs : Vésicules extracellulaires (EVs)

VIH : Virus de l'immunodéficience humaine

VLD : Zelda/Vielfatig

VPS36 : Vacuolar protein-sorting-associated protein 36

vtRNA : Vault RNA

WAK : Cell-wall associated kinase

Wnt : Wingless integration

WT : Wild type

XIST : X-inactive specific transcript

ZGA : Zygotic genome activation

ZLD : Zelda/Vielfatig

Zn : Zinc

ZO-1 : Zonula occludens-1

μ_{NTA} : Diamètre moyen déterminé par Nanoparticle tracking analysis (NTA)

μ_{TEM} : Diamètre moyen déterminé par microscopie électronique (TEM)

Je dédie cette thèse à Jacques, mon père, décédé en juillet 2017. Pendant mon enfance, papa a beaucoup contribué à l'émergence chez moi d'une fascination pour le Vivant. La pêche au crapet-soleil, la chasse aux cigales, ce vivarium d'escargots et notre collection de plumes d'oiseaux ont pavé la voie à la recherche en biologie. Merci de m'avoir transmis l'émerveillement que t'inspirait la nature.

Remerciements

Je remercie en premier lieu Dr Éric Lécuyer, mon superviseur. J'ai d'abord approché Éric en 2011; jeune et naïf étudiant de première année au baccalauréat, j'étais fasciné par la beauté des *in situ* réalisées par Éric pendant son stage post-doctoral chez Dr Henry Krause. J'ai découvert un homme d'une générosité exceptionnelle, un père de famille qui connaît la valeur du travail, du consensus et de l'harmonie. Pendant les nombreuses années où j'ai fréquenté son laboratoire, la porte du bureau d'Éric est toujours restée ouverte. Il fait preuve d'une exceptionnelle disponibilité et d'une précieuse attention à l'endroit de ses étudiants, même pendant la saison des *grants*. En progressant dans mes travaux, j'ai aussi appris à percevoir le grand courage d'Éric. Diriger un laboratoire peut, à l'occasion, évoquer la navigation en eaux troubles. Face aux écueils financiers, académiques et interpersonnels, Éric a toujours fait preuve d'une droiture impressionnante, cultivant le goût du travail acharné et l'importance de protéger ses matelots. C'est grâce à tes efforts insatiables que le laboratoire continue de croître. Quand ma crise d'adolescence académique a surgi et que je me suis révélé critique au seuil de l'insolence, Éric a réagi avec un professionnalisme serein et une grande maturité. Merci de ta patience. Merci de ton indulgence aussi, quand mon père est tombé malade et que j'ai choisi de m'absenter longuement du labo. Tu m'as permis d'accompagner son parcours vers la mort et ces moments, toujours vivides dans ma mémoire, m'ont beaucoup fait grandir. Je suis fier d'être ton étudiant. Merci pour ces centaines de kilomètres que nous avons parcourus, de semaine en semaine, lors de nos courses collectives sur le Mont-Royal. Tu as réveillé chez moi un goût certain pour le sport, qui était décidément profondément enfoui. Merci pour ces soirées arrosées au chalet de l'IRCM en compagnie de ta famille et de nos amis. Elles accompagneront longtemps mes moments de solitude.

Je remercie les organismes subventionnaires qui ont contribué à ma rémunération et m'ont ainsi permis de poursuivre mes travaux, recensés dans cette thèse. Merci aux Instituts de Recherche en Santé du Canada (IRSC), au Fonds de Recherche du Québec – Santé (FRQS), au Conseil de

Recherches en Sciences Naturelles et en Génie du Canada (CRSNG) et à la Faculté des Études Suéprieures et Post-Doctorales de l'Université de Montréal (FESP).

Je remercie Dr Janusz Rak, qui nous a guidés dans la découverte de l'univers des vésicules extracellulaires. Scientifique à la fois incisif et rieur, conteur prolixe qui cultive la justesse et la beauté des mots, nos discussions continueront de m'éclairer longtemps.

Je remercie Louis Philip Benoit Bouvrette, mon grand ami et éternel co-auteur. Nous partageons une rigueur certaine et un goût affirmé pour une certaine forme d'humour acerbe. T'avoir comme compagnon au fil de ces années m'a beaucoup aidé à les traverser. Je n'oublierai jamais ces longues soirées passées à discuter d'alignement de séquences en buvant de la bière dans le glorieux taudis que j'habitais à l'époque.

Je remercie Dr Julie Bergalet, notre bouffée d'air frais national. Ton sourire n'a jamais fait défaut et la rigolade était toujours au rendez-vous. J'ai beaucoup appris de toi, autant à la paillasse que dans la vie. Ta légèreté est une grande vertu et un gage de sagesse. Arpenter Tokyo et les plages de la Crête avec toi s'est révélé un grand plaisir ponctué d'éclats de rire tonitruants.

Je remercie Dr Carole Iampietro, maman-mouche, animée d'un rare idéal stoïcien et d'une grande douceur, combinaison surprenante s'il en est une. Tu m'as beaucoup appris, autant sur les chromosomes balanceurs que sur le sens de la vocation de scientifique.

I would like to thank Dr. Neal Cody. You've been a big brother of mine. You work ethic will remain a reference: you taught me the importance of careful experiments, thorough controls and self-criticism. Thanks for introducing me to the UCSC Genome Browser, which quickly became a strong asset after you left the lab.

I would like to thank our lab's enduring pillar, Xiaofeng (Andy) Wang. Your resilience and wisdom have helped me tremendously throughout the years. You've been my personal Mandarin teacher, always patient and resourceful. 感谢你的智慧和你的喜悦.

Je remercie les trois étudiants que j'ai supervisés au fil des ans : Josée-Anne Boucher, Lilyanne Perras et Juan-Carlos Padilla. Votre détermination et votre goût d'apprendre m'ont inspiré et donné envie de continuer. Travailler à vos côtés aura été un honneur. *Thanks!*

Merci à Dr Delphine Garnier de m'avoir initié au monde merveilleux des vésicules. Tu as été un excellent mentor.

Merci à Dany Gauthier, notre dame en rouge! Tu es une femme merveilleuse et j'ai beaucoup apprécié ta compagnie.

私は親愛なる友人マチューと妻ナツコに感謝したいと思います。私の滞在中の友情とおもてなしは、常に私の心を蘇らせます。私はあなたがモントリオールで私を訪れるのを待つことができません。

Merci à Dr Adeline Gadzinsky pour sa finesse, son idéalisme, son humour et son esprit critique acéré. Ta présence à l'IRCM me manque toujours.

Merci à Renée Francoeur. Quand tu y es, les petits voyages quotidiens à la laverie de l'étage prennent un nouveau sens. J'aime ton sens de la répartie.

J'aimerais remercier Félix Legendre. Ton humour, ta légèreté et ta douce excentricité m'ont longtemps manqué. Nous avions une camaraderie authentique. J'ai appris que tu avais entrepris un doctorat à Ottawa et je te souhaite beaucoup de succès.

Je tiens à remercier Dr Nikolaus Heveker pour son hospitalité, son insatiable goût d'apprendre et son penchant marqué pour la pédagogie. Tu m'as beaucoup éclairé.

Vorrei ringraziare Rodrigo, che era un bidello da alcuni anni all'IRCM. Il tuo umorismo e la tua leggerezza mi mancano ancora. La pratica dell'italiano con te mi ha dato il più grande piacere. Spero che festeggi il tuo ritiro nei Caraibi ora, come previsto.

Je remercie les membres fidèles de mes comités de thèse et de pré-doc, Dr Vincent Archambault, Dr Jacques Drouin et Dr Pascal Chartrand. Avec ténacité, vous m'avez amené à me remettre en question et à poursuivre ma croissance. C'était la bonne approche.

I would like to thank Dr. Lawrence (Lawrie) Rajendran. Your love and guidance will stay with me forever. நான் ராட்டர்டாமில் அந்த இரவு மறக்க மாட்டேன். விரைவில் உங்களை சந்திக்க விரும்புகிறேன்.

Merci à Dr Vladimir (Vlad) Reinharz pour son hospitalité et ses précieux conseils. Tu m'as beaucoup éclairé.

Я хотел бы поблагодарить Илю. Спасибо, что познакомил меня с Тель-Авивом. Иногда мне снится, что я просыпаюсь против тебя. Я надеюсь увидеть тебя снова.

Merci à Caroline et Thierry. Vous visiter est toujours un plaisir et j'ai bien hâte que vous passiez me voir à Montréal.

והערכתי מופתית היתה ברחובות שלכם האורחים הכנסת .אוליצקי איגור ר"ולד פילפל יצחק ר"לד להודות רוצה אני היהודית האורחים הכנסת את מאוד.

Je voudrais remercier Dr Jérôme Waldispuhl (McGill) et Dr Bruno Sargueuil (Paris Descartes). Votre générosité et votre hospitalité m'ont marqué. J'ai adoré ce premier séjour à Paris.

Ich möchte Prof. Dr. Matthias Hentze für das EMBL danken. Die prägnante und kritische Natur deiner Überlegungen hat mich sehr erleuchtet. Ich werde unser kurzes Treffen nicht vergessen.

Merci aux membres volubiles de notre *Axis meeting*, Dr Daniel Zenklussen, Dr Nicole Francis, Dr Daniel (Dan) Scott et tant d'autres. Vous m'avez initié à la collégialité académique, cette chance de se rencontrer autour d'une idée, cette joie d'en débattre à en oublier le temps qui passe. D'abord intimidé, je suis peu à peu sorti de ma coquille pour rejoindre la discussion et je vous remercie d'avoir libéré ma parole.

Merci à mes amis François, Geoffroy, Pierre, Michel et les autres. J'ai toujours trouvé votre porte ouverte pour honorer ces discussions interminables et me surprendre par ces grands élans d'amour qui me chamboulent toujours.

Merci à mon incontournable maman. Tu m'as donné la vie, et ensuite nous sommes devenus des amis. Nous avons traversé beaucoup de pays, de joies et d'écueils, de naissances et de morts ensemble. *Sei davvero una donna eccezionale. Il tuo consiglio è prezioso e il tuo amore è necessario.*

Merci à ma tendre Marie et à notre petite Ysou ; vivre avec vous est un honneur réitéré chaque matin et un fou rire perpétuel. Vous m'avez fait découvrir tout un monde et je vous en serai toujours reconnaissant.

Chapitre 1 : Rétrospective historique de la biologie de l'ARN

1.1. Des origines au dogme Central de la Biologie Moléculaire (1868-1958)

L'histoire de l'ARN débute avec l'identification des acides nucléiques, attribuée à Friedrich Miescher, biologiste suisse qui isole et caractérise en 1868 des analytes riches en phosphate au sein de noyaux de leucocytes [1]. Au cours de la première moitié du XX^{e} siècle, la distinction conceptuelle entre l'ADN et l'ARN émerge progressivement grâce à d'attentifs efforts de caractérisation chimique. Il devient clair qu'un premier acide nucléique présent dans les noyaux eucaryotes, l'ADN, demeure stable en conditions basiques, tandis qu'un second, l'ARN, est dégradé par les composés alcalins. Ainsi, dès le début des années 1940, il est établi que l'ARN et l'ADN diffèrent tant au niveau de leur composition glucidique (ribose chez l'ARN et désoxyribose chez l'ADN) qu'au niveau des bases azotées qui les constituent (uracile chez l'ARN et thymine chez l'ADN).

Au niveau conceptuel, la notion d'ARN messager (ARNm) se révèle à compter de la seconde moitié du siècle et demeure intimement associée au postulat de Francis Crick, qualifié de « Dogme central de la biologie moléculaire ». Cette théorisation du transfert d'information qui orchestre les systèmes biologiques établit l'ADN en tant que support de l'hérédité, lequel est transcrit en un éphémère message d'ARN, lui-même subséquemment traduit sous forme de protéine, charpente et principal effecteur de la physiologie cellulaire [2]. C'est l'étude des gènes de globine, abondamment exprimés par les réticulocytes, qui fournit en 1958 à Schweet, Lamform et Allen les premières évidences expérimentales du rôle de l'ARNm [3], relai à la fois élusif et incontournable du processus d'expression génique.

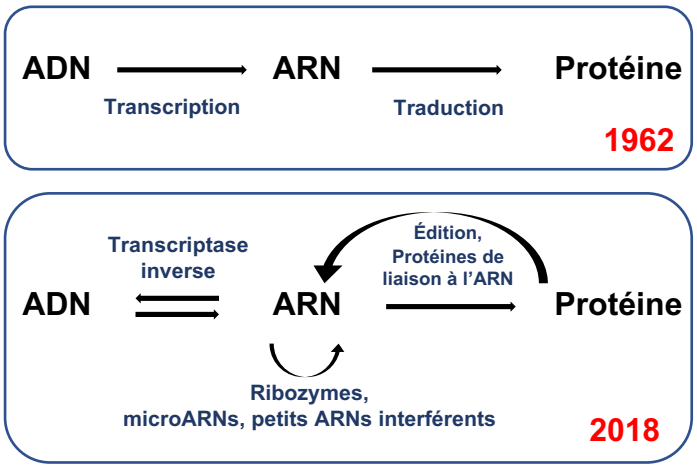


Figure 1. 1. Dogme central de la biologie moléculaire. Représentations schématiques du dogme central de la biologie moléculaire tel que postulé par Francis Crick en 1962 (haut) et tel que revu en 2018, à la lumière de quarante-cinq ans de découvertes en biologie des ARNs.

1.2. Code génétique, séquençage et transcriptase inverse (1958-1970)

Dans les années 1950, l'avènement d'approches d'ultracentrifugation pave la voie à des analyses de sédimentation qui, couplées au développement rapide de la microscopie électronique, débouchent sur l'identification des ribosomes. En parallèle, l'emploi d'acides aminés radioactifs démontre leur association rapide aux ribosomes et leur incorporation subséquente aux chaînes protéiques [1]. En 1966, ce progrès technique conduit à l'identification des polysomes et permet à Khorana, Nirenberg et Holley de découvrir le concept de codon et de décortiquer le sens de la traduction protéique, qui dépend de la lecture des brins d'ARNm par les ribosomes dans le sens 5' vers 3' [4]. L'identification des ARNs de transfert (ARNt), grâce à leur hydrosolubilité marquée et grâce à l'émergence d'approches de traduction *in vitro*, révèle le pairage entre codons et anticodons et complète le dévoilement du code génétique [5]. Au début des années 1960, la purification de l'ARN polymérase bactérienne chez *E. coli*, puis des polymérases eucaryotes, dote l'étude de l'ARN d'un outil puissant qui permettra d'en étendre considérablement la portée au cours de la décennie suivante. De plus, la purification d'ARNt, l'emploi d'une combinatoire d'endoribonucléases et le recours à l'électrophorèse conduit peu

après au premier effort fructueux de séquençage intégral d'un ARN. Quelques années plus tard, la séquence entière d'un premier génome est dévoilée, révélant les 3569 bases du bactériophage MS2. En 1970, Temin, Mizutani et Baltimore rapportent la découverte chez le virus du sarcome de Rous d'une ADN polymérase dépendante de l'ARN, qualifiée de transcriptase inverse, qui est capable de synthétiser de l'ADN à partir d'une molécule d'ARN [1]. L'identification d'une telle enzyme, qui sera plus tard notamment associée au virus de l'immunodéficience humaine (VIH), chamboule le dogme de Crick et dote la biologie moléculaire d'un outil technique de taille qui révèlera plus tard ses atouts dans l'étude de l'expression génique.

1.3. Phylogénie, maturation, épissage et ribozymes (1970-1984)

L'établissement de la séquence des ARN ribosomaux (ARNr) d'un large répertoire de microorganismes au cours des années 1970 pave la voie aux analyses phylogénétiques de Carl Woese. En 1977, il établit une nouvelle taxonomie basée sur la séquence des ARNr 16S qui définit les Archées comme un troisième domaine du Vivant [1]. Ces travaux, rendus possibles grâce aux progrès de la biologie de l'ARN, révolutionnent notre représentation de l'Arbre du Vivant. Par ailleurs, le séquençage d'un inventaire croissant de gènes et d'ARNm conduit, au début des années 1970, à l'identification de la coiffe de 5-methylguanosine et de la queue polyA, des séquences nucléotidiques qui sont ajoutées de manière post-transcriptionnelle aux ARNm, notamment pour promouvoir leur stabilité et leur traduction. De même, l'ajout d'une séquence CCA à l'extrémité 3' des ARNt permet leur association à l'acide aminé correspondant, conduisant à la formation de l'aminoacyl-ARNt, un intermédiaire important du processus de traduction protéique. La découverte de ces séquences nucléotidiques ajoutées après la transcription (*nontemplated nucleotides*) fondera un domaine de recherche fécond, celui de la maturation des ARNs.

Vers la même époque, les progrès soutenus de la cristallographie étendent notre appréciation de la complexité tridimensionnelle de l'ARN, qui peut adopter une structure secondaire, puis une structure tertiaire. L'appréciation de cette versatilité structurale mène plusieurs chercheurs, notamment Carl Woese, à envisager le potentiel catalytique des ARN dès la fin des années 1960. Si la fonction enzymatique avait jusqu'alors été strictement associée à un support peptidique, la preuve expérimentale de rôles catalytiques de l'ARN suivra vers le début

des années 1980, via l'identification d'un phénomène fondamental, l'épissage [1]. En effet, l'analyse comparative de séquences géniques et de séquences d'ARNm matures conduit Phillip Sharp et Richard Roberts à la réalisation que les gènes sont ponctués de séquences non-codantes, qualifiées d'introns, vers la fin des années 1970 [1]. Suite à la transcription d'une unité génique, les introns sont épissés via une séquence coordonnée de réactions impliquant la reconnaissance des frontières entre exons et introns proximaux, leur clivage, l'élimination des introns et la ligation des exons. De plus, l'épissage alternatif, qui correspond à l'exclusion sélective de certains exons, donne lieu à différents isoformes protéiques dont l'activité biologique peut varier, ce qui contribue à enrichir la diversité fonctionnelle du protéome. En étudiant l'épissage des ARNr chez le protozoaire *Tetrahymena thermophila*, Thomas Cech démontre en 1982 que le transcrit précurseur est capable de catalyser son propre épissage *in vitro* [6]. Un an plus tard, Sidney Altman identifie le rôle catalytique de l'ARN au sein d'un complexe de maturation des ARNr procaryotiques, la ribonucléase P [7]. Ces deux découvertes confirment l'hypothèse de Woese sur le potentiel catalytique des transcrits et établissent le concept de « ribozyme », ou ARN enzymatique. Cette propriété de l'ARN devient notamment un argument de poids en faveur de l'hypothèse du monde ARN, qui avance que les premières formes de vie apparues sur Terre auraient été constituées d'un génome d'ARN, capable d'assurer sa propre réplication et de catalyser des réactions métaboliques élémentaires.

1.4. Édition de l'ARN, télomérase, riboswtich et rétrotransposons (1984-1992)

Dès la fin des années 1980, l'étude d'ARNs mitochondriaux met en évidence l'existence de nucléotides modifiés, telle l'inosine, au sein de transcrits endogènes. Ces nucléotides non-canoniques résultent de l'activité post-transcriptionnelle d'enzymes telle ADAR et peuvent viser plus de 50% des nucléotides constituant certains ARNm chez les plastides de protozoaïres. Il s'agit d'un nouveau constat fondamental qui requiert de revisiter le dogme de Crick [8, 9].

Quelques années plus tard, en 1985, le dogme central est à nouveau ébranlé par les travaux de Boeke et Garfinkel, qui découvrent la rétrotransposition en étudiant l'élément répété Ty chez *Saccharomyces cerevisiae* [10]. Barbara McClintock, pionnière de la cytogénétique, avait été la première à identifier dès la fin des années 1960 que des fragments de chromosomes

subissent une transposition chez le maïs, en se détachant de leur locus chromosomique pour en acquérir un nouveau [11]. Or, c'est plus de vingt ans plus tard que Boecke et ses collègues démontrent que plusieurs de ces éléments transposables, qui constituent plus du tiers du génome humain, sont transcrits par la polymérase de l'hôte et nécessitent leurs intermédiaires ARN pour envahir de nouveaux loci génomiques.

En 1985, la découverte de l'activité de la télomérase par Elizabeth Blackburn et Carol Greider révèle la contribution de la biologie de l'ARN au phénomène de la sénescence et permet de résoudre le mystère de la réplication terminale des chromosomes linéaires [12]. En effet, l'extrémité des chromosomes eucaryotiques est constituée de séquences répétitives qui se voient écourtées à chaque cycle de réplication, conséquence de l'orientation 5' vers 3' de l'activité du complexe ADN polymérase. Le raccourcissement progressif des télomères agit à la manière d'une horloge interne, tel un marqueur du nombre de générations cellulaires passées, et l'érosion des télomères conduit à un arrêt réplicatif, qualifié de sénescence. Or, les lignées germinales, de même que plusieurs cancers, expriment une enzyme ribonucléoprotéique, la télomérase, qui présente une activité de transcriptase inverse spécialisée capable d'altérer ce processus en allongeant les télomères. La découverte singulière de ce ribozyme galvanise à la fois le domaine de l'oncologie moléculaire et celui de la médecine régénérative.

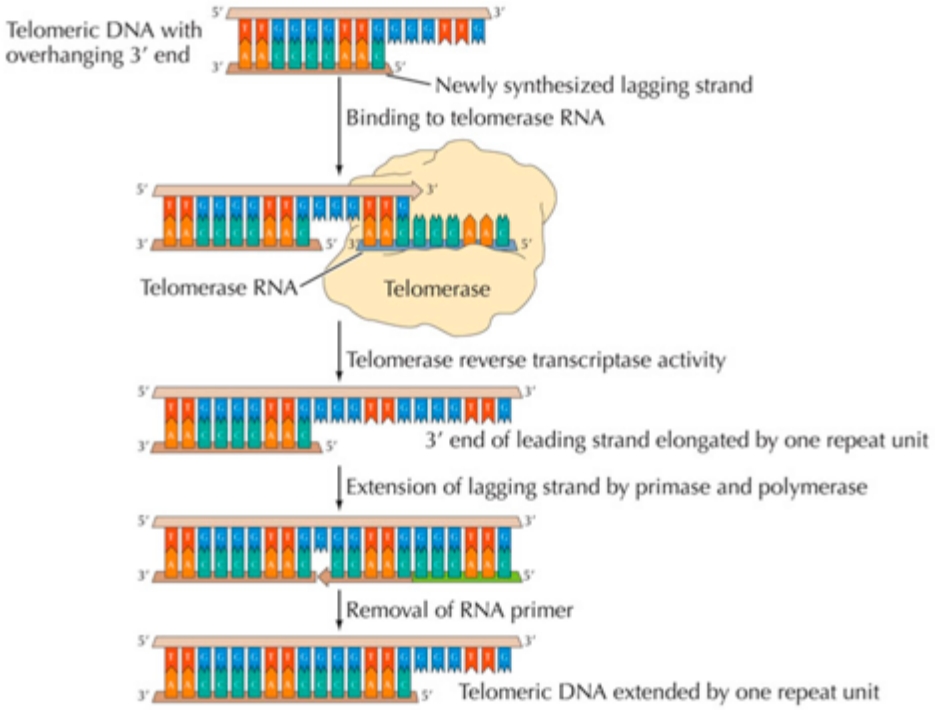


Figure 1. 2. L'activité de transcriptase inverse de la télomérase permet de réparer l'extrémité des chromosomes linéaires suite à la réplication de l'ADN. Schéma récapitulant les lésions qui résultent de la réplication incomplète parle complexe ADN polymérase à l'extrémité des chrosmomes linéaires eucaryotiques, nommées « télomères ». La télomérase, une ribonucléoprotéine dotée d'une activité de transcriptase inverse, reconnaît l'extrémité simple brin des télomère et permet de les réparer. (Figure tirée du site de l'Université de l'Oregon : http://oregonstate.edu/instruction/bi314/summer09/dnafix.html)

Vers la fin du XX^{e} siècle, le cumul des travaux effectués chez divers transcrits bactériens établit progressivement le concept de *riboswitch*, qui correspond à un aptamère d'ARN d'origine endogène. Un *riboswitch* est défini comme un ARN structuré capable d'interagir avec de petites molécules métaboliques, telles la thiamine, la flavine ou la glutamine [1]. Cette interaction conduit à un changement de conformation chez le transcrit, qui mène à son tour à un changement de son activité catalytique. À ce titre, les ribozymes se présentent un peu comme diverses protéines, notamment les récepteurs nucléaires, dont l'interaction avec un ligand conduit à un changement de conformation, qui s'accompagne en retour d'un changement de leur profil fonctionnel.

1.5. L'Ère de la transcription non-codante (1993-2017)

Bien que plusieurs découvertes de taille aient approfondi notre portrait de la biologie de l'ARN depuis le tournant du millénaire, le changement de paradigme le plus profond des dernières années concerne probablement l'ampleur de la couverture transcriptionnelle chez les eucaryotes et la diversité fonctionnelle associée aux ARN non-codants. Tandis qu'environ 1.2% des bases du génome humain encodent les acides aminés au sein de protéines, des recensements de la dernière décennie indiquent que jusqu'à 93% des bases seraient transcrites en ARN [13, 14]. En effet, l'émergence d'approches de séquençage à haut débit et la profondeur d'analyse croissante qui leur est associée ont révélé que la vaste majorité des bases des génomes eucaryotiques est transcrite. Ces avancées techniques ont dévoilé une multitude de nouvelles classes d'ARN non-codants, dont plusieurs accomplissent des fonctions cruciales. Parmi ces transcrits régulateurs, les petits ARNs non-codants occupent une place de choix.

Dans le contexte de l'étude des opérons, Jacob et Monod furent les premiers à envisager, dès 1961, l'existence chez les Eucaryotes de répresseurs capables de cibler les ARNm de manière à moduler l'expression génique. La preuve expérimentale suit, près de 30 ans plus tard, avec les travaux de Victor Ambros, qui démontre que le petit ARN encodé par le gène *lin-4* chez le nématode *C. elegans* régule l'abondance de la protéine lin-14. Sept ans plus tard, la découverte d'un nouveau régulateur, *let-7*, cimente le concept de microARN (miRNA). Ces petits transcrits endogènes constitués de 21 à 24 nucléotides sont capables de guider une ribonucléase de la famille Argonaute (AGO) vers leur cible complémentaire, typiquement un ARNm, de manière à en promouvoir le clivage ou à en déstabiliser la traduction. Plusieurs classes de petits ARNs qui fonctionnent de manière analogue sont ensuite mis à jour, notamment les petits ARN endogènes (siRNAs) et les ARNs associés aux protéines Piwi (piRNAs). De petits transcrits synthétiques sont aujourd'hui couramment utilisés en laboratoire pour cibler l'ARNm endogène qui leur est complémentaire et ainsi mener à une déplétion protéique.

Une multitude de transcrits plus longs, dont le contingent dépasserait les 30 000 au sein du transcriptome humain, ont été répertoriés depuis le début du millénaire, grâce aux approches de séquençage en profondeur [1]. Ces ARNs d'une colossale diversité fonctionnelle sont collectivement qualifiés de *long non-coding RNAs* (lncRNAs) et ne partagent guère que deux caractéristiques, soient leur longueur substantielle, plus de 200 nucléotides, et l'absence d'un cadre de lecture voué à la traduction protéique au sein de leur séquence. Certains présentent une

couverture transcriptionelle qui chevauche intégralement ou partiellement celle d'un ARNm, en orientation codante ou en orientation antisens. La question de la contribution fonctionnelle des lncRNAs à la physiologie cellulaire est controversée et la majorité des dizaines de milliers de transcrits répertoriés par le séquençage à haut débit n'ont pas encore fait l'objet d'une caractérisation fonctionnelle suffisante. Dans d'autres cas, malgré des efforts soutenus, le rôle biologique de ces transcrits non-codants demeure élusif. Ainsi, s'il a été démontré que les régions régulatrices distales qui coordonnent l'expression de nombreux gènes font l'objet d'une transcription active coïncidant avec l'expression de leur gène cible, les évidences supportant une contribution fonctionnelle des *enhancer RNAs* (eRNAs) qui en résultent demeurent faibles [15]. De même, il est trop tôt pour conclure quant à la fonctionnalité des transcrits générés en périphérie des bris double-brin de l'ADN, qualifiés de *double stranded breaks (DSB)-induced RNAs* (diRNAs), bien que leur identification ait conduit à diverses hypothèses d'une éventuelle contribution aux voies de réponse aux dommages à l'ADN. En revanche, d'autres lncRNAs ont révélé bon nombre de leurs secrets et ont fondé une littérature riche et concluante qui démontre une contribution incontournable à la régulation transcriptionnelle et post-transcriptionnelle. Ainsi, les rôles du lncRNA *XIST* (X-inactive specific transcript) dans le remodelage chromatinien qui mène à l'extinction transcriptionnelle d'un chromosome X entier chez les cellules somatiques femelles sont solidement établis [16]. Dans le même ordre d'idée, le lncRNA *HOTAIR*, qui est transcrit dans le locus des gènes Hox, s'est révélé un acteur épigénétique puissant et polyvalent, impliqué en *trans* dans la régulation de centaines de gènes cibles. Finalement, il est clairement démontré que le lncRNA *NEAT1* est requis à la formation de corps nucléaires spécifiques, les paraspeckles, où convergent divers facteurs protéiques associés à la maturation post-transcriptionnelle, notamment PSP1 et p54nrb [17].

En conclusion, l'ère post-génomique que nous traversons constitue une période particulièrement féconde pour la biologie de l'ARN. Les découvertes liées à l'ampleur de la couverture transcriptionnelle des génomes eucaryotes et aux processus de la régulation post-transcriptionelle se sont accélérées au fil des dernières décennies. Il s'agit résolument d'une époque charnière pour ce domaine d'étude et œuvrer dans le monde de la biologie de l'ARN n'a probablement jamais été aussi gratifiant qu'aujourd'hui.

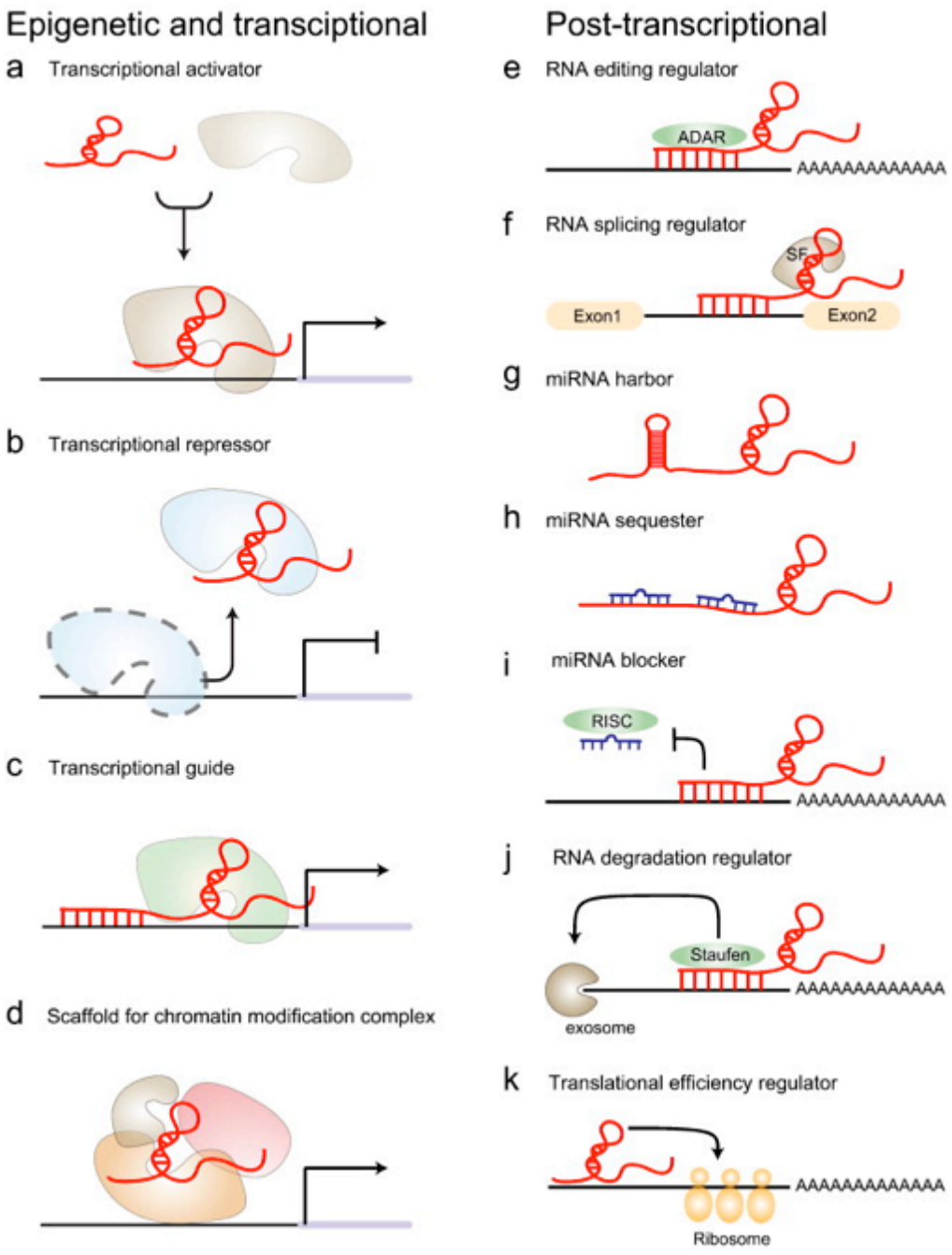


Figure 1.3. **Profils fonctionnels avérés et hypothétiques de divers lncRNAs au niveau transcriptionnel et post-transcriptionnel.** Les lncRNAs ont été associés à la régulation transcriptionnelle, dans le cadre de laquelle ils peuvent contribuer au recrutement d'activateurs ou de répresseur protéiques (a,b), notamment en orientant leur activité vers une région promotrice reconnue par complémentarité de base (c) ou en en servant de plateforme de recrutement pour l'association de facteurs de remodelage de la chromatine (d). Au niveau post-transcriptionnel, les lncRNAs peuvent s'hybrider à un ARNm complémentaire pour former une séquence d'ARN double-brin, substrat d'ADAR (e) ou de régulateurs de l'épissage (f). Les

lncRNAs peuvent également agir à titre de précurseurs de miRNAs (g), d'éponges ou de compétiteurs des miRNAs (h, i). Finalement, ils peuvent réguler l'expression génique en modulant la machinerie de dégradation des ARNs (j) ou l'efficacité de traduction (k). (Figure tirée de Yanga G, Lub X et Yuanc L. (2014) LncRNA: A link between RNA and cancer, Biochimica et Biophysica Acta.)

1.6. Principes et rôles biologiques de la localisation des ARNs

En 1983, Jeffrey et Brodeur identifient une asymétrie marquée dans la distribution subcellulaire des ARNm de la β-actine chez les embryons et les oeufs d'Ascidies, un groupe de vertébrés primitifs associés aux éponges de mer [18]. Cette découverte importante établit un nouveau domaine de la biologie des ARNs en révélant un principe fondateur : la notion de localisation des ARNs. Jusqu'alors, les canons de biologie cellulaire reposaient sur l'idée que la compartimentation des protéines serait le reflet d'un transport strictement post-traductionnel. En effet, les facteurs dotés d'une forte asymétrie spatiale, par exemple des récepteurs transmembranaires, peuvent être traduits de manière anisotrope dans l'espace cytosolique ou réticulaire et acheminés subséquemment à la périphérie cellulaire afin d'y être incorporés à la membrane cytoplasmique. Or, la découverte de Jeffrey et Brodeur sous-tend un principe alternatif : l'ARNm qui code pour de tels récepteurs pourrait lui-même faire l'objet d'un transport actif vers la périphérie de la cellule pour peu que ce transport soit couplé à une traduction localisée. Cette notion implique une économie d'énergie considérable pour la cellule. Puisqu'un ARNm peut être traduit en protéine des centaines de fois, traduire localement semble bien plus efficace que de transporter individuellement chaque peptide de manière post-traductionnelle [19]. Au-delà de l'économie d'énergie, la localisation des ARNm permet un gain en rapidité, notamment dans le système nerveux central, où la traduction d'ARNm préalablement transcrits et localisés aux synapses est souvent initiée strictement en réponse à une stimulation précise [20]. Autre avantage, la localisation des ARNm peut restreindre l'expression d'une protéine potentiellement toxique à un contexte spatiotemporel précis, tel qu'illustré par la localisation du messager *MBP* (*myelin basic protein*) chez les oligodendrocytes [19].

La découverte de Jeffrey et Brodeur est rapidement suivie d'une série d'études qui confirment et étendent le principe de localisation des ARNm à une multitude de systèmes biologiques, notamment dans le contexte de l'embryogenèse (**Figure 1.4**). En effet, la déposition

asymétrique de transcrits spécifiques sera décrite chez l'œuf de la mouche *Drosophila melanogaster* et de la grenouille *Xenopus laevis* avant la fin des années 1980 [21] [22]. Rapidement, il s'avère que l'asymétrie spatiale de certains transcrits de la Drosophile, tels *bicoid, nanos* et *oskar*, contribue à l'expression localisée de déterminants morphogénétiques. Ce processus s'avère crucial dans l'établissement de la polarisation de l'embryon le long des axes antéro-postérieur et dorso-ventral [23] [24]. Son abrogation est typiquement tératogène et conduit à des phénotypes saisissants, tel qu'illustré par le mutant *bicoid*, qui est dénué de structures anatomiques antérieures et se développe notamment en l'absence d'une tête [25] [26]. La contribution de la localisation des ARNm à l'acquisition de la structure et de l'agencement du corps embryonnaire est d'ailleurs conservée chez le Xénope. En effet, la localisation au pôle végétal de l'ARNm *VegT*, qui code pour un facteur de transcription à boîte T, induit l'identité endodermique et mésodermique, contribuant à un événement de différenciation cellulaire fondateur [27] [28].

Vers la fin des années 1980, des ARNm localisés sont également identifiés dans diverses cellules somatiques comportant une polarité marquée, notamment les fibroblastes, les oligodendrocytes, les neurones de mammifères et même la levure en division [29] [30] [31] (**Figure 1.4)**. Chez les fibroblastes motiles, l'acheminement et la traduction localisée de l'ARNm de la β-actine au lamellipode en dynamise et en oriente la chimiotaxie [29] [32]. Une observation similaire est rapportée dans le développement du système nerveux : la traduction localisée de la β-actine au cône de croissance contribue à l'orientation de la croissance axonale [33]. En effet, dans le système nerveux central, la localisation des ARNm et leur traduction localisée s'avère particulièrement répandue et adopte un rôle déterminant dans une panoplie de processus. Lors du développement cérébral, elle coordonne la réponse aux gradients chimiotaxiques et permet aux neurones de croître vers leurs partenaires synaptiques appropriés. Dans le cerveau adulte, la traduction localisée des ARNm acheminés aux synapses favorise la plasticité synaptique. De même, suite à une axotomie, la régénération des axones périphériques est favorisée par une augmentation du taux de traduction localisée [20]. Dans ce contexte, l'ARNm du facteur de transcription STAT3 est traduit aux synapses et fait ensuite l'objet d'un étonnant transport rétrograde vers le noyau, où il promeut la transcription de gènes régénératifs en réponse à la lésion axonale [34]. Par ailleurs, la principale protéine constitutive de la gaine de myéline, Mbp, est nécessaire à la transmission de l'influx nerveux et est exprimée par les oligodendrocytes qui

entourent les axones. Elle agit en compactant les membranes lipidiques, ce qui peut constituer une source de toxicité pour les assemblages membranaires fins du réticulum endoplasmique et de l'appareil de Golgi. Afin de restreindre la diffusion intracellulaire de Mbp qui pourrait s'avérer cytotoxique, son ARNm est acheminé en périphérie et traduit localement, conduisant à la formation d'une gaine de myéline étanche sans pour autant nuire à la dynamique membranaire intracellulaire (**Figure 1.4**) [35]. Finalement, la découverte du rôle clé de la localisation des transcrits *ASH1* dans le cycle reproductif de la levure du boulanger (*Saccharomyces cerevisiae*) par Pascal Chartrand, devenu depuis professeur à notre département de Biochimie, contribue dans les années 1990 à révéler l'étendue phylogénétique de la localisation des ARNm parmi les eucaryotes [36] (**Figure 1.4)**. L'ARNm *ASH1*, qui encode un répresseur transcriptionnel capable d'inhiber la transition de l'identité sexuée (*mating type*), est transporté sélectivement au bourgeon la levure en cours de division. La protéine correspondante, Ash1p, est dès lors uniquement exprimée chez la cellule-fille, qui acquiert donc un type sexuel distinct de celui de la cellule-mère.

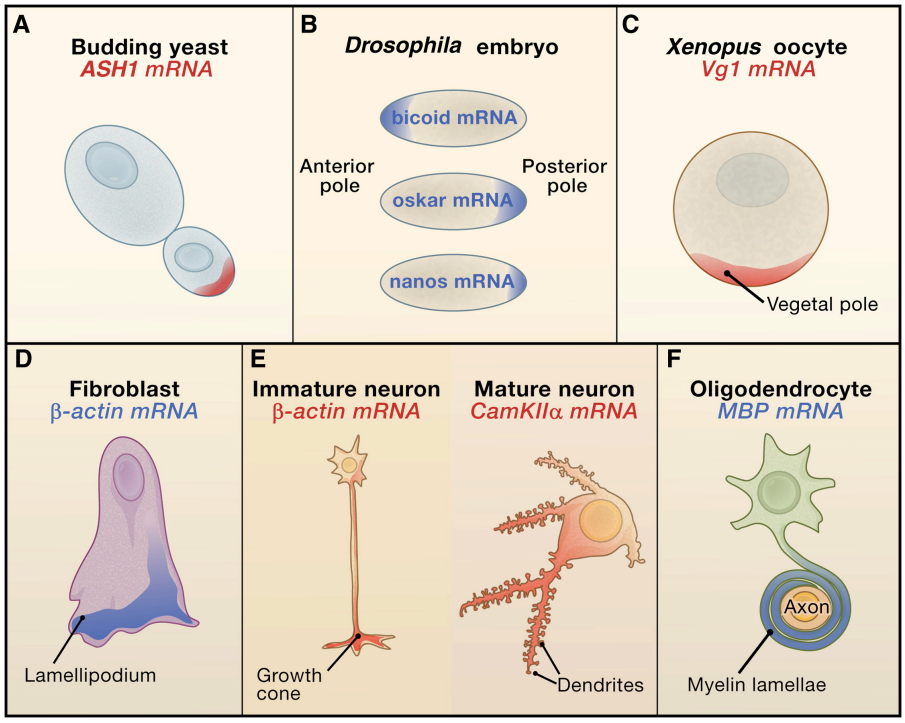


Figure 1. 4. Exemples classiques d'ARNm localisés

. (A) Chez *S. cerevisiae*, la localisation des transcrits *ASH1* contribue à l'établissement de l'identité sexuée de la cellule-fille. (B) Pendant l'embryogenèse de la Drosophile, la localisation de divers transcrits, notamment *bicoid*, *oskar* et *nanos*, détermine l'établissement de l'axe antéro-postérieur. (C) Chez l'oeuf de Xénope, la localisation de l'ARNm, *Veg1* (ou *Vg1*) au pôle végétal contribue à la différenciation tissulaire au cours de l'embryogenèse. (D) Chez les fibroblastes motiles de poulet, l'ARNm de la β-actine est localisé aux lamellipodes de manière à promouvoir la chimiotaxie. (E) L'ARNm de la β-actine est également localisé aux cônes de croissance des neurones de mammifères, tandis que l'ARNm de la kinase CamKIIα se situe aux dendrites distales, propriété qui contribue à la transmission de l'influx nerveux. (F) L'ARNm de la MBP est localisé et traduit à proximité de la membrane cytoplasmique des oligodendrocytes, de manière à prévenir la perturbation du réseau membranaire intracellulaire. (Figure tirée de « Martin, K. C., & Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. Cell, 136(4), 719-730. »)

1.7. Mécanismes de la localisation des ARNs

Peu après l'identification du phénomène de la localisation des ARNm, divers groupes ont tenté de développer des approches expérimentales pour comprendre les codes et les déterminants de ces processus. Ces notions ont fait l'objet de plusieurs excellentes revues de la littérature [19, 37-44] [45]. Ces efforts ont conduit à l'émergence d'un paradigme général: les ARNm localisés comportent des éléments de reconnaissance, qualifiés de *cis-regulatory motifs* (CRMs) ou *cis-acting motifs*. Ces CRMs peuvent être constitués de quelques nucléotides adjacents ou au contraire formés de structures secondaires complexes impliquant des nucléotides éloignés. Souvent retrouvés au sein de la séquence non-codante en 3' (3'UTR) des ARNm, les CRMs sont sélectivement reconnus par des facteurs en *trans*, typiquement des protéines de fixation aux ARNs, ou *RNA-binding proteins* (RBPs) (**Figure 1.5**). La reconnaissance des CRMs par les RBPs peut dépendre des étapes nucléaires de maturation des transcrits. En effet, quelques exemples d'isoformes d'un même ARNm dotés de localisations divergentes ont été répertoriés; c'est notamment le cas de *BDNF*, le *Brain-derived neurotrophic factor*, dont un seul isoforme est ciblé aux dendrites dans les motoneurones de l'hippocampe du rat [46]. Des événements nucléaires d'épissage alternatif ou de polyadénylation alternative donnant lieu à l'inclusion sélective d'un CRM dans un des isoformes et à son exclusion dans l'autre sont à l'origine de ces patrons de localisation distincts [19]. Une fois dans le cytoplasme, les complexes ribonucléoprotéiques formés par les RBPs associées aux ARNs matures figurent en général au sein de structures plus larges, qualifiées de granules de transport. Ces granules font à leur tour l'objet d'un transport actif le long du cytosquelette d'actine ou des microtubules jusqu'à leur destination finale (**Figure 1.5**) [47]. Les granules y seront ensuite désassemblées et l'ARNm ancré *in situ* jusqu'à son éventuelle traduction.

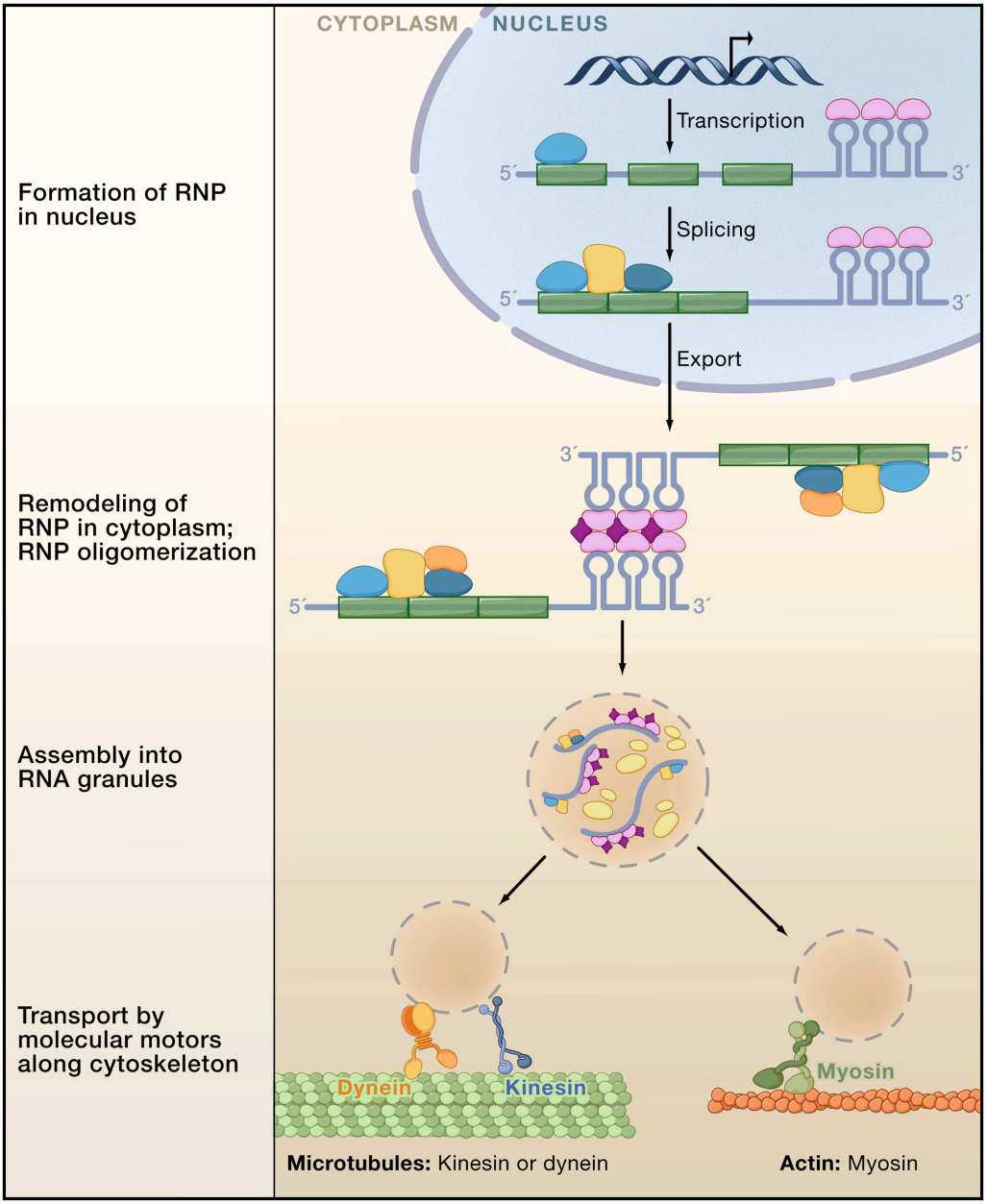


Figure 1. 5. Mécanisme séquentiel général de la localisation des ARNm

Les CRMs, souvent retrouvés dans le 3'UTR des ARNm (représentés ici comme des tige-boucles) sont reconnus par des RBPs spécifiques (en rose) de manière co-transcriptionnelle ou post-transcriptionnelle. Les étapes de maturation nucléaire de l'ARN, tels l'épissage et la polyadénylation alternative, peuvent par ailleurs s'avérer déterminants à l'inclusion ou à l'exclusion d'un CRM au sein du messager mature. Suite à l'export nucléaire, les complexes ribonucléoprotéiques formés par les ARNm et leurs RBPs associées peuvent faire l'objet d'un remodelage, impliquant par exemple l'oligomérisation des transcrits (représentée ici comme une dimérisation via l'interaction de tige-boucles). Les ribonucléoprotéines sont ensuite incorporées au sein de granules de transport, qui sont activement acheminées à leur destination grâce au déplacement de moteurs protéiques le long du cytosquelette. (Figure tirée de « Martin, K. C., &

Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. Cell, 136(4), 719-730. »)

Décrit dès la fin des années 1980, le mécanisme de la localisation de l'ARNm *bicoid* offre une illustration probante des phénomènes évoqués ci-haut [48]. En recourant à la transgénèse pour exprimer chez l'ovaire de Drosophile diverses versions tronquées du transcrit, MacDonald et Struhl ont pu identifier une séquence de 625 nt présente dans le 3'UTR du transcrit qui détermine sa localisation au pôle antérieur de l'oocyte. Au sein de cette séquence, ils ont identifié les éléments de reconnaissance de *bicoid*, ou BLEs (*bicoid localization elements*), dont BLE1, formé de 50 nt qui acquièrent spontanément une structure secondaire en tige-boucle. Cet élément s'est avéré nécessaire au transport de *bicoid* depuis les cellules nourricières ovariennes jusqu'à l'oocyte, destiné à devenir un embryon après la fécondation [49]. D'autres BLEs ont plus tard été associés au transport de *bicoid* dans l'oocyte lui-même ainsi qu'à son ancrage au pôle antérieur [50]. Des mutations choisies de manière à préserver la structure secondaire en tige-boucle ne perturbent pas la localisation de *bicoid*, ce qui indique que, dans ce cas précis, c'est bien cette structure et non la séquence qui fait l'objet d'une reconnaissance spécifique [51]. Subséquemment, des analyses *in vitro* ont révélé la dimérisation des transcrits *bicoid*, par le biais d'interactions spécifiques entre deux tige-boucles [51]. *In vivo*, cette dimérisation est elle-même nécessaire et suffisante à la reconnaissance de *bicoid* par la RBP Staufen, dont l'expression est requise à la localisation de *bicoid* au pôle antérieur lors des stades tardifs de l'oogenèse [52] [53]. L'exemple de *bicoid* révèle donc la complexité et la diversité des phénomènes qui sous-tiennent la localisation d'un seul ARN. Ainsi, afin de comprendre pleinement le mécanisme de localisation de *bicoid*, une dizaine d'études recourant à des essais *in vitro*, de la mutagenèse dirigée et de la transgénèse ont été nécessaires. Il est à noter que les exemples classiques d'ARN localisés discutés dans la section 1.6 tels *nanos*, *VegT*, *MBP*, ou *ASH1* ont chacun fait l'objet d'une imposante série d'études ayant permis de dévoiler le mécanisme de leur localisation [19].

1.8. Méthodologies expérimentales appliquées à l'étude de la localisation des ARN

Nous avons énuméré dans la section 1.6 quelques exemples classiques d'ARNm localisés et avons évoqué dans la section 1.7 la complexité des mécanismes responsables de la localisation de l'un d'entre eux, *bicoid*. Or, dans la dernière décennie, le développement de diverses techniques permettant de documenter la localisation des ARNs a largement étendu l'inventaire des ARNs localisés. En effet, environ 100 ARNs localisés étaient répertoriés au début des années 2000, un nombre qui a depuis crû à plusieurs milliers, surtout grâce à l'adoption des techniques de séquençage à haut débit [54]. Dans bien des cas, les mécanismes responsables de la localisation de ces nouveaux cas d'ARNs demeurent à définir. Dans cette section, nous évoquerons brièvement les approches méthodologiques ayant permis d'identifier de nouveaux ARNs localisés.

L'hybridation *in situ* (*in situ* hybridization - *ISH*) est sans conteste l'approche classique permettant la détection de séquences spécifiques d'acides nucléiques au sein d'un échantillon fixé. Développée en 1969, cette technique implique l'utilisation d'une sonde marquée complémentaire à la séquence ciblée [55]. D'abord de nature radioactive, les méthodes de marquage ont évolué vers la fin des années 1970 pour exploiter la forte affinité de la streptavidine pour la biotine, ou l'immunogénicité marquée d'un épitope puissant, la digoxigénine [56]. Couplées à l'utilisation de fluorophores, ces méthodes sont qualifiées d'hybridation *in situ* en fluorescence (fluorescence *in situ* hybridization - FISH). Bien que les approches traditionnelles de FISH soient limitées à la détection de quelques ARNs spécifiques par échantillon, il est possible d'optimiser cette méthode pour sonder des dizaines d'échantillons dans la même expérience, par exemple dans un format de plaque à 96 puits. Cette astuce a permis à mon directeur de thèse, Éric Lécuyer, de cribler plus de 3000 transcrits chez l'embryon de Drosophile, environ le quart du transcriptome annoté [57] [58]. Cette étude a révélé une proportion étonnement élevée d'ARNs dotés d'une localisation asymétrique (71%) [57] (**Figure 1.6**). Notre laboratoire a publié plusieurs versions du protocole utilisé dans cette étude, auquel j'ai recouru pour les analyses présentées dans le chapitre 8 [59-62].

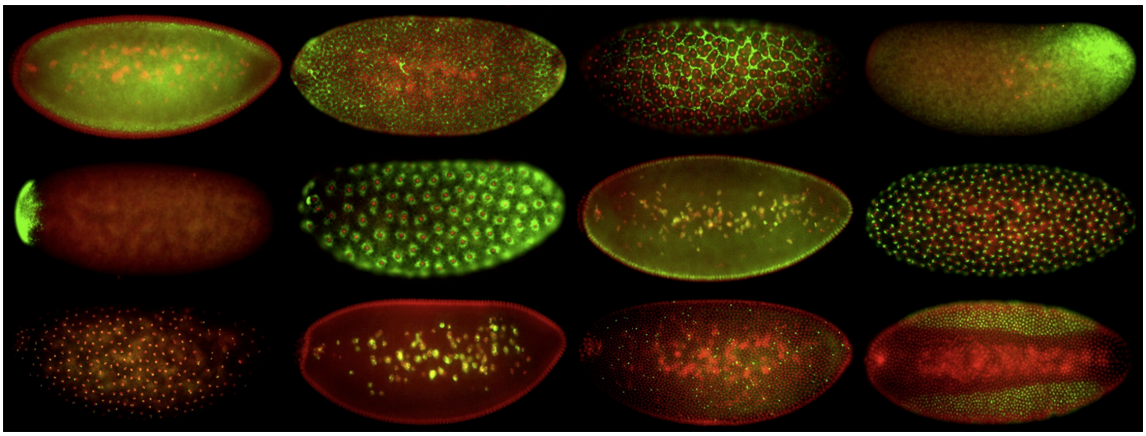


Figure 1. 6. Localisation subcellulaire d'ARNm variés chez l'embryon de Drosophile. Exemples d'embryons traités par FISH de manière à révéler des patrons de localisation variés (ARN en vert, ADN en rouge). Le pôle antérieur est à gauche. (Tiré de Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., ... & Krause, H. M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell*, *131*(1), 174-187.)

Par ailleurs, diverses approches d'imagerie des ARNs ont vu le jour dans les décennies qui ont suivi l'émergence du FISH. Ces méthodologies ont fait l'objet de plusieurs revues de la littérature de qualité [63, 64]. Il est maintenant possible de quantifier les transcrits individuels par des approches dites « *single-molecule* » à la fois sur des échantillons fixés et des organismes vivants, c'est-à-dire en temps réel, ou « *live imaging* ». Le système MS2 constitue un exemple particulièrement puissant de telles approches. Je le décris en détail dans le chapitre 3, afin de couvrir ses contributions à notre compréhension de l'activation du génome zygotique de la Drosophile. Brièvement, le système MS2 est une innovation du laboratoire Singer dont la stratégie consiste à exprimer des transcrits de fusion constitués de la séquence cible flanquée de répétitions des motifs tige-boucles *MS2* ou *PP7*, qui sont exprimées de manière endogène par certains bactériophages [36]. En parallèle, un vecteur ou une insertion chromosomique permet l'expression d'une protéine de fusion, constituée d'une protéine fluorescente telle GFP ou mCherry et de la séquence MCP, qui présente une forte affinité pour les motifs *MS2* [65] (**Figure 1.7A**). L'expression concomitante de ce transcrit de fusion, qui sert de la cible, et de cette protéine de fusion, qui agit comme une sonde, permet de recruter l'activité fluorescente au niveau du transcrit et donc de documenter en temps réel divers aspects de sa dynamique, par exemple son transport. Depuis son développement chez *S. cerevisiae* vers la fin des années 1990,

le système MS2 a bénéficié d'une adoption enthousiaste, conduisant à son adaptation chez divers organismes, notamment la Drosophile [66], la souris [67], le nématode *C. elegans* [68], ou encore le poisson-zèbre [69].

Des analogues ribonucléotidiques de la protéine GFP ont également été développés au début de la décennie 2010, dont le plus connu est l'aptamère *Spinach* [70-72]. Il s'agit d'une courte séquence qui présente une affinité marquée pour un fluorophore et permet de suivre des transcrits endogènes par transgenèse (**Figure 1.7B**). En utilisant des balises moléculaires (*molecular beacons*) en tige-boucle dont les extrémités 5' et 3' sont liées à un fluorophore d'un côté et à un agent d'extinction (*quencher*) de l'autre, il est également possible de marquer des transcrits spécifiques en temps réel [73] (**Figure 1.7C**). Finalement, des approches de micro-injections de transcrits marqués et produits *in vitro* ont permis de mettre en évidence certaines propriétés du transport des ARNs chez l'embryon de Drosophile [74] (**Figure 1.7D**).

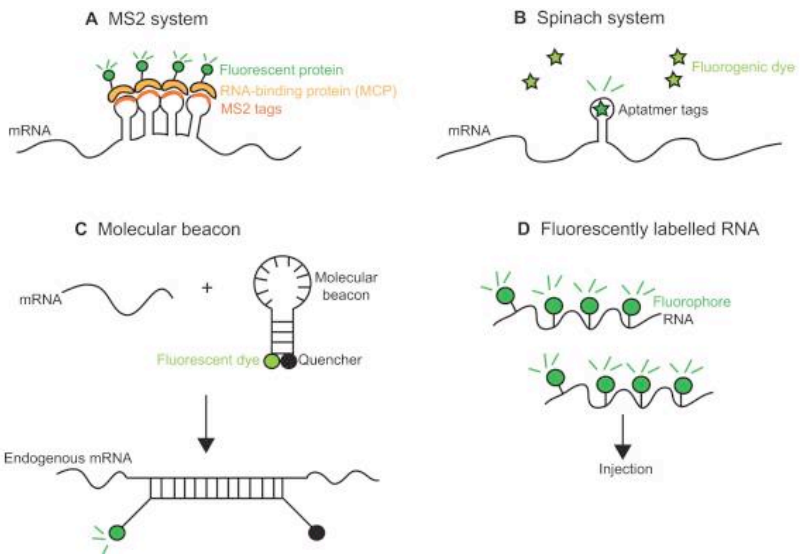


Figure 1. 7. Exemples d'outils méthodologiques dans l'imagerie de l'ARN.

(A) Représentation du système MS2, où la protéine de fusion MCP-GFP s'associe aux tige-boucles *MS2*. (B) Schéma de la tige-boucle de l'aptamère Spinach complexée à son fluorophore. (C) Représentation d'une balise moléculaire et de son changement de conformation suite à la reconnaissance de sa cible et leur hybridation subséquente. (D) Dessin d'un transcrit marqué d'un fluorophore destiné à une micro-injection. (Figure tirée de Medioni, C., Mowry, K., & Besse, F. (2012). Principles and roles of mRNA localization in animal development. Development, 139(18), 3263-3276.)

Aux avancées spectaculaires en imagerie des ARNs qui ont ponctué les dernières décennies s'ajoute le développement d'approches biochimiques pour purifier certaines structures subcellulaires. Ainsi les transcrits associés à des structures aussi variées que l'appareil mitotique [75], les dendrites [76], les pseudopodes [77] ou les vésicules extracellulaires [78-81] ont fait l'objet d'un profilage transcritomique. Mon travail s'inscrit dans cette lignée. Le potentiel des approches de purification couplées à la transcriptomique est donc discuté de manière plus détaillée dans l'introduction du chapitre 4, dédié au protocole CeFra-seq.

Références

[1] de Chadarevian S, Rheinberger HJ. Introduction. History of molecular biology. Stud Hist Philos Biol Biomed Sci. 2009;40:4-5.
[2] Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. 1953;171:737-8.
[3] Schweet R, Lamfrom H, Allen E. The Synthesis of Hemoglobin in a Cell-Free System. Proc Natl Acad Sci U S A. 1958;44:1029-35.
[4] Khorana HG, Buchi H, Ghosh H, Gupta N, Jacob TM, Kossel H, et al. Polynucleotide synthesis and the genetic code. Cold Spring Harb Symp Quant Biol. 1966;31:39-49.
[5] Rich A, RajBhandary UL. Transfer RNA: molecular structure, sequence, and properties. Annu Rev Biochem. 1976;45:805-60.
[6] Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell. 1982;31:147-57.
[7] Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell. 1983;35:849-57.
[8] Kuntz M, Camara B, Weil JH, Schantz R. The psbL gene from bell pepper (Capsicum annuum): plastid RNA editing also occurs in non-photosynthetic chromoplasts. Plant Mol Biol. 1992;20:1185-8.
[9] Bock R, Hermann M, Fuchs M. Identification of critical nucleotide positions for plastid RNA editing site recognition. RNA. 1997;3:1194-200.
[10] Boeke JD, Garfinkel DJ, Styles CA, Fink GR. Ty elements transpose through an RNA intermediate. Cell. 1985;40:491-500.
[11] Mcclintock B. Control of Gene Action in Maize. Brookhaven Sym Biol. 1965:162-+.
[12] Greider CW, Blackburn EH. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell. 1985;43:405-13.
[13] Clark MB, Amaral PP, Schlesinger FJ, Dinger ME, Taft RJ, Rinn JL, et al. The reality of pervasive transcription. PLoS Biol. 2011;9:e1000625; discussion e1102.

[14] Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SP, et al. Large-scale transcriptional activity in chromosomes 21 and 22. Science. 2002;296:916-9.
[15] Lam MT, Cho H, Heinz S, Benner C, Kaikkonen MU, Tanaka-Oishi Y, et al. Negative Regulation of Enhancer-Associated RNA in Macrophages. Faseb J. 2012;26.
[16] Rastan S. X chromosome inactivation and the Xist gene. Curr Opin Genet Dev. 1994;4:292-7.
[17] Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell. 2009;33:717-26.
[18] Jeffery WR, Tomlinson CR, Brodeur RD. Localization of actin messenger RNA during early ascidian development. Dev Biol. 1983;99:408-17.
[19] Martin KC, Ephrussi A. mRNA localization: gene expression in the spatial dimension. Cell. 2009;136:719-30.
[20] Jung H, Yoon BC, Holt CE. Axonal mRNA localization and local protein synthesis in nervous system assembly, maintenance and repair. Nat Rev Neurosci. 2012;13:308-24.
[21] Rebagliati MR, Weeks DL, Harvey RP, Melton DA. Identification and cloning of localized maternal RNAs from Xenopus eggs. Cell. 1985;42:769-77.
[22] Frigerio G, Burri M, Bopp D, Baumgartner S, Noll M. Structure of the segmentation gene paired and the Drosophila PRD gene set as part of a gene network. Cell. 1986;47:735-46.
[23] Berleth T, Burri M, Thoma G, Bopp D, Richstein S, Frigerio G, et al. The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. EMBO J. 1988;7:1749-56.
[24] Ephrussi A, Dickinson LK, Lehmann R. Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. Cell. 1991;66:37-50.
[25] Driever W, Siegel V, Nusslein-Volhard C. Autonomous determination of anterior structures in the early Drosophila embryo by the bicoid morphogen. Development. 1990;109:811-20.
[26] Peng JC, Wang L, Wikramanayake AH. Origins of anterior-posterior polarity by localized activation of Disheveled. Mol Reprod Dev. 2017;84:443.
[27] Zhang J, King ML. Xenopus VegT RNA is localized to the vegetal cortex during oogenesis and encodes a novel T-box transcription factor involved in mesodermal patterning. Development. 1996;122:4119-29.

[28] Zhang J, Houston DW, King ML, Payne C, Wylie C, Heasman J. The role of maternal VegT in establishing the primary germ layers in Xenopus embryos. Cell. 1998;94:515-24.
[29] Lawrence JB, Singer RH. Intracellular localization of messenger RNAs for cytoskeletal proteins. Cell. 1986;45:407-15.
[30] Trapp BD, Moench T, Pulley M, Barbosa E, Tennekoon G, Griffin J. Spatial segregation of mRNA encoding myelin-specific proteins. Proc Natl Acad Sci U S A. 1987;84:7773-7.
[31] Garner CC, Tucker RP, Matus A. Selective localization of messenger RNA for cytoskeletal protein MAP2 in dendrites. Nature. 1988;336:674-7.
[32] Kislauskis EH, Zhu X, Singer RH. beta-Actin messenger RNA localization and protein synthesis augment cell motility. J Cell Biol. 1997;136:1263-70.
[33] Leung KM, van Horck FP, Lin AC, Allison R, Standart N, Holt CE. Asymmetrical beta-actin mRNA translation in growth cones mediates attractive turning to netrin-1. Nat Neurosci. 2006;9:1247-56.
[34] Selvaraj BT, Frank N, Bender FL, Asan E, Sendtner M. Local axonal function of STAT3 rescues axon degeneration in the pmn model of motoneuron disease. J Cell Biol. 2012;199:437-51.
[35] Ainger K, Avossa D, Diana AS, Barry C, Barbarese E, Carson JH. Transport and localization elements in myelin basic protein mRNA. J Cell Biol. 1997;138:1077-87.
[36] Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. Localization of ASH1 mRNA particles in living yeast. Mol Cell. 1998;2:437-45.
[37] Lecuyer E, Yoshida H, Krause HM. Global implications of mRNA localization pathways in cellular organization. Curr Opin Cell Biol. 2009;21:409-15.
[38] Tekotte H, Davis I. Intracellular mRNA localization: motors move messages. Trends Genet. 2002;18:636-42.
[39] Singer RH. The cytoskeleton and mRNA localization. Curr Opin Cell Biol. 1992;4:15-9.
[40] Kloc M, Zearfoss NR, Etkin LD. Mechanisms of subcellular mRNA localization. Cell. 2002;108:533-44.
[41] Kislauskis EH, Singer RH. Determinants of mRNA localization. Curr Opin Cell Biol. 1992;4:975-8.
[42] Jansen RP. mRNA localization: message on the move. Nat Rev Mol Cell Biol. 2001;2:247-56.

[43] Du TG, Schmid M, Jansen RP. Why cells move messages: The biological functions of mRNA localization. Semin Cell Dev Biol. 2007.
[44] Czaplinski K, Singer RH. Pathways for mRNA localization in the cytoplasm. Trends Biochem Sci. 2006;31:687-93.
[45] Bullock SL. Messengers, motors and mysteries: sorting of eukaryotic mRNAs by cytoskeletal transport. Biochem Soc Trans. 2011;39:1161-5.
[46] An JJ, Gharami K, Liao GY, Woo NH, Lau AG, Vanevski F, et al. Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. Cell. 2008;134:175-87.
[47] Kanai Y, Dohmae N, Hirokawa N. Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. Neuron. 2004;43:513-25.
[48] Macdonald PM, Struhl G. cis-acting sequences responsible for anterior localization of bicoid mRNA in Drosophila embryos. Nature. 1988;336:595-8.
[49] Macdonald PM, Kerr K, Smith JL, Leask A. RNA regulatory element BLE1 directs the early steps of bicoid mRNA localization. Development. 1993;118:1233-43.
[50] Macdonald PM, Kerr K. Redundant RNA recognition events in bicoid mRNA localization. RNA. 1997;3:1413-20.
[51] Ferrandon D, Koch I, Westhof E, Nusslein-Volhard C. RNA-RNA interaction is required for the formation of specific bicoid mRNA 3' UTR-STAUFEN ribonucleoprotein particles. EMBO J. 1997;16:1751-8.
[52] St Johnston D, Beuchle D, Nusslein-Volhard C. Staufen, a gene required to localize maternal RNAs in the Drosophila egg. Cell. 1991;66:51-63.
[53] Weil TT, Forrest KM, Gavis ER. Localization of bicoid mRNA in late oocytes is maintained by continual active transport. Dev Cell. 2006;11:251-62.
[54] Medioni C, Mowry K, Besse F. Principles and roles of mRNA localization in animal development. Development. 2012;139:3263-76.
[55] Pardue ML, Gall JG. Molecular hybridization of radioactive DNA to the DNA of cytological preparations. Proc Natl Acad Sci U S A. 1969;64:600-4.
[56] Bauman JG, Wiegant J, Borst P, van Duijn P. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. Exp Cell Res. 1980;128:485-90.

[57] Lecuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, et al. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell. 2007;131:174-87.
[58] Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018;46:D754-D61.
[59] Lecuyer E, Parthasarathy N, Krause HM. Fluorescent in situ hybridization protocols in Drosophila embryos and tissues. Methods Mol Biol. 2008;420:289-302.
[60] Lecuyer E. High resolution fluorescent in situ hybridization in Drosophila. Methods Mol Biol. 2011;714:31-47.
[61] Cédric Diot AC, Éric Lécuyer. Visualizing RNA localization dynamics in Drosophila embryos, tissues and cultured cells. Methods. 2017.
[62] Legendre F, Cody N, Iampietro C, Bergalet J, Lefebvre FA, Moquin-Beaudry G, et al. Whole mount RNA fluorescent in situ hybridization of Drosophila embryos. J Vis Exp. 2013:e50057.
[63] Weil TT, Parton RM, Davis I. Making the message clear: visualizing mRNA localization. Trends Cell Biol. 2010;20:380-90.
[64] Gallardo F, Chartrand P. Visualizing mRNAs in fixed and living yeast cells. Methods Mol Biol. 2011;714:203-19.
[65] Fouts DE, True HL, Celander DW. Functional recognition of fragmented operator sites by R17/MS2 coat protein, a translational repressor. Nucleic Acids Res. 1997;25:4464-73.
[66] Belaya K, St Johnston D. Using the mRNA-MS2/MS2CP-FP system to study mRNA transport during Drosophila oogenesis. Methods Mol Biol. 2011;714:265-83.
[67] Park HY, Lim H, Yoon YJ, Follenzi A, Nwokafor C, Lopez-Jones M, et al. Visualization of dynamics of single endogenous mRNA labeled in live mouse. Science. 2014;343:422-4.
[68] Yan D, Wu Z, Chisholm AD, Jin Y. The DLK-1 kinase promotes mRNA stability and local translation in C. elegans synapses and axon regeneration. Cell. 2009;138:1005-18.
[69] Yasuda K, Kotani T, Ota R, Yamashita M. Transgenic zebrafish reveals novel mechanisms of translational control of cyclin B1 mRNA in oocytes. Dev Biol. 2010;348:76-86.
[70] Paige JS, Wu KY, Jaffrey SR. RNA mimics of green fluorescent protein. Science. 2011;333:642-6.

[71] Warner KD, Chen MC, Song W, Strack RL, Thorn A, Jaffrey SR, et al. Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. Nat Struct Mol Biol. 2014;21:658-63.
[72] You M, Jaffrey SR. Structure and Mechanism of RNA Mimics of Green Fluorescent Protein. Annu Rev Biophys. 2015;44:187-206.
[73] Monroy-Contreras R, Vaca L. Molecular beacons: powerful tools for imaging RNA in living cells. J Nucleic Acids. 2011;2011:741723.
[74] Amrute-Nayak M, Bullock SL. Single-molecule assays reveal that RNA localization signals regulate dynein-dynactin copy number on individual transcript cargoes. Nat Cell Biol. 2012;14:416-23.
[75] Blower MD, Feric E, Weis K, Heald R. Genome-wide analysis demonstrates conserved localization of messenger RNAs to mitotic microtubules. J Cell Biol. 2007;179:1365-73.
[76] Moccia R, Chen D, Lyles V, Kapuya E, E Y, Kalachikov S, et al. An unbiased cDNA library prepared from isolated Aplysia sensory neuron processes is enriched for cytoskeletal and translational mRNAs. J Neurosci. 2003;23:9409-17.
[77] Mili S, Moissoglu K, Macara IG. Genome-wide screen reveals APC-associated RNAs enriched in cell protrusions. Nature. 2008;453:115-9.
[78] Choi DS, Kim DK, Kim YK, Gho YS. Proteomics, transcriptomics and lipidomics of exosomes and ectosomes. Proteomics. 2013;13:1554-71.
[79] Chevillet JR, Kang Q, Ruf IK, Briggs HA, Vojtech LN, Hughes SM, et al. Quantitative and stoichiometric analysis of the microRNA content of exosomes. P Natl Acad Sci USA. 2014;111:14888-93.
[80] Lasser C, O'Neil SE, Ekerljung L, Ekstrom K, Sjostrand M, Lotvall J. RNA-containing exosomes in human nasal secretions. Am J Rhinol Allergy. 2010;25:89-93.
[81] Lambertz U, Ovando MEO, Vasconcelos EJR, Unrau PJ, Myler PJ, Reiner NE. Small RNAs derived from tRNAs and rRNAs are highly enriched in exosomes from both old and new world Leishmania providing evidence for conserved exosomal RNA Packaging. Bmc Genomics. 2015;16.

Préface au chapitre 2

Ce chapitre d'introduction est présenté sous la forme d'un article de revue publié dans *Frontiers in Microbiology* (Référence : Lefebvre, F. A., & Lécuyer, E. (2017). Small Luggage for a Long Journey: Transfer of Vesicle-Enclosed Small RNA in Interspecies Communication. Frontiers in microbiology, 8, 377.)

L'article traite de la contribution des petits ARNs non-codants dans les phénomènes de communication entre espèces. Il explore le rôle possible des vésicules extracellulaires comme véhicules de communication intercellulaire. À ce titre, il fait état des propeiétés généerales de ces structures en insistant sur les mécanismes impliqués dans le ciblage de transcrits spécifiques aux vésicules extracellulaires.

Dans le contexte de la présente thèse, les notions exposées dans cet article de revue permettent de contextualiser mes travaux portant sur les vésicules extracellulaires (Chapitres 4, 5 et 6).

J'ai écrit l'intégralité du manuscrit, qui a fait l'objet d'une relecture et d'une série de corrections proposées par mon directeur de thèse, Dr Éric Lécuyer, dans le contexte de sa soumission à *Frontiers in Microbiology*.

Chapitre 2 : Small Luggage for a Long Journey: Small RNA and Membrane Vesicles in Interspecies Communication (Article #1)

Chapitre 2 : Article #1 (Revue de la littérature)

Small Luggage for a Long Journey: Small RNA and Membrane Vesicles in Interspecies Communication

Fabio Alexis Lefebvre[1,2], Éric Lécuyer[1,2,3,4]

1- Institut de Recherches Cliniques de Montréal (IRCM)
Montréal, Québec, Canada

2- Département de Biochimie
Université de Montréal, Montréal, Québec, Canada

3- Division of Experimental Medicine
McGill University, Montréal, Québec, Canada

4- Address correspondences to: Dr. Eric Lécuyer
IRCM, RNA Biology Laboratory
110 Avenue des Pins, Ouest
Montréal, Québec, Canada
H2W 1R7
Tel: 514-987-5646, Fax: 514-987-5752
Email: Eric.Lecuyer@ircm.qc.ca

Abstract

In the evolutionary arms race, symbionts have evolved means to modulate each other's physiology, oftentimes through the dissemination of biological signals. Beyond small molecules and proteins, recent evidence shows that small RNA molecules are transferred between organisms and transmit functional RNA interference signals across biological species. However, the mechanisms through which specific RNAs involved in cross-species communication are sorted for secretion and protected from degradation in the environment remain largely enigmatic. Over the last decade, extracellular vesicles have emerged as prominent vehicles of biological signals. They can stabilize specific RNA transcripts in biological fluids and selectively deliver them to recipient cells. Here, we review examples of small RNA transfers between plants and bacterial, fungal and animal symbionts. We also discuss the transmission of RNA interference signals from intestinal cells to populations of the gut microbiota, along with its roles in intestinal homeostasis. We suggest that extracellular vesicles may contribute to inter-species crosstalk mediated by small RNA. We review the mechanisms of RNA sorting to extracellular vesicles and evaluate their relevance in cross-species communication by discussing conservation, stability, stoichiometry and co-occurrence of vesicles with alternative communication vehicles.

Introduction

The ‘one gene, one enzyme’ paradigm has long dominated our understanding of molecular biology. Although peptides are key effectors of cell physiology, strictly protein-centrist portraits of life have encountered early criticism and been deemed reductive since the 1950s [1]. In the post-genomic era, it has become increasingly clear that the bulk of eukaryotic genomes – loci previously dubbed ‘junk DNA’ or ‘dark matter’ - undergo pervasive transcription, yielding thousands of non-coding (nc)RNAs, many of which are conserved and tissue-specific [2-5]. In particular, small non-coding (s)RNAs transcribed from intergenic, intronic and repeated regions can exert RNA interference (RNAi) by guiding Argonaute ribonucleases (RNAses) to specific complementary targets[6]. The pivotal role of sRNA in a broad range of biological contexts is well established: estimates suggest that up to 60% of mammalian mRNA is subjected to RNAi by sRNA [7].

Symbiotic relationships favour the intricate proximity of multiple species in biological niches. To sustain the evolutionary arms race, symbionts have evolved means to influence each other via secreted signals. In line with the ‘one gene, one enzyme’ paradigm, communication across species was long taught to strictly involve peptides and small metabolites. Over the last decade, however, reports of communication across species via transfers of RNA silencing signals have surged in diverse biological niches, prompting a re-evaluation of cross-species communication[8] [9].

Naked RNA transcripts are rapidly degraded in the human systemic circulation[10] but extracellular vesicles (EVs), ribonucleoprotein complexes (RNPs) and lipoproteins can stabilize transcripts and protect them from RNAse degradation[11-14]. Microarray and deep-sequencing approaches have revealed that repertoires of secreted sRNA don't mirror cellular populations, suggesting the involvement of selective sorting mechanisms[11-14]. Various protein components of RNAi machinery have been found in mammalian EVs and can perform cell-independent sRNA maturation [15]. RNA silencing activity can be transferred across tissues, with emerging implications in early development[16, 17], cancer biology[15, 18], immunology[19], regenerative medicine [20] and gene therapy[21]. Recent reports show that EVs enable inter-organismal and long-range transfers of functional RNA and protein signals respectively in *C.*

elegans [22] and mammals [23], suggesting EVs may contribute to cross-species transfer of RNA silencing activity.

Here, we survey the properties of eukaryotic sRNA and review emerging evidence of inter-organismal RNAi across species and kingdoms, including crosstalk between plants, bacterial, fungal and metazoan pathogens and host-microbiota interactions in the gut. We discuss the mechanisms of RNA sorting to mammalian EVs for secretion. We suggest that sRNA-loaded EVs may contribute to cross-species RNAi activity. To put this hypothesis in perspective, we discuss contrasting evidence challenging the efficiency of EV-mediated sRNA transfer in light of deficient stability and stoichiometry.

Overview of gene silencing by sRNA

Small RNAs have three defining features: (1) they are short (21-31 nucleotides), (2) don't encode peptides and (3) can associate with RNAses of the Argonaute family (AGO) to modulate gene expression by targeting complementary transcripts. sRNA can impact gene expression through at least four distinct mechanisms: (1) AGO-dependent cleavage of target RNA, (2) destabilization of target mRNA through polyA tail shortening, (3) translational inhibition via polysomal protein interactions and (4) transcriptional silencing through chromatin modifications [24] [25] [26]. The field of RNAi truly emerged in the late 1990s, after the discovery that double-stranded (ds) RNA can specifically silence complementary transcripts in *C. elegans*[27]. An expanding repertoire of sRNA classes has since emerged, including microRNA (miRNA), endogenous small interfering RNA (endo-siRNA) and Piwi-interacting RNA (piRNA). In addition, infrastructural non-coding RNA, such as transfer RNA (tRNA) and vaultRNA (vtRNA) can be processed and recognized by the RNAi machinery to silence diverse mRNA targets[16, 28]. The protein machinery involved in sRNA nuclear export, maturation and RNAi is highly conserved and was likely present in the last common ancestor of eukaryotes[29], emphasizing the potential functionality of RNAi in cross-species communication. The biogenesis of sRNA in mammals has been clarified over the last decade[30-32] and reviewed in detail elsewhere[25, 26, 33, 34] **(Figure 2.1).**

Most miRNAs are transcribed from intergenic or intronic regions, with a few examples derived from exons of protein-coding genes. Hairpins found on primary (pri)-miRNA transcripts (≈1000 nt) bind to the nuclear microprocessor complex consisting of the RNAse III enzyme Drosha and the RBP DGCR8/Pasha. The microprocessor cleaves pri-miRNAs into precursor (pre)-miRNAs (≈70 nt), which undergo nuclear export via Exportin-5 [35]. Pre-miRNA and long dsRNA sequences are recognized by Dicer and subsequently cleaved ('diced') to generate sRNA duplexes (≈22 nt). Recent evidence suggests that structured dsRNA regions of tRNA and vtRNA can also be recognized and processed by Dicer, yielding small sequences similar to a mature miRNA [28]. Argonaute 2 (AGO2) is recruited to the complex by the Dicer-binding protein TRBP, enabling the transfer of the leading (or guide) RNA strand to the PAZ domain of AGO2. TRBP, PACT and C3PO are involved in leading strand selection and re-positioning [36, 37]. They also clear the remaining complementary single-stranded RNA copy, the passenger strand, or miRNA*, which is typically targeted for degradation[38]. The resulting minimal RNA-induced silencing complex (RISC) consists of AGO2 bound to the mature miRNA.

RNAi in crosstalk between plants, bacteria, fungi, insects and nematodes

Plants rely on RNAi for various endogenous processes, including defense against viral parasites. The genome of *Arabidopsis thaliana* encodes over 10 different AGO proteins, presumably reflecting the emergence of new functions [39]. In 1990, *Napoli et al.* reported an unexpected block in anthocyanin biosynthesis upon the introduction of a chimeric chalcone synthase construct in petunia [40]. The resulting crop exhibited white and patterned flowers presenting pale nonclonal sectors on a pigmented WT background. The mechanism involved was deemed unclear at the time and only became apparent after the realisation that dsRNA expression leads to gene silencing[27]. More recently, engineering of plant RNAi has been described as an effective and 'eco-friendly' method to modulate crop phenotypes in the aim of increasing productivity [41]. In particular, host gene silencing–hairpin RNAi (HGS-hpRNAi), wherein a transgenic sRNA-encoding hairpin is expressed in plants to impact pathogen resistance has emerged over the last 15 years[23]. HGS-hpRNAi is reportedly effective against diverse

pathogens, including bacteria (e.g. *Agrobacterium*), fungi (e.g. *Fusarium*), insects (e.g. *Helicoverpa*) and nematodes (e.g. *Meloidogyne*) (**Table 2.I**). However, the mechanisms through which plant-encoded dsRNA and/or sRNA molecules are transferred to symbiotic neighbours remain largely unclear and could involve EVs, RNPs and/or lipoproteins.

The soil bacterium *Agrobacterium tumefaciens* causes crown gall disease through disruptions of host's auxin and cytokinin biosynthesis, leading to the formation of tumour in various species of *Eudicotidae* flowering plants [42]. The pathogenesis of *Agrobacterium* crown gall disease is well characterized and involves the bacterial Tumor-inducing (Ti) plasmid [43]. Ti ssDNA is trafficked from the bacterium to the host plant via a conjugation pilus [43], which is a prevalent vehicle of nucleic acid exchange among bacteria. Whether the bacterial conjugation pilus can enable host-encoded sRNA populations to enter *Agrobacterium* remains unclear. Ti is integrated in the host genome by recombination and encodes opine synthesis oncogenes (*iaaM* and *ipt*). *Escobar et al.* reported resistance to crown gall tumorigenesis in transgenic *Arabidopsis* expressing self-complementary constructions designed to initiate HGS-hpRNAi against *iaaM* and *ipt* [44]. This finding was later confirmed in apple (*Malus pumila*) and walnut (*Juglans regia*) trees [23] [45].

Ascomycete pathogens of the *Fusarium* genus release mycotoxins and cause 'root rot' and *Fusarium* head blight pathologies, leading annually to severe loss in cereal crop productions. *Koch et al.* showed that *Arabidopsis thaliana* and *Hordeum vulgare* (barley) plants expressing dsRNA targeting the fungal gene *CYP51* were completely immune to *Fusarium graminearum*[46] (**Figure 2.2A**). Fungal growth was strongly restricted (≥99%) and only present in the vicinity of inoculation sites. The authors reported increased sporulation and altered morphology of *Fusarium* exposed to transgenic plants, consistent with compromised levels of cytochrome P450 lanosterol C-14α-demethylase, the enzyme encoded by *CYP51*. Similarly, *Ghag et al.* achieved effective resistance to *Fusarium oxysporum cubense* in transgenic bananas (*Musa paradisiaca*) expressing hairpins to target two vital fungal genes, *velvet* and *Fusarium transcription factor 1*[47]. Additional demonstrations of HGS-hpRNAi from plants to fungi have involved the mutualistic mycorrihzal genus *Glomus*. *Helber et al.* characterized the *Monosaccharide Transprter2* (*MST2*) gene in *Glomus* species and provided evidence of its

requirement for mycorrhiza formation through a HGS-hpRNAi loss-of-function model [48]. A similar approach was used by *Vega-Arreguin et al.* to demonstrate the role of the fungal gene *Avr3a1* in the hypersensitive response exhibited by diverse species of the genus *Nicotiana* to the pathogenic oomycete *Phytophthora capsici*[49].

Nematodes are appealing models to study RNAi and HGS-hpRNAi. In the nematode *C. elegans*, where RNAi was discovered [27], sRNA induces a systemic, amplified and heritable response[34]. Strong evidence indicates that dsRNA expressed in *E.coli* fed to *C. elegans* can transmit systemic and heritable silencing activity upon ingestion [50]. In *C. elegans*, the intestinal transmembrane protein SID-1 binds and imports dsRNA [51], while SID-2 is associated with cellular export of RNAi signals and required for systemic environmental RNAi [52]. Parasitic nematodes of the *Meloidogyne* genus are soilborne root pathogens that feed on diverse plants, notably potato (*Solanum tuberosum*) and soybean (*Glycine max*). *Meloidogyne* damages roots and compromises the plant's ability to absorb water and nutrients, affecting crop productivity [53]. *Dinh et al.* showed that *Arabidopsis* and potato plants expressing dsRNA constructs that target the nematode gene *16D10L* develop resistance to *Meloidogyne chitwoodi*. Interestingly, RNAi against *16D10L* was transferred to the progeny of worms feeding on transgenic roots, consistent with reports of heritable RNAi in nematodes[54]. In addition, *Ibrahim et al.* assessed the efficiency of HGS-hpRNAi at reducing galls formed by *Meloidogyne incognita* in soybean roots. The authors tested four potential targets in *M. incognita* and reported that HGS-hpRNAi against transcripts encoding Tyrosine Phosphatase (TP) and Mitochondrial Stress-70 Protein Precursor (MSP) were highly efficient at reducing galls in infected plants [55].

Like nematodes, several arthropods can internalize dietary dsRNA molecules [56]. Indeed, orthologs of the *sid-1 C. elegans* gene have been identified in several insects, including *Apis mellifera* (honeybee), *Bombyx mori* (silkworm) and *Tribolium castaneum* (red flour beetle) [57]. In light of these findings, *Baum et al.* provided a general assessment of HGS-hpRNAi usefulness for the control of coleopteran insects. *Baum et al.* showed that the western corn rootworm *Diabrotica virgifera virgifera* Leconte is sensitive to orally provided dsRNA, exhibiting dramatic suppression of 17 endogenous targets within 24 hours of ingestion [58]. The authors demonstrated the potential of oral dsRNA delivery for insect pest control by determining

the lethal dose of sequences targeting diverse protein-coding genes. Among the 290 dsRNA tested, 125 showed significant larval mortality. In a HGS-hpRNAi assay, maize plants expressing dsRNA targeting coleopteran *v-ATPase A* were protected from *Diabrotica* feeding damage. *Thakur et al.* and *Mutti et al.* provided additional evidence that expression of dsRNA targeting insect genes can improve crop resistance in various models [59, 60]. Multiple plants release secondary metabolites or phytochemicals that promote resistance to parasites. For example, cotton plants (*Gossypium* genus) synthesize gossypol, a toxic sesquiterpene compound that detracts most herbivores[61]. However, the cotton bollworm, *Helicoverpa armigera*, tolerates high concentrations of gossypol. *Mao et al.* showed that a cytochrome P450 gene, *CYP6AE14*, is induced by gossypol and required for insect tolerance to the compound[61]. When fed *Arabidposis* or *Nicotiana* plant material expressing a *dsRNA* construct raised against *CYP6AE14*, the sensitivity of *Helicoverpa* larvae to gossypol was markedly increased.

Recent evidence suggests that fungal pathogens can transfer RNAi signals to modulate the immunity of the plants they parasite. Indeed, *Weiberg et al.* reported that *Botyris cinerea*, the causative agent of gray mold disease, encodes sRNA populations derived from retrotransposons which can silence *Arabidopsis* and *Solanum* genes involved in immunity[62] (**Figure 2.2B**). In this study, *Weiberg et al*. generated sRNA sequencing libraries from *B. cinerea* mycelia, conidiospores and total biomass. In parallel, they profiled leaves from *B. cinerea*-infected *Arabidopsis* and tomato plants (*Solanum lycopersicum*). Interestingly, a total of 832 *B. cinerea*-encoded sRNAs were tracked and overrepresented in infected plant extracts, 52 of which mapped to a six different fungal long terminal repeat (LTR) retrotransposons. Among predicted *Arabidopsis* and *Solanum* targets, reporter assays confirmed silencing of *mitogen activated protein kinases* (*MPK1*, *MPK2*, *MAPKKK4*), *peroxiredoxin* (*PRXIIF*) and *cell-wall associated kinase* (*WAK*). Transgenic *Arabidopsis* plants ectopically expressing three *B. cinerea* sRNAs (*Bc-siRNA3.1*, *Bc-siRNA3.2*, *Bc-siRNA5*) displayed normal morphology but enhanced disease susceptibility upon pathogen challenge. A consistent phenotype was observed in *mpk1 mpk2* double mutants, suggesting an involvement of these factors in host immunity against *B. cinerea*. Immunoprecipitation of AGO1 in *B. cinerea*-infected *Arabidopsis* retrieved *Bc-siRNA3.1*, *Bc-siRNA3.2 and Bc-siRNA5*. *Arabidopsis* mutants of *Ago1* (*ago1-27*) showed reduced disease susceptibility to *B. cinerea*, whereas the Dicer-like mutant *dcl1-7* rather displayed enhanced

disease phenotype. By contrast, knock out of *B. cinerea dcl* genes depleted sRNA pools and reduced virulence upon *Arabidopsis* and *Solanum* inoculation. Together, these results show that fungal Dicer and plant Ago1 are involved in *Bc-siRNA3.1-*, *Bc-siRNA3.2- and Bc-siRNA5-* induced gene silencing in *Arabidopsis*. *Weiberg et al.* thus unravelled a mechanism whereby fungal sRNA populations hijack the plant host's RNAi machinery to subvert plant immunity and promote disease progression.

RNAi in crosstalk between intestinal cells and the gut microbiota

With over 100 trillion organisms representing 1000 species, the gut microbiota plays pivotal roles in human health and disease [63] [64]. Inflammatory bowel disease [65], diabetes [66], obesity [67] [68], diverse malignancies [69, 70] and neurological disorders [71] have been linked to disruptions in intestinal homeostasis. Several studies have reported an increase in susceptibility to parasites in mice bearing a conditional *Dicer* deletion in intestinal epithelial cells (*Dicer1*$^{\Delta gut}$), suggesting that RNAi is involved in mucosal immunity, possibly affecting communication with gut microorganisms.

In *Dicer1*$^{\Delta gut}$ mice, *Biton et al.* documented a strong depletion of mucus-secreting goblet cells required for intestinal homeostasis [72]. Silencing of the microprocessor component Drosha in the colorectal cancer line HT-29 led to impaired differentiation. Higher levels of miR-375 were observed in HT-29 clones expressing goblet cell markers than in enterocyte clones. Algorithmic approaches and luciferase assays identified the transcription factors *klf4* and its antagonist *klf5* as targets of miR-375. Since KLF4 is involved in goblet cell differentiation, this result suggests that downregulation of *klf5* by miR-375 could promote goblet cell commitment. Goblet-cell depletion is a hallmark of gut inflammation and immune mobilisation[73]. *Biton et al.* thus investigated the impact of miR-375 deletion on mucosal immunity and compared the phenotype of conditional knock out strains of *Dicer1* and *miR-375*. The levels of T_H2 cytokines (IL-4, IL-5, IL-13) were hampered in the gut of *Dicer1*$^{\Delta gut}$ and *miR-375*$^{\Delta gut}$ mice, whereas no change was observed in T_H1 cytokine abundance. Consistently, TSLP, a cytokine involved in

T_H2 activation, and RELMβ, a T_H2-controlled antiparasitic factor, were both down-regulated in *miR-375*$^{\Delta gut}$ and *Dicer1*$^{\Delta gut}$ samples, which failed to clear the parasitic nematode *Trichuris muris*.

In the colorectal line HT-29, IL-13 treatments increased the levels of both RELMβ and KLF4, suggesting a role in goblet cell differentiation. Previous work had shown that the PI(3)K pathway mediates intracellular responses to IL-13[74]. *Biton et al.* thus examined *Pten1*$^{\Delta gut}$ mice and found increased levels of miR-375, RELMβ protein and goblet cell markers in the gut of these animals. Since PTEN is a negative regulator of PI3K, this result supports an active role for the PI(3)K pathway upstream of miR-375 in goblet cell differentiation. When challenged with *T. muris or* IL-13, athymic nude mice, which fail to mount an effective T_H2 response[75], exhibited a phenotype reminiscent of *Dicer1*$^{\Delta gut}$ and *miR-375*$^{\Delta gut}$ mice, including compromised levels of RELMβ and TSLP. These findings provide strong evidence that miRNA contribute regulation in mucosal immunity and in coordinating the immune response to helminths.

Singh et al. identified 16 miRNA transcripts differentially expressed in caecum samples from germ-free and conventionally fed mice. Computational approaches pointed to over 2,000 putative mRNA targets, including factors involved in intestinal barrier function and immune regulation [76]. The authors found a strong overlap between their list of target mRNA and a previous survey of factors deregulated in the mucosa of *Dicer1*$^{\Delta gut}$ mice. This observation suggests that the gut microbiota modulates miRNA expression in the host, impacting intestinal barrier integrity. Along with several other studies [77] [78] [79], *Biton et al.* and *Singh et al.* provide evidence that RNAi modulates crosstalk between gut epithelial cells and the microorganisms that surround them. However, the prevalence and relevance of direct transfers of sRNA molecules from host to gut microbiota remain unclear. In a recent study, *Liu et al.* contributes to bridge that gap, showing that commensal gut bacteria uptake host miRNAs secreted in feces and that host miRNAs can exert RNAi in *E. coli* and *F. nucleatum* [80] (**Figure 2.3**). *Liu et al.* isolated and characterized miRNA populations contained within EVs in human and mice fecal samples. They identified over 180 miRNA in feces, which were differentially distributed in gut luminal content from the distal ileum and colon of mice. They show that intestinal epithelial cells, Paneth cells and goblet cells all contribute miRNA transcripts that account for fecal populations.

Liu et al. compared the gut microbiota in fecal matter from *Dicer1*$^{\Delta gut}$ and control (*Dicer1*$^{fl/fl}$) mice by sequencing the V4 region of rRNA 16S. Several differences were noted: representation of the bacterial phyla Firmicutes and Proteobacteria was notably increased in *Dicer1*$^{\Delta gut}$ samples. *Liu et al.* then submitted seven abundant bacterial RNA sequences from *E. coli* and *F. nucleatum* to a miRBase analysis [81] and identified numerous putative base-pairing events with host miRNA. Synthesized miRNA mimics of hsa-miR-1226-5p promoted the growth of *E. coli,* while hsa-miR-515-5p favoured growth of *F. nucleatum in vitro*. Mutated controls preventing base pairing of miRNA mimics to bacterial targets had no impact. Fluorescent Cy3-conjugated miRNA entered *E. coli* and *F. nucleatum*, co-localized with nucleic acids and increased the 16S rRNA/23S rRNA ratio in *F. nucleatum*. In *E. coli*, RNAseP levels were increased by miR-4747-3p while the bacterial transcripts *rutA* and *fucO* levels were decreased by miR-1224-p and miR-623, respectively. Target levels where not affected by mice miRNA mimics bearing mutations in predicted base-pairing nucleotides. Fecal gavage of *Dicer1*$^{\Delta gut}$ mice with WT samples led to a restoration of WT microbiota populations after seven days, as determined by 16S rRNA sequencing.

Next, *Liu et al.* investigated the phenotype of *Dicer1*$^{\Delta gut}$ mice and found evidence of reduced MHC-II levels in intestinal lymphoid tissue inducer cells. Diverse cytokines were also decreased in the ileum and colon, including LT-β, IFN-γ and TGF-β. The resistin-like molecules Relm-α and Relm-β, which are critical for maintenance of intestinal barrier integrity, were compromised in the ileum of *Dicer1*$^{\Delta gut}$ mice, along with Occludin-1, ZO-1 and Claudin-1, -2 and -5, echoing previous reports of *Dicer1*$^{\Delta gut}$ phenotypes [82]. Based on these findings, the authors suspected increased susceptibility to colitis in *Dicer1*$^{\Delta gut}$ mice, and tested the hypothesis by inducing the disease through oral administration of dextran sulphate sodium. As expected, *Dicer1*$^{\Delta gut}$ mice exhibited greater body weight loss, colon shortening and colon infiltration in response to dextran sulphate than WT mice. However, gavage of *Dicer1*$^{\Delta gut}$ mice with fecal matter from WT mice prior to dextran sulphate treatment alleviated the severity of these phenotypes, suggesting that fecal miRNA can attenuate the colonic alterations seen in *Dicer1*$^{\Delta gut}$ mice.

Together, these observations strongly suggest that host miRNA can be internalized, exert RNAi and mediate compositional changes in gut bacterial populations to promote intestinal homeostasis. Although *Liu et al.* does not directly demonstrate that host miRNA populations are transferred via EVs, the study shows that sequences enriched in fecal EVs are involved in cross-kingdom RNAi. Interestingly, in mice, EVs have been identified as a communication vehicle between intestinal epithelial cells and the immune system enabling MHC-II protein transfers [83].

Diverse subpopulations of secreted vesicles contain sRNA

In mammalian cells, the release of membranous vesicles upon exocytosis of vesicular endosomes was first reported in 1983 by *Harding et al.* [84]. Long dismissed as cellular debris, EVs have emerged over the last decade as key vehicles of biological signals, notably sRNA. Transcripts enriched in EVs include specific mRNA and full-length and fragmented non-coding transcripts, such as ribosomal (r)RNA, long non-coding (lnc)RNA, transfer (t)RNA, vault (vt)RNA, Y RNA, small nuclear (sn)RNA and small nucleolar (sno)RNA populations[14] [85] [86] [87] [88]. Viral transcripts have been identified in the EVs of cells infected with Epstein-Barr virus [89] [90]. Although specific miRNAs are overrepresented in EVs, diverse studies indicate that cumulative miRNA abundance is lower in EVs than in cells [91] [92].

The EV field have largely focussed on mammalian systems. In humans, EVs have notably been described as promising sources of biomarkers for diverse diseases [93]. EVs and EV-associated RNA populations have been identified and profiled by RNA-seq in multiple human biological fluids, including blood [94] [95], milk [96], semen [97], saliva [98], cerebral spinal fluid [99], urine [100] and ascitic fluids [101]. The release of exosome-like vesicles carrying sRNA populations have also been described in the nematode *Caenorhabditis elegans*, the arthropod *Drosophila melanogaster* [14] and the unicellular fungi *Cryptococcus neoformans*, *Paracoccidiodes brasiliensis*, *Canadida albicans* and *Saccharomyces cerevisae* [102]. Similarly, specific populations of sRNA have been defined in protozoans of the *Leishmania* [103] and *Trypanosoma* [104] genera. Outer membrane vesicles released by Gram-negative

bacteria, notably *Vibrio cholera*, have been shown to contain specific RNA populations and suggested to function as an RNA delivery system during infection. Release of MVE-associated exosomes in plants has been suggested 40 years ago [105] and is consistent with electron microscopy evidence but remains controversial[106].

'EV' is an umbrella term referring to diverse subpopulations of membrane-enclosed vesicles, often co-purified together in protocols that involve sequential ultracentrifugation of biological fluids [107]. Exosomes are small EVs (40-120 nm) that originate in endosomes and are released in the extracellular space upon exocytosis of multivesicular endosomes (MVEs). Vesicles shed by the plasma membrane through an actin-dependent abscission are typically larger (50-1000 nm) and have been called microvesicles, ectosomes or microparticles [108]. Apoptotic cells release small vesicles (50-500 nm) in addition to large apoptotic bodies containing organelles [109] [110] (50-5000 nm). Membrane-enclosed particles with retroviral-like composition and morphology (90-100 nm) have also been identified in cancer cell media [111] and in plasma samples of lymphoma patients [112]. Beyond EVs, sRNA can be stably shuttled in biological fluids in association with lipoproteins and RNP complexes, some containing AGO2 and nucleophosmin-1 (NPM1). At least two studies suggest that extracellular AGO2 and miRNA are more abundant in soluble complexes than within EVs [113] [114]. In addition, 'tunnelling nanotubes' are actin-rich protrusions that can bridge eukaryotic cells and may provide an alternative route for nucleic acid transfers [115]. First described in 2004 in cultures of rat pheochromocytoma cells [116], tunnelling nanotubes have since been shown to enable transmission of HIV particles between T cells and Jurkat cells [117].

It is technically challenging to separate subpopulations of EVs. *Kowal et al.* developed a tailored approach to separate EV subpopulations using continuous density gradients and immuno-isolation[118]. Mass spectrometry revealed distinct but partially overlapping protein profiles in human exosomes, microvesicles and apoptotic bodies. Other studies reported divergent nucleic acid imprints in EV subpopulations [119, 120]. *Crescitelli et al.* found that ribosomal RNA are more abundant in apoptotic bodies and in microvesicles than in exosomes, which reportedly contain more RNA than microvesicles. However, *Kanada et al.* showed that

microvesicles could target reporter molecules to recipient cells more efficiently than exosomes, which appeared ineffective at delivering nucleic acids [121].

Mechanisms of sRNA sorting to EVs

Our understanding of EVs in health and disease has expanded rapidly over the last years and has been frequently reviewed[122, 123]. In this section, we will focus on the mechanisms involved in sorting sRNA to mammalian EVs (**Figure 2.4**). Spatially resolved subcellular targeting of RNA molecules is often mediated by sequence or structure motifs found in the transcript. Called *cis*-acting elements, these sequences specifically interact with *trans*-acting factors, usually RBPs[124] [125]. At the subcellular level, mRNA localization is a prevalent process with key functional contributions, notably in embryogenesis[125-127] and synaptogenesis[128-130]. *Batagov et al.* [131] extended the rationale of subcellular localization and submitted a list of EV-targeted transcripts inferred from microarray datasets [132] to motif search algorithms. Although multiple alignments and position-specific scoring approaches failed to identify shared signatures among EV RNA, the authors found that EV-enriched transcripts display significantly shorter half-lives than cell-retained transcripts. By contrast, *Bolukbasi et al.* identified a 25 nt motif in the 3'UTR of mRNAs enriched in EVs from glioblastoma and melanoma cell lines[133]. Mutagenesis and reporter assays confirmed the functionality of the sequence in targeting mRNA to EVs. Interestingly, the authors found that the motif encompasses the seed region for miR-1289 along with a core CUGCC sequence. Furthermore, altering the levels of miR-1289 was sufficient to modulate artificial target levels in EVs, suggesting a role for miRNA in sorting complementary mRNA to EVs. It should be noted that the studies discussed above used microarray data focussed on mRNA and long non-coding RNA. Other sources have since emphasized the enrichment of shorter sequences in EVs, including specific miRNA [134]. *Hung and Leonard* showed that RNA length alone strongly modulates targeting efficiency [135]. They developed a tailored approach based on the MS2-GFP system and confirmed that long sequences (1.5 kb) are poorly loaded in EVs.

The EV-associated miRNA repertoires exhibit considerable cell-type specificity and dramatic alterations in these populations have been reported upon cell fate commitment or changes in environmental status and stimuli. Hypoxic conditions have been shown to increase exosomal release and modulate associated miRNA in breast cancer cell lines, notably leading to a strong increase in miR-210 targeting [136]. Interleukin treatment also promotes activation-dependent changes in miRNA populations released by macrophages through EVs[137]. The miRNA repertoire of colon cancer cell EVs is profoundly affected by mutations in the transcription factor *KRAS*. *Cha et al.* profiled the transcriptome of EVs released by cell lines differing only in *KRAS* status and notably showed that levels of the pro-metastatic miR-100 are decreased in mutant *KRAS* EVs, whereas miR-10b abundance is increased in these samples [138].

Unlike EVs released by non-malignant cells, breast cancer exosomes contain the proteins Dicer, TRBP and AGO2, which can reportedly perform cell-independent miRNA processing within EVs[15]. Several studies have identified populations of pre-miRNA, mature miRNA, miRNA* strands, hairpins loops and pre-miRNA cleavage products in EVs[15] [139]. Over the course of 48h, *Melo et al.* identified a sharp decrease in pre-miRNA abundance in previously purified exosomes, which coincided with a marked increase in corresponding mature miRNA levels. In addition, *Melo et al.* identified a role of the sialoglycoprotein CD43 in recruiting Dicer to cancer cell exosomes. Indeed, co-immunoprecipiation revealed an interaction between the two proteins and silencing of CD43 severely compromised the recruitment of Dicer to exosomes. Moreover, inhibiting Dicer activity in breast cancer exosomes significantly impaired growth in recipient malignant cells, providing evidence of its involvement in tumour progression.

Villarroya-Beltri et al. provided robust evidence of a sequence-specific mechanism involving the RBP hnRNPA2B1 in miRNA sorting to EVs [140]. The authors investigated activation-dependent changes in the miRNA repertoire of lymphoblasts and observed divergent trends in cells and EVs, consistent with active, sequence-specific loading. Sequence alignments and targeted mutagenesis revealed a role of the GGAG motif in miRNA targeting to EVs. RNA pull-down experiments coupled to mass spectrometry identified three hnRNP factors specifically

bound to EV-targeted miRNA. The author focused on hnRNPA2B1 and confirmed specific miRNA association by immunoprecipitation coupled to qPCR. They also showed that hnRNPA2B1 targeting to EVs is regulated by SUMO conjugation. Annexin A2 is a Ca^{2+}-binding protein that contributes to link membrane-associated complexes to cytoskeletal components [141]. Annexin A2 exhibits sequence-specific RNA-binding activity and is involved in *c-myc* post-transcriptional regulation[142]. Proteomic studies have revealed that Annexin A2 is among the most abundant proteins in EVs[143]. *Hagiwara et al.* provided evidence that Annexin A2 can bind miRNA in the presence of Ca^{2+} in diverse cancer cell lines[143]. The authors reported a global decrease in miRNA loading to cancer cell EVs upon Annexin A2 silencing.

Lipidomics studies based on mass spectrometry and nuclear magnetic resonance have unravelled profound differences in the composition of EV and plasma membranes [144]. Lipid rafts are dynamic, detergent-resistant membrane microdomains enriched in sphingomyelin and depleted in in phosphatidylcholine [145]. Lipid rafts are overrepresented in EVs and have been involved in sorting proteins and RNAs to exosomes [145] [146]. The sphingolipid ceramide is enriched in lipid rafts and implicated in membrane sorting during exosome budding to the lumen of MVEs [147] [148]. Neutral sphingomyelinase (nSMase) is the rate-limiting enzyme in ceramide biogenesis and its inhibitor GW4869 has been used by several groups to restrict exosome release *in vitro* [148, 149] [150]. The 'ceramide pathway' has emerged as an important route for miRNA loading to exosomes [150]. *Kosaka et al.* provided evidence that metastatic cancer cells exert microenvironment remodelling of endothelial cells through exosome-associated miR-210 [151]. This phenotype was abrogated by silencing nSMase in breast cancer cell lines, consistent with a role of the ceramide pathway in exosomal miRNA sorting.

Koppers-Lalic et al. reported that 3' end uridylated miRNA isoforms are enriched in B cell exosomes, whereas 3' end adenylated isoforms are poorly targeted and relatively enriched in cells [92]. The authors extended their finding in EVs purified from human urine samples and concluded that nontemplated terminal uridylation promotes miRNA sorting to EVs. Previous studies have shown that terminal adenylation increases transcript stability while uridylation has a destabilizing effect [152], bridging the findings of *Koppers-Lalic et al.* to *Batagov et al.*'s conclusions that RNAs with short half-lives are enriched in EVs. *Squadrito et al.* investigated co-

dependencies in miRNA and target mRNA levels in bone marrow-derived macrophages and corresponding EVs [137]. The authors used IL-4 and genetic perturbations to alter the expression of miRNA and their target mRNA. Their observations suggest that the levels of endogenous target modulate miRNA sorting to EVs, likely through a relocation of RISC from P-bodies to MVEs. Indeed, several studies suggest that intracellular targeting of AGO2 complexes to endolysosomal compartments and exosomes reflects the dynamics of membrane-less organelles such as P-bodies and GW182-bodies [153].

Recruitment of AGO2-bound miRNA to MVEs

Gibbings et al. investigated RISC subcellular distribution using cell fractionation, immunofluorescence and qPCR [154]. They showed that 'GW-bodies' containing AGO2 and its cofactor GW182 are distinct from canonical P-bodies and selectively congregate with MVEs. Immunogold labeling of monocoyte-derived exosomes revealed enrichments for GW182 but not DCP1A, a canonical P-body marker. Diverse miRNA and their target mRNA were enriched in the vicinity of GW182-positive MVEs. To understand how GW-182 bodies are recruited to MVEs, the authors silenced components of the endosomal sorting complexes required for transport (ESCRT), a highly conserved multisubunit machinery that localizes to MVEs and performs bending and scission of the membrane involved in protein and exosome release [155]. They showed that depleting ESCRT components severely impairs miRNA silencing activity in the cell by monitoring let-7-a and miR-206 repression through reporter assays. *Gibbings et al*. thus established two principles: (1) miRNA-loaded RISCs congregate at the site of exosome biogenesis and (2) ESCRT components regulate both exosome biogenesis and RNAi.

Independently of its involvement in exosome secretion, the ESCRT-II complex exhibits sequence-specific RNA-binding activity in metazoans. *Irion et al.* [156] focused on *bicoid* mRNA localization during *Drosophila* development. The study revealed that mutations in all three subunits of the ESCRT-II complex abolish the localization of *bicoid* mRNA at the anterior pole of the egg. The authors demonstrated a direct interaction between the N-terminal GLUE domain of VPS36 and stem-loop V in *bicoid* 3'UTR using UV-crosslinking and a yeast three-hybrid assay. They extended their finding in *Xenopus*, establishing conservation of the

interaction in Vertebrates. ESCRT-II is thus at the crossroads of exosome biogenesis, RNAi and subcellular RNA localization, prompting speculations that the complex may contribute to sRNA sorting to exosomes. *Kosaka et al*. tested the hypothesis and depleted an ESCRT component, Alix, in HEK293 cells. In agreement with *Gibbings et al.*, luciferase assays showed a reduction in intracellular silencing activity by miR-146. However, the amount of miR-146 in EVs was not altered by Alix depletion. EVs from Alix-depleted HEK293 cells contained miR-146 and silenced a reporter gene in recipient cells as efficiently as EVs released by untreated cells. Further efforts are required to elucidate the involvement of ESCRT components in sRNA sorting to EVs.

Recent work by *McKenzie et al.* provides an alternative mechanism of AGO2-miRNA relocation from P-bodies to MVEs [157]. Echoing *Cha et al.*'s identification of KRAS signalling as a modulator of miRNA sorting to EVs, *McKenzie et al.* showed that KRAS-dependent activation of the MEK-ERK pathway inhibits AGO2 sorting to EVs. This work revisits a previously identified KRAS-dependent phosphorylation of serine residue 387 on AGO2 and demonstrates its implication in excluding AGO2-miRNA complexes from MVE association and exosome targeting. AGO2 targeting to exosomes thus reflects KRAS-MEK-ERK signalling status, which is impacted by environmental cues. These reconciliatory findings provide a possible explanation for discrepancies in previous reports regarding AGO2 levels in exosomes[15, 154].

Contrasting and 'EV sceptic' perspectives

We have reviewed examples of RNAi activity transfers across diverse species spanning the eukaryotic and prokaryotic domains of life. We then envisioned possible vehicles of sRNA transfer, including EVs, lipoproteins, soluble RNPs and tunnelling nanotubes. We emphasized emerging mechanisms of sRNA sorting to EVs in mammalian system, suggesting that these vesicles may contribute to cross-species sRNA transfers. EV association strongly enhances the stability of RNA molecules in the extracellular environment. In addition, examples of long-range transfers of biomolecules through EVs have been reported in diverse systems. In *C. elegans*, EVs

transferred between worms contribute to the specification of male sexual behaviour[22], while functional EV-associated transcripts encoding a Cre recombinase are reportedly shuttled across distant tumours in mice [158]. Transfers of EVs and delivery of molecular cargo from human to mouse cultured cells and from the protozoan pathogen *Trypanosoma cruzi* to human erythrocytes have been documented, suggesting that EVs can indeed serve as widespread mediators of interspecies RNA transfers [13] [159, 160].

Numerous studies thus support the functionality of EV-associated sRNA populations in intercellular communication [15, 89, 161]. However, contrasting reports resulting from careful quantitative assessments argue that EVs are poor vehicles for RNA transfers due to degradation upon recipient cell entry and/or insufficient cargo abundance. *Kanada et al.* examined the fate of nucleic acids contained in HEK293FT small exosome-like EVs and larger microvesicle-like EVs upon recipient cell entry[121]. They found that exosome-like EVs fail to transfer nucleic acids to murine 4TI recipient cells. Microvesicle-like EVs delivered reporter RNA, which was successfully amplified using a nested PCR approach 24h after delivery. However, full-length and fragmented reporter RNA was undetectable 48h after transfer assays, likely due to degradation in acidic lysosomal compartments. Plasmid-encoded Cre recombinase was efficiently loaded in microvesicle-like EVs as plasmidic (p)DNA, RNA and protein. Recombinase activity was stably transmitted to recipient cells, but exclusively through pDNA. These findings suggest that pDNA transfers may have confounded the conclusions of several studies using reporter constructions to investigate RNA transfers.

Chevillet et al. purified EVs from five human biological fluids and cell line media and used quantitative approaches to determine miRNA stoichiometric occupancy in these samples[91]. Regardless of the source, they found that EV samples contain low counts of individual miRNA, amounting at most to an average of (0.00825 ± 0.02) miRNA molecules per EV. While this result suggests that EVs are poor miRNA transfer vehicles *in vivo*, *Chevillet et al.* discussed diverse stoichiometric models to reconcile their assessments with reports of functional miRNA delivery. Indeed, population analyses are unable to determine the distributions of miRNA molecules in individual EVs. A low occupancy/high miRNA concentration model, wherein most EVs contain no miRNA molecule but a single EV carries several copies could be

compatible with reports of functional transfers. Similarly, the kinetics of EV uptake could impact functionality in recipient cells, provided they internalize EVs at a high frequency. In the context of interspecies communication, it should be noted that organisms able to amplify systemic RNAi responses such as plants and nematodes may exhibit enhanced sensitivity to signals conveyed by a low number of initial miRNA copies.

Contrasting and sceptical reports are highly valuable to the EV field and should inform attentive methodological choices for future experiments. In transfer assays, reporter genes expressed from chromosomal insertions should be favoured rather than plasmid-based approaches, effectively ruling out pDNA transfers. DNAse treatments should be used when specifically investigating the roles of EV-associated RNA. Imaging EV transfers through time-lapse analysis of high-resolution microscopy data could illuminate the kinetics of internalization. In addition, the fate of transferred biomolecules should be considered over the course of several days, documenting the association of transferred material with subcellular compartments of recipient cells that may impact stability. Dutiful application of these principles will reveal whether membrane vesicles can indeed spread silencing instructions across phylogenetic boundaries.

Figure Legends

Figure 2.1. Overview of sRNA biogenesis in mammals

Pri-miRNA is cleaved into pre-miRNA by the microprocessor complex, consisting of two nuclear proteins, Drosha and its cofactor DGCR8. Pre-miRNA is exported to the cytoplasm through Exportin-5 (Exp-5), readily bound and processed into short dsRNA sequences of approximately 22 nt by the dsRBP Dicer and its associated factors TRBP and PACT. Mounting evidence suggests that structured ncRNA encompassing stretches of paired nucleotides such as tRNAs can be recognized and processed as Dicer substrates. Dicer recruits AGO2 and its cleavage yields two single-stranded RNA sequences, called the leading strand and the guide strand (or miRNA*). The leading strand is actively repositioned in the complex and loaded onto AGO2 to form a RISC, which can exert RNA silencing.

Figure 2.2. Host-induced hairpin RNA-mediated silencing confers resistance to the fungal pathogen *Fusarium* and *Botrytis cinerea* sRNA populations hijack *Arabidopsis* RNAi pathways to suppress plant immunity.

(A) Host-induced hairpin RNA-mediated silencing enables plant to resist to the fungal pathogen Fusarium. In Arabidopsis, expression of a dsRNA construct complementary to fungal CYP51 transcripts can immunize transgenic plants to the pathogenic ascomycete Fusarium graminearum by inhibiting fungal growth [46]. The vehicles through which transgenic dsRNA and/or sRNA is transferred are unknown and possibly include plant EVs, secreted RNPs and/or lipoproteins. (B) Botrytis cinerea sRNA populations hijack Arabidopsis RNAi pathways to suppress plant immunity. Populations of sRNA derived from a B. cinerea retrotransposon are shuttled to infected Arabidopsis and Solanum lycopersicum. In plants, fungal sRNA are loaded onto AGO1 and direct the silencing of diverse proteins, including Mitogen-activated kinases, which impact the host's immune response[62]. The vehicles through which dsRNA and/or sRNA are transferred from fungus to plant are unknown and possibly include fungal EVs, secreted RNPs and/or lipoproteins.

Figure 2.3. Host miRNA targets microbiota gene expression

Gut epithelial cells release miRNAs that can be recovered in murine and human fecal matter. Fecal miRNA populations are likely stabilized through EVs and possibly through lipoproteins or RNPs containing AGO2. Host miRNA enters *E. coli* and *F. nucleatum* where it co-localizes with bacterial nucleic acids and impacts bacterial growth by interacting with nucleic acids [80].

Figure 2.4. Mechanisms of sRNA loading to EVs

Schematic view of a mammalian cell releasing sRNA through lipoproteins and AGO2 RNPs (left), exosomes (center) and membrane-shed microvesicles (right). Properties broadly associated with RNA targeting to EVs are listed on top (white font). Mechanisms, lipid structures and RBPs involved in sorting RNA molecules to exosome and microvesicles are portrayed.

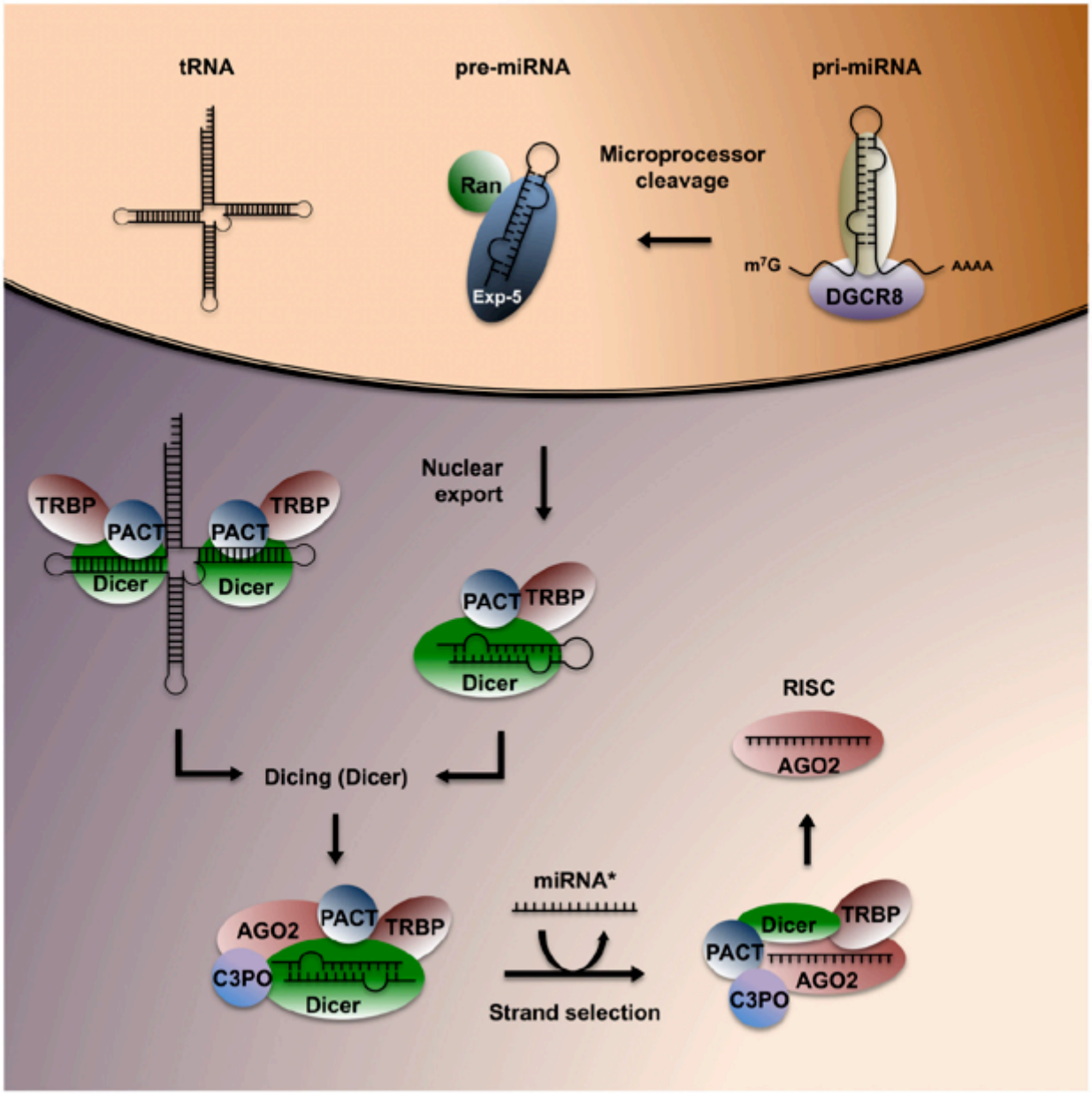


Figure 2. 1. Overview of sRNA biogenesis in mammals.

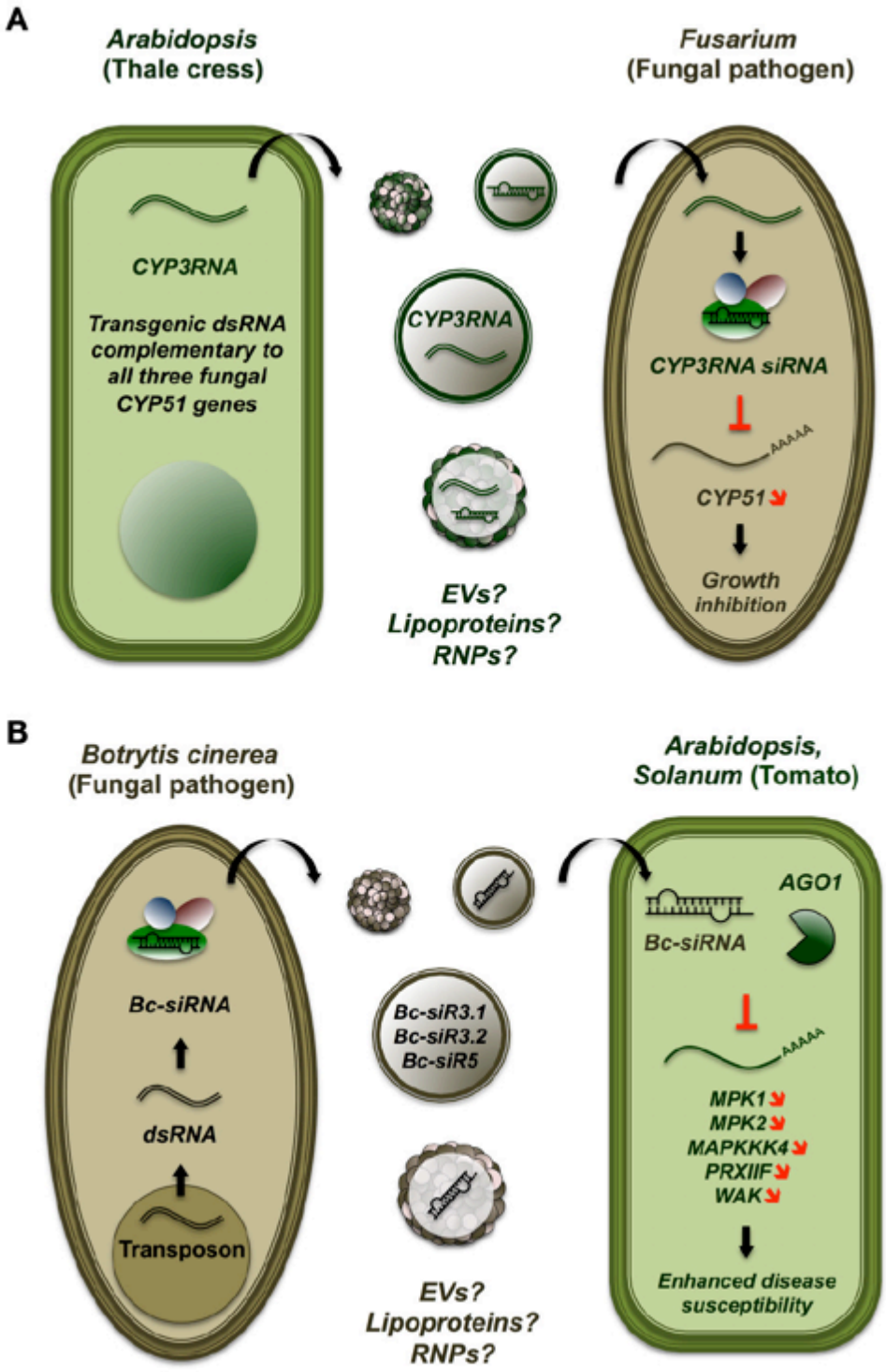


Figure 2. 2. Host-induced hairpin RNA-mediated silencing confers resistance to the fungal pathogen *Fusarium*

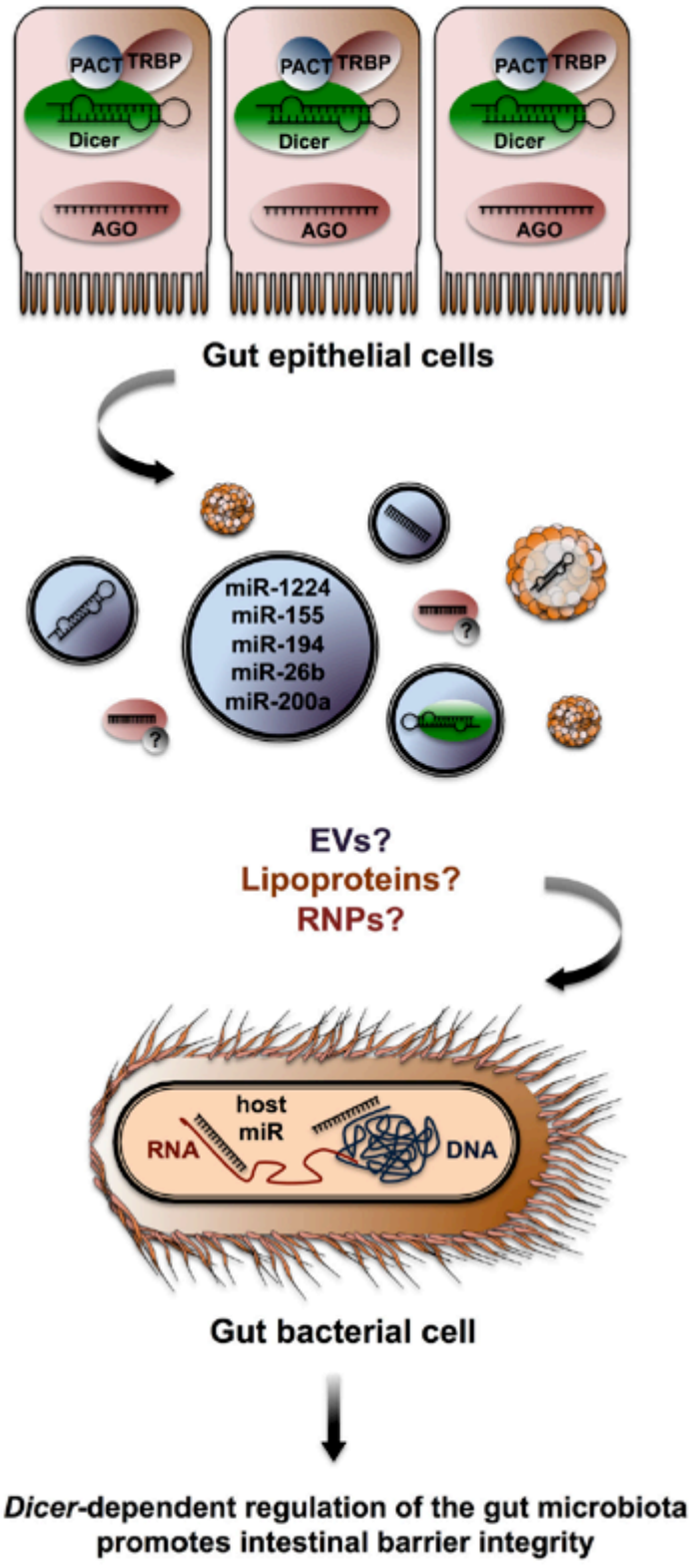


Figure 2. 3. Host miRNA targets microbiota gene expression

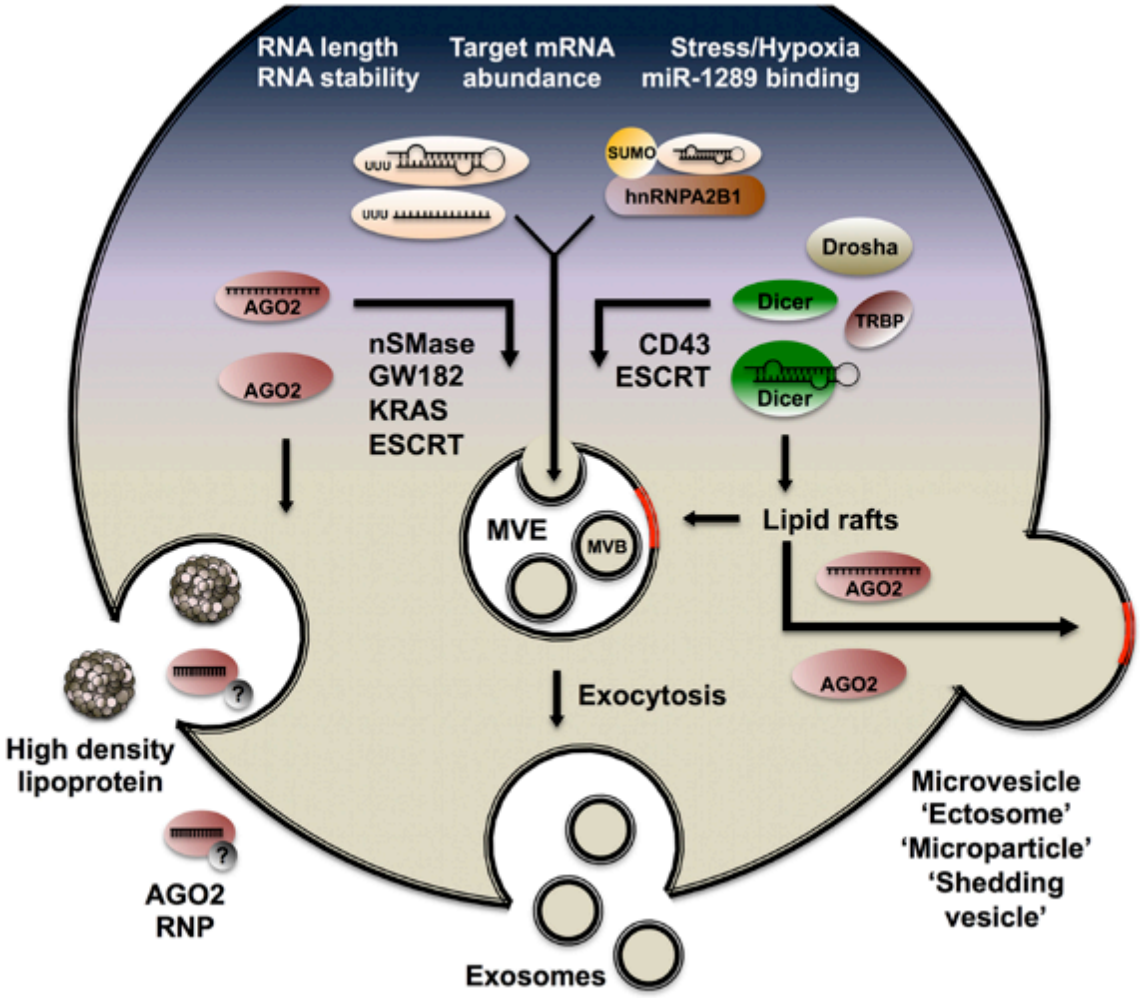


Figure 2. 4. Mechanisms of sRNA loading to EVs

Table 2. I. Evidence of RNAi activity transfers from plants to bacteria, fungi and metazoans

Donor plant	Recipient pathogen	Targets	Outcome	Ref
Arabidopsis thaliana (Thale cress)	*Agrobacterium tumefaciens* (Bacteria)	*iaaM*, *ipt*	Resistance to crown gall disease	[44]
Malus genus (Apple)	*Agrobacterium tumefaciens* (Bacteria)	*iaaM*, *ipt, iaaH*	Resistance to crown gall disease	[23]
*Juglans regia (*Walnut*)*	*Agrobacterium tumefaciens* (Bacteria)	*iaaM*, *ipt*	Resistance to crown gall disease	[45]
Hordeum vulgare (Barley)	*Fusarium graminearum* (Fungus)	*CYP51*	Inhibition of fungal growth	[46]
Musa paradisiaca (Banana)	*Fusarium oxysporum* (Fungus)	*Velvet, FTG1*	Effective resistance to *F. oxysporum*	[47]
Nicotiana genus (Tobacco)	*Phytophtora capsici* (Fungus)	*PcAvr3a1*	Infection of resistant tobacco	[49]
Medicago truncatula (Barrelclover)	*Glomus intraradices* (Fungus)	*MST2*	Impaired mycorrhiza formation	[48]
Malus genus (Apple)	*Venturia inequalis* (Fungus)	*THN*	Light brown phenotype	[162]
Solanum tuberosum (Potato)	*Meloidogyne* sp (Nematode)	*Mc16D10L*	Egg count reduction	[54]
Glycine max (Soybean)	*Meloidogyne incognita* (Nematode)	*TP, MSP*	Reduction of *incognita* gall count	[55]
Zea mays (Maize)	*Diabrotica virgifera* (Insect)	*v-ATPaseA/E β-tubulin*	Increased larval mortality	[58]
Nicotiana benthamiana	*Myzus persicae* (Insect)	*Rack1 MpC002*	Reduced aphid fecundity	[163]
Arabidopsis thaliana (Thale cress)	*Helicoverpa armigera* (Insect)	*CYP6AE14*	Larval growth retardation	[61]
Nicotiana rustica (Tobacco)	*Helicoverpa armigera* (Insect)	*EcR*	Improvement of pest resistance	[164]
Nicotiana rustica (Tobacco)	*Bemisia tabaci* (Insect)	*v-ATPseA*	Improvement of pest resistance	[59]
Medicago sativa (Lucerne)	*Acyrthosiphon pisum* (Insect)	*COO2*	Lethality of *A. pisum*	[60]
Oryza sativa (Rice)	*Nilaparvata lugens* (Insect)	*NIHT1, Nlcar Nltry*	No phenotype reported	[165]

Acknowledgments

FAL is supported by a scholarship from the Canadian Institute of Health Research (CIHR).

References

[1] Mc CB. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A. 1950;36:344-55.

[2] Coffey AJ, Kokocinski F, Calafato MS, Scott CE, Palta P, Drury E, et al. The GENCODE exome: sequencing the complete human exome. Eur J Hum Genet. 2011;19:827-31.

[3] Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012;22:1775-89.

[4] Jalali S, Gandhi S, Scaria V. Navigating the dynamic landscape of long noncoding RNA and protein-coding gene annotations in GENCODE. Hum Genomics. 2016;10:35.

[5] Clark MB, Amaral PP, Schlesinger FJ, Dinger ME, Taft RJ, Rinn JL, et al. The reality of pervasive transcription. PLoS Biol. 2011;9:e1000625; discussion e1102.

[6] Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ. Argonaute2, a link between genetic and biochemical analyses of RNAi. Science. 2001;293:1146-50.

[7] Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005;120:15-20.

[8] Knip M, Constantin ME, Thordal-Christensen H. Trans-kingdom cross-talk: small RNAs on the move. PLoS Genet. 2014;10:e1004602.

[9] Weiberg A, Bellinger M, Jin HL. Conversations between kingdoms: small RNAs. Curr Opin Biotech. 2015;32:207-15.

[10] Tsui NB, Ng EK, Lo YM. Stability of endogenous and added RNA in blood specimens, serum, and plasma. Clin Chem. 2002;48:1647-53.

[11] Hunter MP, Ismail N, Zhang X, Aguda BD, Lee EJ, Yu L, et al. Detection of microRNA expression in human peripheral blood microvesicles. PLoS One. 2008;3:e3694.

[12] Lasser C, Alikhani VS, Ekstrom K, Eldh M, Paredes PT, Bossios A, et al. Human saliva, plasma and breast milk exosomes contain RNA: uptake by macrophages. J Transl Med. 2011;9:9.

[13] Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, Lotvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat Cell Biol. 2007;9:654-9.

[14] Lefebvre FA, Benoit Bouvrette LP, Perras L, Blanchet-Cohen A, Garnier D, Rak J, et al. Comparative transcriptomic analysis of human and Drosophila extracellular vesicles. Sci Rep. 2016;6:27680.

[15] Melo SA, Sugimoto H, O'Connell JT, Kato N, Villanueva A, Vidal A, et al. Cancer exosomes perform cell-independent microRNA biogenesis and promote tumorigenesis. Cancer Cell. 2014;26:707-21.

[16] Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. Science. 2016;351:391-6.

[17] da Silveira JC, de Andrade GM, Nogueira MF, Meirelles FV, Perecin F. Involvement of miRNAs and Cell-Secreted Vesicles in Mammalian Ovarian Antral Follicle Development. Reprod Sci. 2015;22:1474-83.

[18] Dror S, Sander L, Schwartz H, Sheinboim D, Barzilai A, Dishon Y, et al. Melanoma miRNA trafficking controls tumour primary niche formation. Nat Cell Biol. 2016;18:1006-17.

[19] Montecalvo A, Larregina AT, Shufesky WJ, Stolz DB, Sullivan ML, Karlsson JM, et al. Mechanism of transfer of functional microRNAs between mouse dendritic cells via exosomes. Blood. 2012;119:756-66.

[20] Hergenreider E, Heydt S, Treguer K, Boettger T, Horrevoets AJ, Zeiher AM, et al. Atheroprotective communication between endothelial cells and smooth muscle cells through miRNAs. Nat Cell Biol. 2012;14:249-56.

[21] Mizrak A, Bolukbasi MF, Ozdener GB, Brenner GJ, Madlener S, Erkan EP, et al. Genetically engineered microvesicles carrying suicide mRNA/protein inhibit schwannoma tumor growth. Mol Ther. 2013;21:101-8.

[22] Wang J, Silva M, Haas LA, Morsci NS, Nguyen KC, Hall DH, et al. C. elegans Ciliated Sensory Neurons Release Extracellular Vesicles that Function in Animal Communication. Curr Biol. 2014;24:519-25.

[23] Viss WJ, Pitrak J, Humann J, Cook M, Driver J, Ream W. Crown-gall-resistant transgenic apple trees that silence Agrobacterium tumefaciens oncogenes. Mol Breeding. 2003;12:283-95.

[24] Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science. 2002;297:1833-7.

[25] Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet. 2009;10:94-108.

[26] Rissland OS, Lai EC. RNA silencing in Monterey. Development. 2011;138:3093-102.

[27] Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature. 1998;391:806-11.

[28] Persson H, Kvist A, Vallon-Christersson J, Medstrand P, Borg A, Rovira C. The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. Nat Cell Biol. 2009;11:1268-71.

[29] Shabalina SA, Koonin EV. Origins and evolution of eukaryotic RNA interference. Trends Ecol Evol. 2008;23:578-87.

[30] Forstemann K, Tomari Y, Du T, Vagin VV, Denli AM, Bratu DP, et al. Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. PLoS Biol. 2005;3:e236.

[31] Yeom KH, Lee Y, Han J, Suh MR, Kim VN. Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. Nucleic Acids Res. 2006;34:4622-9.

[32] Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, et al. Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell. 2004;117:69-81.

[33] Mattick JS, Makunin IV. Small regulatory RNAs in mammals. Hum Mol Genet. 2005;14 Spec No 1:R121-32.

[34] Collins RE, Cheng X. Structural and biochemical advances in mammalian RNAi. J Cell Biochem. 2006;99:1251-66.

[35] Yi R, Qin Y, Macara IG, Cullen BR. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. 2003;17:3011-6.

[36] Noland CL, Doudna JA. Multiple sensors ensure guide strand selection in human RNAi pathways. RNA. 2013;19:639-48.

[37] Noland CL, Ma E, Doudna JA. siRNA repositioning for guide strand selection by human Dicer complexes. Mol Cell. 2011;43:110-21.

[38] Liu Y, Ye X, Jiang F, Liang C, Chen D, Peng J, et al. C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. Science. 2009;325:750-3.

[39] Zhang H, Xia R, Meyers BC, Walbot V. Evolution, functions, and mysteries of plant ARGONAUTE proteins. Curr Opin Plant Biol. 2015;27:84-90.
[40] Napoli C, Lemieux C, Jorgensen R. Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in Trans. Plant Cell. 1990;2:279-89.
[41] Younis A, Siddique MI, Kim CK, Lim KB. RNA Interference (RNAi) Induced Gene Silencing: A Promising Approach of Hi-Tech Plant Breeding. Int J Biol Sci. 2014;10:1150-8.
[42] Hoekema A, Hirsch PR, Hooykaas PJJ, Schilperoort RA. A Binary Plant Vector Strategy Based on Separation of Vir-Region and T-Region of the Agrobacterium-Tumefaciens Ti-Plasmid. Nature. 1983;303:179-80.
[43] McCullen CA, Binns AN. Agrobacterium tumefaciens and plant cell interactions and activities required for interkingdom macromolecular transfer. Annu Rev Cell Dev Bi. 2006;22:101-27.
[44] Escobar MA, Civerolo EL, Summerfelt KR, Dandekar AM. RNAi-mediated oncogene silencing confers resistance to crown gall tumorigenesis. P Natl Acad Sci USA. 2001;98:13437-42.
[45] Escobar MA, Leslie CA, McGranahan GH, Dandekar AM. Silencing crown gall disease in walnut (Juglans regia L.). Plant Sci. 2002;163:591-7.
[46] Koch A, Kumar N, Weber L, Keller H, Imani J, Kogel KH. Host-induced gene silencing of cytochrome P450 lanosterol C14alpha-demethylase-encoding genes confers strong resistance to Fusarium species. Proc Natl Acad Sci U S A. 2013;110:19324-9.
[47] Ghag SB, Shekhawat UKS, Ganapathi TR. Host-induced post-transcriptional hairpin RNA-mediated gene silencing of vital fungal genes confers efficient resistance against Fusarium wilt in banana. Plant Biotechnol J. 2014;12:541-53.
[48] Helber N, Wippel K, Sauer N, Schaarschmidt S, Hause B, Requena N. A Versatile Monosaccharide Transporter That Operates in the Arbuscular Mycorrhizal Fungus Glomus sp Is Crucial for the Symbiotic Relationship with Plants. Plant Cell. 2011;23:3812-23.
[49] Vega-Arreguin JC, Jalloh A, Bos JI, Moffett P. Recognition of an Avr3a Homologue Plays a Major Role in Mediating Nonhost Resistance to Phytophthora capsici in Nicotiana Species. Mol Plant Microbe In. 2014;27:770-80.

[50] Liu H, Wang X, Wang HD, Wu J, Ren J, Meng L, et al. Escherichia coli noncoding RNAs can affect gene expression and physiology of Caenorhabditis elegans. Nat Commun. 2012;3:1073.
[51] Jose AM, Smith JJ, Hunter CP. Export of RNA silencing from C-elegans tissues does not require the RNA channel SID-1. P Natl Acad Sci USA. 2009;106:2283-8.
[52] Winston WM, Sutherlin M, Wright AJ, Feinberg EH, Hunter CP. Caenorhabditis elegans SID-2 is required for environmental RNA interference. P Natl Acad Sci USA. 2007;104:10565-70.
[53] Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EGJ, Deleury E, et al. Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. Nature Biotechnology. 2008;26:909-15.
[54] Dinh PTY, Brown CR, Elling AA. RNA Interference of Effector Gene Mc16D10L Confers Resistance Against Meloidogyne chitwoodi in Arabidopsis and Potato. Phytopathology. 2014;104:1098-106.
[55] Ibrahim HMM, Alkharouf NW, Meyer SLF, Aly MAM, El-Din AEYG, Hussein EHA, et al. Post-transcriptional gene silencing of root-knot nematode in transformed soybean roots. Experimental Parasitology. 2011;127:90-9.
[56] Khila A, Grbic M. Gene silencing in the spider mite Tetranychus urticae: dsRNA and siRNA parental silencing of the Distal-less gene. Dev Genes Evol. 2007;217:241-51.
[57] Gramates LS. Insights into social insects from the genome of the honeybee Apis mellifera (vol 443, pg 931, 2006). Nature. 2006;444:512-.
[58] Baum JA, Bogaert T, Clinton W, Heck GR, Feldmann P, Ilagan O, et al. Control of coleopteran insect pests through RNA interference. Nature Biotechnology. 2007;25:1322-6.
[59] Thakur N, Upadhyay SK, Verma PC, Chandrashekar K, Tuli R, Singh PK. Enhanced Whitefly Resistance in Transgenic Tobacco Plants Expressing Double Stranded RNA of v-ATPase A Gene. Plos One. 2014;9.
[60] Mutti NS, Park Y, Reese JC, Reeck GR. RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, Acyrthosiphon pisum. J Insect Sci. 2006;6.
[61] Mao YB, Cai WJ, Wang JW, Hong GJ, Tao XY, Wang LJ, et al. Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. Nature Biotechnology. 2007;25:1307-13.

[62] Weiberg A, Wang M, Lin FM, Zhao H, Zhang Z, Kaloshian I, et al. Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. Science. 2013;342:118-23.
[63] Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, et al. The Long-Term Stability of the Human Gut Microbiota. Science. 2013;341:44-+.
[64] Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell. 2006;124:837-48.
[65] Halfvarson J, Dicksved J, Rosenquist M, Jarnerot G, Tysk C, Engstrand L, et al. Molecular fingerprinting of the gut microbiota of twins reveals differences according to Crohn's disease. Gastroenterology. 2008;134:A650-A.
[66] Patterson E, Ryan PM, Cryan JF, Dinan TG, Ross RP, Fitzgerald GF, et al. Gut microbiota, obesity and diabetes. Postgrad Med J. 2016;92:286-300.
[67] Tilg H, Moschen AR, Kaser A. Obesity and the Microbiota. Gastroenterology. 2009;136:1476-83.
[68] Nehra V, Allen JM, Mailing LJ, Kashyap PC, Woods JA. Gut Microbiota: Modulation of Host Physiology in Obesity. Physiology. 2016;31:327-35.
[69] Alterations in the Intestinal Microbiota Promote Colorectal Cancer. Cancer Discov. 2012;2:866-.
[70] Franceschi F, Roccarina D, Gigante G, De Marco G, Marini M, Giupponi B, et al. Microbiota and colon cancer. Eur Rev Med Pharmaco. 2009;13:95-.
[71] Moos WH, Faller DV, Harpp DN, Kanara I, Pernokas J, Powers WR, et al. Microbiota and Neurological Disorders: A Gut Feeling. Bioresearch Open Acc. 2016;5:137-45.
[72] Biton M, Levin A, Slyper M, Alkalay I, Horwitz E, Mor H, et al. Epithelial microRNAs regulate gut mucosal immunity via epithelium-T cell crosstalk. Nat Immunol. 2011;12:239-46.
[73] Clevers H. Identification of stem cells in small intestine and colon by a single marker gene Lgr5. Ejc Suppl. 2008;6:1-.
[74] Wang ML, Keilbaugh SA, Cash-Mason T, He XC, Li LH, Wu GD. Immune-mediated signaling in intestinal goblet cells via PI3-kinase-and AKT-dependent pathways. Am J Physiol-Gastr L. 2008;295:G1122-G30.
[75] Cliffe LJ, Grencis RK. The Trichuris muris system: a paradigm of resistance and susceptibility to intestinal nematode infection. Adv Parasit. 2004;57:255-307.

[76] Singh N, Shirdel EA, Waldron L, Zhang RH, Jurisica I, Comelli EM. The Murine Caecal MicroRNA Signature Depends on the Presence of the Endogenous Microbiota. Int J Biol Sci. 2012;8:171-86.

[77] Dalmasso G, Hang TTN, Yan YT, Laroui H, Charania MA, Ayyadurai S, et al. Microbiota Modulates Host Gene Expression via MicroRNAs. Gastroenterology. 2011;140:S663-S.

[78] Runtsch MC, Hu RZ, Alexander M, Wallace J, Kagele D, Petersen C, et al. MicroRNA-146a constrains multiple parameters of intestinal immunity and increases susceptibility to DSS colitis. Oncotarget. 2015;6:28556-72.

[79] Dai X, Chen X, Chen Q, Shi L, Liang HW, Zhou Z, et al. MicroRNA-193a-3p Reduces Intestinal Inflammation in Response to Microbiota via Down-regulation of Colonic PepT1. J Biol Chem. 2015;290:16099-115.

[80] Liu SR, da Cunha AP, Rezende RM, Cialic R, Wei ZY, Bry L, et al. The Host Shapes the Gut Microbiota via Fecal MicroRNA. Cell Host & Microbe. 2016;19:32-43.

[81] Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Research. 2014;42:D68-D73.

[82] Braniste V, Al-Asmakh M, Kowal C, Anuar F, Abbaspour A, Toth M, et al. The gut microbiota influences blood-brain barrier permeability in mice. Sci Transl Med. 2014;6.

[83] Van Niel G, Mallegol J, Bevilacqua C, Candalh C, Brugiere S, Tomaskovic-Crook E, et al. Intestinal epithelial exosomes carry MHC class II/peptides able to inform the immune system in mice. Gut. 2003;52:1690-7.

[84] Harding C, Heuser J, Stahl P. Receptor-Mediated Endocytosis of Transferrin and Recycling of the Transferrin Receptor in Rat Reticulocytes. Journal of Cell Biology. 1983;97:329-39.

[85] Nolte-'t Hoen EN, Buermans HP, Waasdorp M, Stoorvogel W, Wauben MH, t Hoen PA. Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions. Nucleic Acids Res. 2012;40:9272-85.

[86] Li CC, Eaton SA, Young PE, Lee M, Shuttleworth R, Humphreys DT, et al. Glioma microvesicles carry selectively packaged coding and non-coding RNAs which alter gene expression in recipient cells. RNA Biol. 2013;10:1333-44.

[87] Kalra H, Simpson RJ, Ji H, Aikawa E, Altevogt P, Askenase P, et al. Vesiclepedia: a compendium for extracellular vesicles with continuous community annotation. PLoS Biol. 2012;10:e1001450.
[88] Xiao D, Ohlendorf J, Chen Y, Taylor DD, Rai SN, Waigel S, et al. Identifying mRNA, microRNA and protein profiles of melanoma exosomes. PLoS One. 2012;7:e46874.
[89] Pegtel DM, Cosmopoulos K, Thorley-Lawson DA, van Eijndhoven MA, Hopmans ES, Lindenberg JL, et al. Functional delivery of viral miRNAs via exosomes. Proc Natl Acad Sci U S A. 2010;107:6328-33.
[90] Nanbo A, Kawanishi E, Yoshida R, Yoshiyama H. Exosomes Derived from Epstein-Barr Virus-Infected Cells Are Internalized via Caveola-Dependent Endocytosis and Promote Phenotypic Modulation in Target Cells. Journal of Virology. 2013;87:10334-47.
[91] Chevillet JR, Kang Q, Ruf IK, Briggs HA, Vojtech LN, Hughes SM, et al. Quantitative and stoichiometric analysis of the microRNA content of exosomes. P Natl Acad Sci USA. 2014;111:14888-93.
[92] Koppers-Lalic D, Hackenberg M, Bijnsdorp IV, van Eijndhoven MAJ, Sadek P, Sie D, et al. Nontemplated Nucleotide Additions Distinguish the Small RNA Composition in Cells from Exosomes. Cell Rep. 2014;8:1649-58.
[93] Skog J, Noerholm M, Bentink S, Romain C, Fishbeck J, Sinclair I, et al. Development of a urine microvesicle/exosome RNA biomarker panel to identify prostate cancer. Mol Cancer Ther. 2013;12.
[94] Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. P Natl Acad Sci USA. 2008;105:10513-8.
[95] Huang XY, Yuan TZ, Tschannen M, Sun ZF, Jacob H, Du MJ, et al. Characterization of human plasma-derived exosomal RNAs by deep sequencing. Bmc Genomics. 2013;14.
[96] Chen T, Xi QY, Ye RS, Cheng X, Qi QE, Wang SB, et al. Exploration of microRNAs in porcine milk exosomes. Bmc Genomics. 2014;15.
[97] Vojtech L, Woo S, Hughes S, Levy C, Ballweber L, Sauteraud RP, et al. Exosomes in human semen carry a distinctive repertoire of small non-coding RNAs with potential regulatory functions. Nucleic Acids Research. 2014;42:7290-304.

[98] Michael A, Bajracharya SD, Yuen PST, Zhou H, Star RA, Illei GG, et al. Exosomes from human saliva as a source of microRNA biomarkers. Oral Dis. 2010;16:34-8.
[99] Baraniskin A, Kuhnhenn J, Schlegel U, Chan A, Deckert M, Gold R, et al. Identification of microRNAs in the cerebrospinal fluid as marker for primary diffuse large B-cell lymphoma of the central nervous system. Blood. 2011;117:3140-6.
[100] Nilsson J, Skog J, Nordstrand A, Baranov V, Mincheva-Nilsson L, Breakefield XO, et al. Prostate cancer-derived urine exosomes: a novel approach to biomarkers for prostate cancer. Brit J Cancer. 2009;100:1603-7.
[101] Kahlert C, Kalluri R. Exosomes in tumor microenvironment influence cancer progression and metastasis. J Mol Med. 2013;91:431-7.
[102] Peres da Silva R, Puccia R, Rodrigues ML, Oliveira DL, Joffe LS, Cesar GV, et al. Extracellular vesicle-mediated export of fungal RNA. Sci Rep. 2015;5:7763.
[103] Lambertz U, Ovando MEO, Vasconcelos EJR, Unrau PJ, Myler PJ, Reiner NE. Small RNAs derived from tRNAs and rRNAs are highly enriched in exosomes from both old and new world Leishmania providing evidence for conserved exosomal RNA Packaging. Bmc Genomics. 2015;16.
[104] Fernandez-Calero T, Garcia-Silva R, Pena A, Robello C, Persson H, Rovira C, et al. Profiling of small RNA cargo of extracellular vesicles shed by Trypanosoma cruzi reveals a specific extracellular signature. Mol Biochem Parasit. 2015;199:19-28.
[105] Halperin W, Jensen WA. Ultrastructural Changes during Growth and Embryogenesis in Carrot Cell Cultures. J Ultra Mol Struct R. 1967;18:428-&.
[106] Tanchak MA, Fowke LC. The Morphology of Multivesicular Bodies in Soybean Protoplasts and Their Role in Endocytosis. Protoplasma. 1987;138:173-82.
[107] Hill AF, Pegtel DM, Lambertz U, Leonardi T, O'Driscoll L, Pluchino S, et al. ISEV position paper: extracellular vesicle RNA analysis and bioinformatics. J Extracell Vesicles. 2013;2.
[108] Akers JC, Gonda D, Kim R, Carter BS, Chen CC. Biogenesis of extracellular vesicles (EV): exosomes, microvesicles, retrovirus-like vesicles, and apoptotic bodies. J Neurooncol. 2013;113:1-11.
[109] Ihara T, Yamamoto T, Sugamata M, Okumura H, Ueno Y. The process of ultrastructural changes from nuclei to apoptotic body. Virchows Arch. 1998;433:443-7.

[110] Elmore S. Apoptosis: a review of programmed cell death. Toxicol Pathol. 2007;35:495-516.
[111] Muster T, Waltenberger A, Grassauer A, Hirschl S, Caucig P, Romirer I, et al. An endogenous retrovirus derived from human melanoma cells. Cancer Res. 2003;63:8735-41.
[112] Contreras-Galindo R, Kaplan MH, Leissner P, Verjat T, Ferlenghi I, Bagnoli F, et al. Human endogenous retrovirus K (HML-2) elements in the plasma of people with lymphoma and breast cancer. J Virol. 2008;82:9329-36.
[113] Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. P Natl Acad Sci USA. 2011;108:5003-8.
[114] Turchinovich A, Weiz L, Langheinz A, Burwinkel B. Characterization of extracellular circulating microRNA. Nucleic Acids Research. 2011;39:7223-33.
[115] Belting M, Wittrup A. Nanotubes, exosomes, and nucleic acid-binding peptides provide novel mechanisms of intercellular communication in eukaryotic cells: implications in health and disease. J Cell Biol. 2008;183:1187-91.
[116] Rustom A, Saffrich R, Markovic I, Walther P, Gerdes HH. Nanotubular highways for intercellular organelle transport. Science. 2004;303:1007-10.
[117] Sowinski S, Jolly C, Berninghausen O, Purbhoo MA, Chauveau A, Kohler K, et al. Membrane nanotubes physically connect T cells over long distances presenting a novel route for HIV-1 transmission. Nat Cell Biol. 2008;10:211-9.
[118] Kowal J, Arras G, Colombo M, Jouve M, Morath JP, Primdal-Bengtson B, et al. Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. Proc Natl Acad Sci U S A. 2016;113:E968-77.
[119] Lazaro-Ibanez E, Sanz-Garcia A, Visakorpi T, Escobedo-Lucea C, Siljander P, Ayuso-Sacido A, et al. Different gDNA content in the subpopulations of prostate cancer extracellular vesicles: apoptotic bodies, microvesicles, and exosomes. Prostate. 2014;74:1379-90.
[120] Crescitelli R, Lasser C, Szabo TG, Kittel A, Eldh M, Dianzani I, et al. Distinct RNA profiles in subpopulations of extracellular vesicles: apoptotic bodies, microvesicles and exosomes. J Extracell Vesicles. 2013;2.

[121] Kanada M, Bachmann MH, Hardy JW, Frimannson DO, Bronsart L, Wang A, et al. Differential fates of biomolecules delivered to target cells via extracellular vesicles. Proc Natl Acad Sci U S A. 2015;112:E1433-42.
[122] Stoorvogel W, Kleijmeer MJ, Geuze HJ, Raposo G. The biogenesis and functions of exosomes. Traffic. 2002;3:321-30.
[123] Raposo G, Stoorvogel W. Extracellular vesicles: exosomes, microvesicles, and friends. J Cell Biol. 2013;200:373-83.
[124] Martin KC, Ephrussi A. mRNA localization: gene expression in the spatial dimension. Cell. 2009;136:719-30.
[125] Cody NA, Iampietro C, Lécuyer E. The many functions of mRNA localization during normal development and disease: From pillar to post. WIREs Dev Biol. 2013.
[126] MacDonald PM. bicoid mRNA localization signal: phylogenetic conservation of function and RNA secondary structure. Development. 1990;110:161-71.
[127] Lecuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, et al. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell. 2007;131:174-87.
[128] Latham VM, Jr., Kislauskis EH, Singer RH, Ross AF. Beta-actin mRNA localization is regulated by signal transduction mechanisms. J Cell Biol. 1994;126:1211-9.
[129] Du TG, Schmid M, Jansen RP. Why cells move messages: The biological functions of mRNA localization. Semin Cell Dev Biol. 2007.
[130] Czaplinski K, Singer RH. Pathways for mRNA localization in the cytoplasm. Trends Biochem Sci. 2006;31:687-93.
[131] Batagov AO, Kuznetsov VA, Kurochkin IV. Identification of nucleotide patterns enriched in secreted RNAs as putative cis-acting elements targeting them to exosome nano-vesicles. BMC Genomics. 2011;12 Suppl 3:S18.
[132] Skog J, Wurdinger T, van Rijn S, Meijer DH, Gainche L, Sena-Esteves M, et al. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. Nat Cell Biol. 2008;10:1470-6.
[133] Bolukbasi MF, Mizrak A, Ozdener GB, Madlener S, Strobel T, Erkan EP, et al. miR-1289 and "Zipcode"-like Sequence Enrich mRNAs in Microvesicles. Mol Ther-Nucl Acids. 2012;1.

[134] Bellingham SA, Coleman BM, Hill AF. Small RNA deep sequencing reveals a distinct miRNA signature released in exosomes from prion-infected neuronal cells. Nucleic Acids Research. 2012;40:10937-49.
[135] Hung ME, Leonard JN. A platform for actively loading cargo RNA to elucidate limiting steps in EV-mediated delivery. Journal of Extracellular Vesicles. 2016;5.
[136] King HW, Michael MZ, Gleadle JM. Hypoxic enhancement of exosome release by breast cancer cells. Bmc Cancer. 2012;12:421.
[137] Squadrito ML, Baer C, Burdet F, Maderna C, Gilfillan GD, Lyle R, et al. Endogenous RNAs Modulate MicroRNA Sorting to Exosomes and Transfer to Acceptor Cells. Cell Rep. 2014;8:1432-46.
[138] Cha DJ, Franklin JL, Dou Y, Liu Q, Higginbotham JN, Demory Beckler M, et al. KRAS-dependent sorting of miRNA to exosomes. Elife. 2015;4:e07197.
[139] Chen TS, Lai RC, Lee MM, Choo AB, Lee CN, Lim SK. Mesenchymal stem cell secretes microparticles enriched in pre-microRNAs. Nucleic Acids Res. 2010;38:215-24.
[140] Villarroya-Beltri C, Gutierrez-Vazquez C, Sanchez-Cabo F, Perez-Hernandez D, Vazquez J, Martin-Cofreces N, et al. Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs. Nature Communications. 2013;4.
[141] Gerke V, Creutz CE, Moss SE. Annexins: Linking Ca2+ signalling to membrane dynamics. Nat Rev Mol Cell Bio. 2005;6:449-61.
[142] Filipenko NR, MacLeod TJ, Yoon CS, Waisman DM. Annexin A2 is a novel RNA-binding protein. J Biol Chem. 2004;279:8723-31.
[143] Hagiwara K, Katsuda T, Gailhouste L, Kosaka N, Ochiya T. Commitment of Annexin A2 in recruitment of microRNAs into extracellular vesicles. Febs Lett. 2015;589:4071-8.
[144] Choi DS, Kim DK, Kim YK, Gho YS. Proteomics, transcriptomics and lipidomics of exosomes and ectosomes. Proteomics. 2013;13:1554-71.
[145] de Gassart A, Geminard C, Fevrier B, Raposo G, Vidal M. Lipid raft-associated protein sorting in exosomes. Blood. 2003;102:4336-44.
[146] Dubois L, Ronquist KK, Ek B, Ronquist G, Larsson A. Proteomic Profiling of Detergent Resistant Membranes (Lipid Rafts) of Prostasomes. Mol Cell Proteomics. 2015;14:3015-22.
[147] Megha, London E. Ceramide selectively displaces cholesterol from ordered lipid domains (rafts): implications for lipid raft structure and function. J Biol Chem. 2004;279:9997-10004.

[148] Trajkovic K, Hsu C, Chiantia S, Rajendran L, Wenzel D, Wieland F, et al. Ceramide triggers budding of exosome vesicles into multivesicular Endosomes. Science. 2008;319:1244-7.
[149] Essandoh K, Yang L, Huang W, Wang X, Qin D, Wang Y, et al. Blockade of Exosome Release with Gw4869 Dampens Sepsis-Induced Inflammation and Cardiac Dysfunction. Shock. 2015;43:123-.
[150] Yuyama K, Sun H, Mitsutake S, Igarashi Y. Sphingolipid-modulated Exosome Secretion Promotes Clearance of Amyloid-beta by Microglia. J Biol Chem. 2012;287:10977-89.
[151] Kosaka N, Iguchi H, Hagiwara K, Yoshioka Y, Takeshita F, Ochiya T. Neutral Sphingomyelinase 2 (nSMase2)-dependent Exosomal Transfer of Angiogenic MicroRNAs Regulate Cancer Cell Metastasis. J Biol Chem. 2013;288:10849-59.
[152] Scott DD, Norbury CJ. RNA decay via 3' uridylation. Biochim Biophys Acta. 2013;1829:654-65.
[153] Siomi H, Siomi MC. RISC hitches onto endosome trafficking. Nat Cell Biol. 2009;11:1049-51.
[154] Gibbings DJ, Ciaudo C, Erhardt M, Voinnet O. Multivesicular bodies associate with components of miRNA effector complexes and modulate miRNA activity. Nat Cell Biol. 2009;11:1143-9.
[155] Schmidt O, Teis D. The ESCRT machinery. Curr Biol. 2012;22:R116-R20.
[156] Irion U, St Johnston D. bicoid RNA localization requires specific binding of an endosomal sorting complex. Nature. 2007;445:554-8.
[157] McKenzie AJ, Hoshino D, Hong NH, Cha DJ, Franklin JL, Coffey RJ, et al. KRAS-MEK Signaling Controls Ago2 Sorting into Exosomes. Cell Rep. 2016;15:978-87.
[158] Zomer A, Maynard C, Verweij FJ, Kamermans A, Schafer R, Beerling E, et al. In Vivo imaging reveals extracellular vesicle-mediated phenocopying of metastatic behavior. Cell. 2015;161:1046-57.
[159] Deolindo P, Evans-Osses I, Ramirez MI. Microvesicles and exosomes as vehicles between protozoan and host cell communication. Biochem Soc Trans. 2013;41:252-7.
[160] Evans-Osses I, Reichembach LH, Ramirez MI. Exosomes or microvesicles? Two kinds of extracellular vesicles with different routes to modify protozoan-host cell interaction. Parasitol Res. 2015;114:3567-75.

[161] Katsuda T, Ikeda S, Yoshioka Y, Kosaka N, Kawamata M, Ochiya T. Physiological and pathological relevance of secretory microRNAs and a perspective on their clinical application. Biol Chem. 2014;395:365-73.
[162] Fitzgerald A, van Kan JAL, Plummer KM. Simultaneous silencing of multiple genes in the apple scab fungus, Venturia inaequalis, by expression of RNA with chimeric inverted repeats. Fungal Genet Biol. 2004;41:963-71.
[163] Pitino M, Coleman AD, Maffei ME, Ridout CJ, Hogenhout SA. Silencing of Aphid Genes by dsRNA Feeding from Plants. Plos One. 2011;6.
[164] Zhu JQ, Liu SM, Ma Y, Zhang JQ, Qi HS, Wei ZJ, et al. Improvement of Pest Resistance in Transgenic Tobacco Plants Expressing dsRNA of an Insect-Associated Gene EcR. Plos One. 2012;7.
[165] Zha WJ, Peng XX, Chen RZ, Du B, Zhu LL, He GC. Knockdown of Midgut Genes by dsRNA-Transgenic Plant-Mediated RNA Interference in the Hemipteran Insect Nilaparvata lugens. Plos One. 2011;6.

Préface au Chapitre 3 : Flying the RNA Nest: Drosophila Reveals Novel Insights into the Transcriptome Dynamics of Early Development (Article #2)

Ce chapitre d'introduction est présenté sous la forme d'un article de revue publié dans *Journal of Developmental Biology* (Référence : Lefebvre, F. A., & Lécuyer, E. (2017). Flying the RNA Nest : *Drosophila* Reveals Novel Insights into the Transcriptome Dynamics of Early Development. Journal of Develoopmental Biology, doi:10.3390/jdb6010005.)

L'article traite des changements dynamiques des profils transcriptionels et post-transcriptionels qui se déroulent chez l'embryon précoce suite à la fertilisation. Il explore le rôle de la localisation des ARNs, de la dégradation régulée des transcrits maternels et de l'activation du génome zygotique en insistant sur les avancées récentes liées au développement de nouvelles approches de séquençage et d'imagerie des acides nucléiques.

Dans le contexte de la présente thèse, les notions exposées dans cet article de revue permettent de contextualiser mes travaux portant sur l'embryogenèse de la Drosophile (Chapitres 7 et 8).

J'ai écrit l'intégralité du manuscrit, qui a fait l'objet d'une relecture et d'une série de corrections proposées par mon directeur de thèse, Dr Éric Lécuyer, dans le contexte de sa soumission à *Journal of Developmental Biology*.

Chapitre 3 : Flying the RNA Nest: Drosophila Reveals Novel Insights into the Transcriptome Dynamics of Early Development (Article #2)

Chapitre 3 : Article #2

Flying the RNA Nest: *Drosophila* Reveals Novel Insights into the Transcriptome Dynamics of Early Development

Fabio Alexis Lefebvre[1,2], Éric Lécuyer[1,2,3,4]

1- Institut de Recherches Cliniques de Montréal (IRCM)
Montréal, Québec, Canada

2- Département de Biochimie
Université de Montréal, Montréal, Québec, Canada

3- Division of Experimental Medicine
McGill University, Montréal, Québec, Canada

4- Address correspondences to: Dr. Eric Lécuyer
IRCM, RNA Biology Laboratory
110 Avenue des Pins, Ouest
Montréal, Québec, Canada
H2W 1R7
Tel: 514-987-5646, Fax: 514-987-5752
Email: Eric.Lecuyer@ircm.qc.ca

Abstract

Early development is punctuated by a series of pervasive and fast paced transitions. These events reshape a differentiated oocyte into a totipotent embryo and allow it to gradually mount a genetic program of its own, thereby framing a new organism. Specifically, developmental transitions entail a deep remodelling of transcriptional landscapes at egg activation, during early cleavage cycles and at the midblastula stage. *Drosophila* provides an elegant and genetically tractable system to investigate these conserved changes at a dazzling developmental pace. Here, we review recent studies applying emerging technologies such as ribosome profiling, *in situ* Hi-C chromatin probing and the MS2 RNA imaging system to investigate the transcriptional dynamics at play during *Drosophila* embryogenesis. In light of this new literature, we revisit the main models of zygotic genome activation (ZGA) and attempt to conciliate their diverging propositions. We also review the contributions played by zygotic transcription in shaping embryogenesis and explore emerging properties of zygotic transcription, including transcriptional bursting and transcriptional memory.

Introduction

Early development unfolds through exquisitely coordinated and deeply conserved biological transitions. Fertilization entails the remodeling of a differentiated oocyte into a totipotent embryo, which involves profound genomic, transcriptomic and proteomic re-organization [1]. Early embryos execute a dazzling proliferative push driven by maternally provided gene products to increase cell number. The rapid pace of these early divisions, termed cleavage cycles, is achieved through copious supplies of maternal cyclins, an abbreviated DNA replication execution and the absence of active growth and mitotic checkpoints [2]. In most systems, zygotic nuclei remain transcriptionally silent during this period. As embryogenesis proceeds, cell cycle duration progressively lengthens, reflecting the gradual emergence of the DNA replication checkpoint and the increasing destabilization of maternal cyclins [3] (**Figure 3.1**). Interphase lengthening broadly coincides with progressive zygotic genome activation (ZGA), i.e. *de novo* expression of robust populations of functional transcripts [4] [5, 6]. New zygotic products gradually take over the pool of maternal RNAs, which undergo progressive clearance through regulated degradation mechanisms [7] [8] [9] (**Figure 3.1**). The juxtaposition of ZGA and maternal clearance gradually remodels the transcriptome, a process termed the mother-to-zygotic transition (MZT). The MZT ends at a key developmental time point called the midblastula transition (MBT), which typically involves a dramatic cascade of anatomical and physiological changes that are dependent on zygotic transcription.

Drosophila and many arthropods display a facultative parthenogenetic mode of reproduction, meaning that egg activation can take place independently of sperm entry, although this rarely occurs. The egg-to-embryo transition is triggered by changes in pressure and osmotic balance as the mature oocyte exits through the uterus [10] [11]. Egg activation involves the completion of meiosis and the initiation of fast-paced mitotic divisions, thereby setting the onset of embryogenesis. In *Drosophila*, the first rounds of nuclear divisions arise every 9 minutes, leading to the formation of 6,000 nuclei in only 2 hours [2] (Foe and Alberts, 1983). To facilitate the rapid pace of these divisions and the synchronization of mitotic entry, nuclear cycles (NC) take place within a single large syncytial cell [12] [13] (**Figure 3.1**). The first syncytial divisions

are metasynchronous, and proceed under maternal control, while zygotic nuclei remain largely transcriptionally quiescent [14] [15].

During the interphase of NC8 and NC9, nuclei start migrating to the cortical periphery, forming a syncytial blastoderm embryo [16]. This process coincides with gradual interphase lengthening, reflecting increasing long periods of Cyclin-dependent kinase 1 (Cdk1) inhibition. With its cyclin partners, Cdk1 acts as the chief regulator of cell cycle progression through the phosphorylation of a wide range of protein targets, which notably mediate S phase initiation, spindle assembly and sister chromatid alignment. Cdk1 inhibition expands as the cytoplasmic pool of maternal cyclins, notably String and Twine, is progressively depleted and as the DNA replication checkpoints emerge (**Figure 3.1**). In hiatus during the first cycles, the checkpoint safeguards genome integrity by preventing mitotic entry when single-stranded (ss)DNA is sensed, underlying incomplete replication or extensive damage [17] [18]. When the MBT takes place at NC13, interphase duration suddenly triples as mitotic synchrony is lost and cortical nuclei secede from the syncytium to form a well-defined primordial epithelium, a process termed cellularization [2]. *En masse* zygotic transcription ensues in cells now endowed with motility and a susceptibility for apoptosis, new-found attributes that will play crucial roles in gastrulation, neurulation and organogenesis [19] [20].

Evolutionarily conserved features of the MZT have been well reviewed by Tadros and Lipsitz (2009) and more recently by Langley and colleagues (2014). This exciting area of developmental biology has further expanded over the last few years, along with our understanding of the complex cross-talks resulting in the interrelated emergence of ZGA, checkpoint activation and cellularization. Indeed, the deployment of disruptive technologies to track translation, probe chromosome conformation and image single RNA molecules in *Drosophila* have revealed new insights into the organization of the transcriptome during embryogenesis. Here, we focus on key findings established in recent years that paint an increasingly resolved picture of transcriptome dynamics during early development. We first review the changes in transcriptomic landscapes that take place as maternally-deposited transcripts are selectively localized, translated and degraded. We discuss the different models proposed to regulate the onset of zygotic transcription and focus on the remodelling of chromatin during early embryogenesis. We then review the roles

played by zygotic transcription in promoting cellularization, maternal clearance and replication checkpoint activation. Finally, we outline novel properties of zygotic transcription revealed by real-time, *in vivo* imaging, including transcriptional bursting and transcriptional memory.

Deposition, localization, translation and clearance of maternal transcripts

Early *Drosophila* development is driven by maternal proteins and RNA transcripts synthesized by multiploid nurse cells and deposited in the egg during oogenesis [21] [22]. Long before fertilization, structural and biosynthetic factors such as ribosomal and glycolytic proteins are endowed in the oocyte along with their corresponding mRNAs [23]. These factors will direct rapid mitotic cycling and sustain DNA replication in early embryogenesis, while zygotic nuclei remain transcriptionally quiescent. Maternal deposition is widespread : up to 65% of all annotated *Drosophila* mRNAs can be detected during the first syncytial cycles [24] [25] [26]. Among these products, a set of maternal mRNAs acquire spatially-resolved localization in the oocyte through a series of symmetry breaking events during ovarian development. After egg activation, the asymmetric distributions of these transcripts, notably *bicoid*, *oskar* and *gurken*, defines anteroposterior and dorsoventral axes, which will later determine the body plan of the developing fly [27, 28] [29]. Aside from these classically-defined examples, recent large-scale *fluorescent in situ hybridization* (FISH) surveys have demonstrated that subellular localization is a pervasive feature of the *Drosophila* transcriptome. Indeed, detailed profiling of RNA expression/localization features in fly oocytes and embryos, as tabulated in the Fly-FISH (http://fly-fish.ccbr.utoronto.ca/) and Dresden Ovary Table/DOT (http://tomancak-srv1.mpi-cbg.de/DOT/main.html) database repositories, revealed that the vast majority of mRNAs and long non-coding RNAs are localized in a striking array of distribution patterns [24] [30]. While these studies underline the dynamic localization properties of both maternal and zygotic RNA populations, the underlying regulatory mechanisms and functions for the most part remain to be characterized.

The transcriptomic landscape is highly dynamic during the cleavage cycles, as several RNA degradation pathways operate to selectively remove large sets of deposited transcripts. Indeed, a

study using chromosomal deletions to track the post-transcriptional dynamics of maternal and zygotic transcripts has shown that approximately 35% of maternally-deposited mRNAs are cleared out by the MBT [31]. Comparisons of degradation dynamics in unfertilized and fertilized embryos have revealed the prevalence of at least two pathways. The early onset maternal pathway operates independently of ZGA and accounts for the destabilization of approximately 20% of all mRNAs [8] (**Figure 3.2**). This maternal degradation program reflects the coordinated activity of RNA-binding proteins (RBPs) that associate with specific subsets of maternal transcripts and recruit the CCR4-NOT deadenylase complex to initiate their degradation [32]. Maternal clearance is a highly-coordinated process determined by the interplay of *cis*-acting motifs, generally found in the 3'UTR of the target transcripts, and RBPs that adopt tightly regulated spatial and temporal distributions. Indeed, mRNAs encoding the RBP Smaug (SMG) form an anterior-to-posterior gradient in the oocyte and are translationally regulated at egg activation [33]. SMG activity peaks at NC10, enacting an elegant spatial and temporal regulation of maternal clearance [32-34]. In addition, three RBPs, Pumilio (PUM), Brain tumor (BRAT) and ME31B have more recently been shown to have non-overlapping roles in mediating the decay of hundreds of maternal transcripts [35] [36].

Strong evidence in zebrafish, *Xenopus* and mouse has long suggested that some of the zygotic actors involved in maternal clearance are microRNAs (miRNAs) [37]. Small-interfering RNAs are versatile regulators that can destabilize vast pools of mRNA targets [38] [37]. In support of this model, SMG is required for the zygotic expression of the miR-309 family [39]. Among the miRNAs expressed during the MZT, over 70 exhibit compromised levels in *smg* mutants. In addition, Argonaute-1, a key component of the miRNA-induced silencing complex, showed decreased levels in *smg* mutants. Furthermore, the clearance of predicted maternal targets of SMG-dependent miRNAs is hampered in *smg* mutants, suggesting that these small RNAs contribute to the zygotic component of maternal mRNA degradation [40] [39]. As a consequence of the specificity and diversity of clearance pathways effective during the MZT, the degradation kinetics of maternal transcripts exhibit striking diversity (**Figure 3.3**). The abundance of transcripts strictly targeted by maternal degradation RBPs, such as nanos (nos), starts declining linearly after egg activation. By contrast, targets of zygotic degradation effectors such as bicoid (bcd) are untouched during the first cycles and their removal begins after ZGA. Finally,

transcripts targeted through both maternal and zygotic pathways include Hsp83 and adopt highly specific degradation kinetics that reflect the contributions and levels of both maternal and zygotic factors.

The execution of maternal clearance is tightly coupled to translational regulation of RBPs that mediate mRNA degradation. Indeed, the levels of SMG depend on the PAN GU (PNG) Ser/Thr kinase complex, a key translational regulator activated at the oocyte-to-embryo transition. The PNG complex is involved in reprogramming the proteome as the egg becomes an embryo by regulating the translation of hundreds of maternal mRNAs (**Figure 3**). Its notable targets include *cyclin B* (*cycB*), which encodes a key regulator of the embryonic cell cycle [41] [42] [43]. The complex consists of three proteins, PNG, GNU and PLU. Mutations in any of these genes leads to a profound decrease in Cyclin B protein levels, without affecting the corresponding mRNA levels. A recent study revealed a feedback loop involved in the regulation of PNG activity by the Cyclin B/CDK1 complex at the oocyte-to-embryo transition [44]. In mature oocytes, PNG activity is kept in check through inhibitory CyclinB/CDK1-dependent phosphorylation of the GNU activating subunit, preventing its association with PNG. Meiosis completion coincides with a decrease in CyclinB/CDK1 activity, leading to GNU dephosphorylation, which can then activate the PNG kinase, unleashing its widespread translational activation. However, sustained PNG kinase activity leads to a decrease in GNU levels, providing a mechanism to end PNG kinase function after egg activation and restrict its activity to the oocyte-to-embryo transition period (**Figure 3.3**).

In another study, ribosome profiling performed on mature oocytes and activated eggs derived from *png* mutant mothers revealed compromised translational efficiencies in nearly 1,000 mRNAs [45]. Surprisingly, it *was* found that translational upregulation poorly reflects on protein levels in fertilized eggs, leading the authors to propose the existence of a ‘resetting’ process in which enhanced translation is counterbalanced by proteasomal degradation, perhaps enabling the removal of proteins bearing oocyte-specific posttranslational modifications. Interestingly, this model is reminiscent of the well-characterized MZT phenomenon that takes place later in embryogenesis, as many maternal transcripts are degraded and subsequently expressed *de novo* from the zygotic genome [22].

Overall, spatiotemporal regulation enacted through mRNA localization and translation control are key features of embryonic transitions. In the absence of large-scale zygotic transcription, maternally deposited mRNAs and their regulated translation drive the developmental program of early embryogenesis. These maternal transcripts are then selectively removed, a process relying on the recognition of *cis*-acting motifs by maternal RBPs that are tightly regulated in time and space. As zygotic transcription emerges, a second phase of maternal clearance unfolds through the activity of zygotically expressed determinants (e.g. miRNAs and RBPs).

Models of zygotic genome activation

As maternal clearance takes place during the cleavage cycles, the transcriptomic landscape is remodelled and new populations of transcripts arise upon activation of the zygotic genome. The mechanisms behind ZGA onset are poorly understood and remain an outstanding question in developmental biology, although several models have been proposed [7]. Seminal work involving the injection of plasmids in *Xenopus* embryos showed that early transcription is possible prior to ZGA [46, 47]. However, the expression of plasmid DNA is rapidly silenced and resumes at the normal timing of transcription initiation. This observation suggests that the early zygotic genome is transcriptionally competent and actively repressed during the cleavage cycles. Moreover, ZGA is a gradual process and delineating its onset has remained challenging. Indeed, the emergence of the first zygotic transcripts has long been associated with the acquisition of the syncytial blastoderm morphology, at NC8. However, the detection of a small subset of zygotic genes before NC7, including the transcription factor *engrailed* (*en*), has recently been reported in preblastoderm embryos, with key implications in establishing the synchrony of early mitotic cycles [48]. Regardless of the exact onset of their expression, the first zygotic products display a conserved tendency to encode few exons and their length is shorter than maternally provided mRNAs [5]. In line with this observation, one hypothesis is that transcription is systematically attempted during the first cleavage cycles but that nascent transcripts are largely aborted, due to excessively fast cycling [49] [46] [50] [51] [15] [7]. Hence, zygotic genes may be shorter than

populations contributed maternally because Pol II can complete their transcription prior to an intervening mitotic entry [5, 6].

One long-standing model posits that ZGA is triggered once a critical value of nucleocytoplasmic (N/C) ratio is breached in the syncytium [52] [53] [4] [2] [15]. Indeed, the ratio of nuclei to cytoplasm increases rapidly during early embryogenesis, as the nuclear count expands exponentially and the volume of syncytial cytoplasm is kept constant due to growth inhibition. Tampering with this ratio by constricting embryos and compounding or reducing chromosome size impacts the number of syncytial divisions and the onset of cellularization [53]. It is thought that the increasing DNA mass titrates a maternal factor required to sustain fast-paced proliferation, until its concentration has decreased sufficiently to terminate the early cell cycle program. Similarly, a transcriptional repressor inherited maternally could be embedded in chromatin and diluted as the mass of zygotic DNA increases. After a certain N/C threshold is breached, the concentration of this repressor would have sufficiently decreased to allow zygotic transcription to take place. One enduring rival of the N/C proposition has been called the "molecular clock" model; it proposes that egg activation sets a chronological countdown that times MZT events, including ZGA [2] [15]. A study comparing the onset of transcription in haploid and diploid embryos, which differ markedly in N/C ratio, found little difference in the expression dynamics of most zygotic genes, suggesting that the molecular clock model is the best overall predictor of ZGA onset in *Drosophila*. Interestingly, the authors identified a small subset of genes regulated in a N/C-dependent manner, suggesting that both models prevail, albeit at distinct loci and frequencies [54]. Nevertheless, the molecular identity that senses the clock to activate zygotic transcription independently of the N/C ratio remains elusive.

Hence, diverse mechanisms have been proposed to account for ZGA onset, each of which is supported by empirical evidence [7]. How these different propositions converge to enact a precise execution of genome activation remains elusive. The key to ZGA regulation might lie in the reorganization of chromatin. Interestingly, a rapidly expanding literature is helping to define how chromatin states relate to transcription in early embryos (**Figure 3.4-3.5**).

Chromatin rearrangements and zygotic genome activation

Histone proteins are fundamental components of chromatin and developmentally-regulated changes in their expression could have profound impacts on genome activation. Indeed, most metazoans express tissue-specific variants of the linker histone H1 and the somatic H1 is often replaced by a developmental variant during early embryogenesis [55] [56]. In a recent article, *Pérez-Montero et al.* identified the first H1 variant in *Drosophila*, called dBigH1, and demonstrate its involvement in ZGA regulation [57]. Ubiquitous in preblastoderm and syncytial blastoderm embryos, dBigH1 is progressively replaced by dH1 in somatic cells at the MBT, except in primordial germ cells (PGCs), in which it is retained well after gastrulation (**Figure 5**). $BigH1^{00}$ mutants exhibit high embryonic lethality and a range of developmental defects, including altered nuclear distributions and highly asynchronous divisions. Interestingly, ChIP showed that Pol II is recruited to chromatin earlier in $BigH1^{00}$ mutants than in *wt* embryos. In addition, zygotic mRNAs were more abundant in $BigH1^{00}$ mutants than in *wt* embryos 2h after fertilization. Together, these results show that BigH1, a novel *Drosophila* variant of the linker histone, regulates ZGA onset and is removed from chromatin prior to the MBT.

The histone code refers to a set of posttranslational modifications that modulate chromatin compaction and the accessibility of DNA elements [58]. Changes in this epigenetic landscape likely contribute to the emergence of zygotic transcription. In a recent study, *Li et al.* investigated the genome-wide distribution of 9 histone marks using staged embryo collections at NC8, NC12 and NC14. Prior to ZGA (NC8), chromatin exists in a relatively simple state, lacking histone methylation (me) and displaying low levels of histone acetylation (ac) or nucleosome free regions (NFRs), a hallmark of transcriptional activity [59]. The acetylation marks H4K8ac, H3K18ac and H3K27ac appeared along with transcription by NC12. By contrast, H3K9ac and the methylation marks H3K4me1, H3K4me3, H3K27me3 and H3K36me3 are only apparent after the MBT, at NC14 (**Figure 3.4**). As reported by previous studies, NFRs are prevalent upstream of maternally deposited genes even in absence of zygotic transcription, suggesting that nucleosome depletion is stable across development [60]. To identify putative blastoderm enhancers, *Li et al.* calculated the cumulative binding of 16 early transcription factors and examined the sites showing the highest cumulative occupation, excluding known promoters

and coding regions. They found that putative enhancers display relatively high nucleosome density at NC8, with the appearance of acetylation marks by NC12 and H3K4me1 by early NC14, whereas the repressive mark H3K27me3 only spreads in surrounding regions by late NC14 (**Figure 3.5**).

The factors that act between NC8 and NC14 to deplete enhancer-associated nucleosomes represent major instigators of ZGA. By definition, transcription factors (TFs) capable of recognizing their binding sites in a closed chromatin context to promote chromatin remodelling are termed "pioneer TFs". Several studies have identified such pioneer activity for the Zn-finger transcription factor Zelda/Vielfaltig (ZLD/VFL), a master regulator of early zygotic gene expression [61] [62]. Indeed, ZLD is detected by NC2 in syncytial embryos and its binding displays a striking correlation with the timing and magnitude of early zygotic transcription [63]. ZLD has been shown to prime enhancers by lowering the nucleosome barrier sufficiently to promote the accession of specific binding motifs by their associated TFs [64]. These observations suggest that ZLD may act as a global genome activator in *Drosophila*, like Nanog, Pouf5f3 and SoxB1 in vertebrates [65, 66]. Indeed, *Li et al.* found that nearly all the putative blastoderm enhancers identified through cumulative TF binding at NC14 are already bound by ZLD at NC8. Moreover, the ZLD consensus motif CAGGTAG was the single most enriched sequence associated to the early enhancer marks H3K27ac, H3K18ac and H4K8ac. Finally, H3K4me1 was lost and H3K18ac strongly compromised at ZLD-bound regions in embryos obtained from *zld*$^{-}$ germline clones. Overall, *Li et al.* show that histone marks are depleted during the first cleavage cycles and emerge between NC8 and NC14. Importantly, ZLD is a pioneer TF of the MZT: it can bind its genomic sites in condensed chromatin at NC8 and promote the recruitment of other factors that carry out profound chromatin remodeling at the MBT [67, 68].

In addition to histone modifications, the three-dimensional folding of chromosomes can bring distant genomic loci in close physical proximity, with profound impacts on gene expression [69]. Topologically associated domains (TADs) are regions of high contact probabilities that display significant insulation from neighboring loci, enabling enhancer-promoter contacts and the coordination of gene expression programs [70] [71]. Chromosome conformation capture (3C), its adaptations (4C and 5C) and the recent genome-wide variant Hi-C can reveal TADs with

increasing resolution. They have been optimized in *Drosophila* embryos, enabling investigations into the developmental implications of genome architecture [72] [73] [74, 75]. Recently, *Hug et al.* performed Hi-C at time-points surrounding the ZGA to determine when chromatin architecture is established during development and how its emergence relates to the onset of zygotic transcription [76]. NC8 embryos display poorly organized chromatin, exhibiting broadly uniform contact probabilities through large genomic distances. By contrast, NC13, NC14 and gastrula embryos revealed increasingly strong enrichments of chromatin associations within TADs and sharply declining contact frequencies with the loci surrounding TADs. This picture suggests that chromatin architecture is rapidly remodelled from an unordered state in preblastoderm embryos to a structured organization by NC14 (**Figure 3.5**). These boundaries are tightly maintained in later-stage embryos and in Kc167 cells, consistent with highly stable TAD boundaries described in other models (Dixon et al. 2012).

ChIP-seq revealed a strong dose-dependent correlation between Pol II occupancy and TAD boundary-like regions, especially at housekeeping genes and across developmental stages. Analysis of the early zygotic *Bsg25*/*Elba3* locus, which is switched off before gastrulation, showed that loss of Pol II occupancy at NC14 coincides with the loss of its boundary-like structure [77]. Together, these results suggest that Pol II binding contributes to chromatin conformation reorganization. To test the role of transcription, *Hug et al.* injected NC8 embryos with the Pol II inhibitors α-amanitin and triptolide before performing Hi-C to examine chromatin architecture at the MBT. Inter-TAD insulation was compromised at NC14, as well as the co-localization of housekeeping gene boundaries, although extensive long-distance contacts were still prevalent independent of transcription. ZLD occupancy showed striking correlations with TAD boundaries by NC12, hinting at a potential role in their establishment. To explore this hypothesis, *Hug et al.* performed *in situ* Hi-C on NC14 *zld*$^{-}$ embryos, which revealed a loss of insulation of TAD-boundaries at strong ZLD sites, especially at boundaries established in early cycles. Collectively, *Hug et al.* provide strong evidence that the establishment of long-range interactions broadly coincide with ZGA. Although transcription is not required for the emergence of chromatin conformation, loci transcribed early act as nucleation sites and contribute to the establishment of TAD boundaries. Similarly, ZLD binding contribute

significantly to TAD boundary insulation, consistent with ZLD roles as global activator of the zygotic genome.

In brief, ZGA coincides with a profound reorganization of chromatin. Prior to NC8, chromatin exhibits a simple and disorganized state, with few histone modifications, NFRs or TADs. The germline-specific histone variant dBigH1 is embedded in chromatin, possibly contributing to its transcriptional silencing. Through pioneer TF activity, factors such as ZLD disrupt the nucleosome barrier between NC8 and NC12 to expose zygotic enhancers and promote the transcription of their target genes. Concomitantly, the activating histone marks H4K8ac, H3K18ac and H3K27ac appear and TADs emerge. By the MBT, dBigH1 has been replaced by histone H1, long-range interactions have gained complexity and stability and the histone marks H3K4me1, H3K4me3, H3K27me3 and H3K36me3 are established.

ZGA as a driver of embryonic development

After its emergence, zygotic transcription becomes a major driver of embryonic development. Its contribution is twofold: zygotic products directly enact important functions, notably transcription factors that reshape the developmental program and miRNAs that contribute to maternal clearance. In addition, the process of transcription itself seems to mediate changes in the biology of the embryo. Indeed, active transcription can expose ssDNA and may cause replication stalling when facing a replication fork. These processes have recently been linked to the activation of the DNA replication checkpoint before the MBT. Indeed, studies taking advantage of mutants with impaired ZGA have revealed that transcription contributes to maternal clearance and determines the onset of cellularization and replication checkpoint activation.

Sung et al. characterized a fly model exhibiting a point mutation in the 3'untranslated region (3'UTR) of the *RNPII215* gene, which encodes the large subunit of Pol II [78]. These mutants, termed X161 embryos, undergo premature zygotic transcription onset, providing an appealing model to investigate the complex relationships between zygotic transcription and other key events of the MBT (**Figure 3.6**). Interestingly, X161 embryos terminate the syncytial stage after the completion of 12 NC rather than 13 NC, suggesting that interfering with ZGA onset impacts

the timing of cellularization. To confirm this observation, *Sung et al.* considered mutants for the master transcription factor Zelda (ZLD), which fail to transcribe a broad set of early zygotic genes. They found that X161 *zld* double mutants all undergo 13 syncytial NCs, like *wt* embryos. Since ZLD loss-of-function rescues the premature transcription phenotype of X161 mutants, a normal number of syncytial cycles in X161 *zld* double mutants suggests that early transcription causes the precocious cellularization phenotype of X161 single mutants. In addition, *Sung et al.* used the X161 model to test the contribution of the nucleocytoplasmic ratio on cellularization. Haploid X161 embryos, which present a lowered N/C ratio, underwent only 12 NC, the same number as diploid X161 mutants. This result suggests that the N/C ratio acts independently of ZGA and does not directly regulate the onset of cellularization in *Drosophila*. Together, these experiments provide strong evidence that the onset of zygotic transcription times key events of early embryonic development.

As discussed earlier, maternal clearance is a complex process relying on factors contributed maternally and on the expression of zygotic products. Sung et al. surveyed the levels of three canonical targets of maternal clearance, *string*, *twine* and *smaug* in X161 mutants to monitor the impact of ZGA onset on maternal clearance. They found that the degradation of these maternal transcripts, which starts during the 14^{th} interphase in *wt* embryos, is already well advanced by NC13 in X161 embryos, in agreement with reports of a zygotic contribution to maternal clearance. The authors also found that premature ZGA leads to a precocious requirement for a functional replication checkpoint. Checkpoint activation pauses M phase entry until the completion of DNA replication to safeguard genome integrity. Its emergence is a gradual process completed at the MBT. The Ser/Thr kinase Chk1, encoded by the *Drosophila* gene *grapes* (*grp)*, is a key component of the DNA damage response (DDR), signal transduction cascades that sense DNA lesions to halt mitotic entry [17, 79]. Chk1 activity is required for progression through the MBT and its loss leads to genomic instability exemplified by chromatin defects and embryonic lethality. Chk1 activity is dispensable prior to NC13, but necessary around the MBT, when its loss leads to the apparition of genotoxic lesions. Therefore, the requirement for Chk1 activity can be used as a proxy to score the onset of checkpoint activation. *Sung et al.* found that X161 *grp* double mutants display nuclear envelope and chromatin condensation defects by the 13^{th}

interphase, one cycle earlier than *grp* single mutants, supporting a role of transcription in the onset of the DNA replication checkpoint activation.

Interestingly, the nuclear retention of zygotic transcripts has been identified as a new facet of the DDR during early embryogenesis. Indeed, *Iampietro et al.* showed that syncytial-stage embryos challenged with genotoxic stress undergo extensive nuclear fallout at the MBT, a mechanism of programmed elimination [18]. The authors showed that fallout nuclei display widespread nuclear retention of diverse zygotic transcripts, including histone mRNAs. The nuclear retention of histone mRNAs is linked to a Chk2-mediated phosphorylation of the stem loop binding protein (SLBP), which orchestrates the posttranscriptional processing and nuclear export of histone mRNAs. In turn, the nuclear retention of essential mRNAs such as histones leads to a local depletion of their corresponding proteins in the vicinity of damaged nuclei, promoting their fallout and elimination from the somatic pool. Prior to the establishment of a robust DNA replication checkpoint, the propensity of syncytial embryos to the accumulation of DNA lesion is thus mitigated through a Chk2-mediated nuclear fallout process that relies on the nuclear retention of essential mRNAs. These results reveal a novel role of posttranscriptional transport routes in ensuring genome integrity surveillance during embryogenesis.

Blythe and Wieschaus provided further evidence of the interplay between zygotic transcription and replication checkpoint activation. These authors found that checkpoint activation correlates with the amount of DNA engaged by Pol II, independently of the N/C ratio [80] [81]. Through ChIP-seq analyses, then found that Pol II distributions are not severely impaired in *grp* mutants, with widespread genomic occupancy at NC12, NC13 and NC14. This result suggests that the transcriptional machinery is in place independently of the functionality of the replication checkpoint. To investigate the links between checkpoint activation and ZGA at the molecular level, the authors took advantage of RPA70, an important effector of the DDR. RPA70 binds stress-induced ssDNA produced when replication is stalled, leading to ATR (*mei-41*) recruitment and checkpoint activation. Assessment of RPA70 occupancy through fluorescent microscopy and ChIP-seq revealed a strong correlation with Pol II binding sites, consistent with the hypothesis that Pol II engagement activates the checkpoint at the MBT. In *zld*⁻ mutants, ChIP-seq analyses

revealed altered RPA70 occupancy at *zld*-dependent promoters. This result suggests that transcription contributes to checkpoint activation.

To test this hypothesis, the authors attempted to rescue the lethality phenotype associated to mutations in the DDR factor ATR by altering ZGA through different approaches. They showed that most embryos from double *zld mei-41* (ATR) mutants complete cleavage cycles and that many escape the mitotic catastrophe that characterizes *mei-41* mutants. In addition to the *zld*- model, they used an heterozygous deficiency in the transcriptional activator Trithorax-like/GAGA (*Trl*), associated to defects in the genomic recruitment of poised Pol II, to interfere with ZGA. They found that *mei-41*; *Df(3L)ED4545/+ (Df(trl)/+)* embryos complete cleavage cycles without a mitotic catastrophe after a slightly lengthened NC13 and eventually yield hatching larvae. In addition, heterozygosity of the *cyclin B* gene (*Df(cycB)/+*), which lengthens NC13 time, effectively suppressed the mitotic catastrophe of *mei-41* mutants. Together, these rescue experiments show that reducing the source of replication stalling by interfering with transcription (in *zld* and *Trl* mutants) and providing more time to allow DNA replication (*in cycB mutants*) can bypass the MBT requirement for a functional replication checkpoint. In conjunction with evidence of RPA70 colocalization with the transcriptional machinery, these results strongly suggest that the replication checkpoint is activated in response to ZGA.

In brief, recent studies have used loss-of-function analyses to reveal the contributions of ZGA to key facets of embryonic development. Mutants exhibiting a premature ZGA undergo early cellularization, promptly enact maternal clearance and acquire a precocious requirement for effectors of the DDR. These effects are independent of the N/C ratio and can be rescued by modulating zygotic transcription. Moreover, effectors of the DDR are recruited to chromatin at Pol II occupied loci after ZGA and interfering with the scope of zygotic transcription can bypass the requirement for a functional checkpoint. In addition, DNA damage elicits a Chk2-dependent clearance of damaged nuclei in syncytial embryos through the nuclear retention of essential mRNAs, providing an elegant mechanism to safeguard genome integrity prior to the establishment of a robust DNA replication checkpoint.

Emerging properties of zygotic transcription

While zygotic transcription drives key events of embryonic development, the properties and dynamics of the emerging transcriptional process itself have been challenging to study. The advent of approaches enabling RNA labelling *in vivo* and in real time has provided a clearer picture of zygotic transcription. Originally developed by Singer and colleagues in yeast, the MS2 system takes advantage of the strong affinity of bacteriophage coat proteins (eg. MS2, PP7) for specific RNA stem-loops [82]. For imaging purposes, MS2 phage coat protein fused to a fluorescent reporter (eg. GFP, mCherry) are co-expressed in a transgenic organism along with an RNA fusion that encompasses the target transcript and MS2 stem-loop repeats (**Figure 3.7**). Stable tethering of the coat fusion protein allows for durable tracking of the target RNA, which can be expressed in its endogenous regulatory context to recapitulate physiological properties [83]. Over the last decade, several groups have harnessed the power of the MS2 imaging system to study RNA dynamics and localization during *Drosophila* development [84, 85] [86] [87, 88] [89]. This system has notably been used to investigate the dynamics of transcriptional bursting, calculate Pol II elongation rate at the MBT and monitor post-mitotic transcriptional reactivation.

Quantitative RNA detection methods suggest that transcriptional bursting is a key property of gene expression in diverse systems [90] [91]. The term "bursting" refers to the episodic, discontinuous emergence of nascent transcripts at Pol II-bound loci. To investigate the links between enhancer control and transcriptional bursting at the MBT, *Fukaya et al.* placed well characterized enhancers upstream and downstream of reporter genes flanked by *MS2* and *PP7* stem loops [92]. They performed live-embryo imaging of the *MS2-yellow* reporter containing different *snail* (*sna*) enhancers of varying strength, along with its proximal promoter in different configurations. They observed major differences in bursting frequencies produced by the *sna* primary and shadow enhancer, which were correlated to the discrepancy in total RNA outputs. This analysis was extended to the *rhomboid* (*rho*), *Krüppel*; (*Kr*) and *Abdominal-B* (*Abd-B*) enhancers. By testing a set of conditions, *Fukaya et al.* showed that differential core promoter, distal enhancer and anteroposterior gradient positioning all affect bursting frequency, in line with discrepancies in total RNA outputs. Thus, *Fukaya et al.* identify the regulation of transcriptional bursting frequency as a key determinant of developmental gene activity at the MBT.

Puzzling disparities have long prevailed between reported rates of Pol II elongation (1.1-1.5 kb/min) and robust detection of several long *de novo* transcripts before the MBT [14, 93]. Indeed, established elongation rates can't account for the zygotic transcription of the 22 kb-long unit of Short gastrulation (*sog*) in NC13, when the time window permissive to transcription is narrowly restricted by a hasty interphase (10-12 min). *Fukaya et al.* solved this long-lasting paradox by revisiting Pol II elongation rates in early embryogenesis using dual-fluorescence through the MS2 imaging system [94] (**Figure 3.7A**). They measured an elongation rate of 2.4 kb/min, nearly twice that of previous estimates. This figure is compatible with endogenous *sog* transcription during the 13th interphase. In addition, they found that replacing the promoter or introducing a reporter containing an intron had little impact on elongation rate measurements, suggesting that elongation is not the rate-limiting step in transcription.

The inheritance of transcriptional states from mother to daughter cells, termed transcriptional memory, has been documented in the amoeba *Dictyostelium* [95] [96]. In a recent study, *Ferraro et al.* monitored post-mitotic transcriptional reactivation of stochastically expressed transgenes using the MS2 imaging system [97]. This work provided the first evidence that transcriptional memory prevails at the massive wave of zygotic expression between the 13th and the 14th division. The authors used sensitized transgenes exhibiting patterns of sporadic expression to individually image the behaviour of single cell lineages (**Figure 3.7B**). Daughter cells derived from nuclei that expressed the transgene during NC13, called memory mothers, were four times more likely to show early reactivation during NC14 interphase than daughters arising from non-memory mothers. Quantitative analyses of average fluorescence intensities revealed that memory nuclei produce, on average, two-fold more total mRNA than non-memory nuclei during NC14. These results provide strong evidence that transcriptional memory prevails during *Drosophila* MBT and impacts total RNA output, likely through modifications incurred at the level of nucleosomes, bound TFs or histone modifications following a first round of transcription. *Ferraro et al.* envision this emerging property of early transcription as a mechanism of developmental homeostasis, which could help ensure that cells retain the properties of their progenitors.

In brief, the recent deployment of *in vivo* imaging to document the transcriptional process in real time has revealed new insights into the dynamics of zygotic transcription. *In vivo* imaging has established the notion of transcriptional bursting, and shown that total gene-specific outputs at ZGA reflect the frequency of transcriptional bursts. It has enabled the revision of Pol II elongation rate at the MBT, reconciling the expression of lengthy genes such as *sog* with a short NC13 interphase. Quantitative imaging also showed that transcriptional memory prevails in *Drosophila* embryogenesis, promoting the rapid post-mitotic re-activation of sequences expressed during NC13, which likely contributes to developmental homeostasis.

Conclusion

In this review, we aimed to provide an overview of recent findings relevant to the transcriptome dynamics of early *Drosophila* development. Maternal control is essential in syncytial-stage embryos to sustain fast-paced proliferation in absence of a sizeable transcriptional output. Thus, we discussed the processes of RNA maternal deposition, localization and targeted clearance. Indeed, *Drosophila* embryos host a wealth of complex posttranscriptional regulatory processes. The spatiotemporal dosage of RBPs such as SMG reflect translational fine-tuning which, in turn, modulates the dynamics of hundreds of maternal mRNAs. RNA localization is highly prevalent in the large syncytial embryo and likely plays key roles in orchestrating the developmental program. Indeed, protein-coding transcripts adopt a large diversity of spatial distributions in early embryos, including subembryonic and exclusionary patterns, asymmetric anteroposterior localization and more resolved patterns such as membrane, microtubule or mitotic apparatus associations [24]. In many cases, the functional relevance of these mRNA localization events remains untapped, and could be addressed through loss-of-function analyses. Indeed, RNA localization largely depends on the interactions of *cis*-acting elements and *trans*-acting factors, typically RBPs, providing a functional interface that can be disrupted to assess its roles.

Maternal control is gradually met with an increasing contribution of the zygotic genome, as it progressively acquires transcriptional competence. The mechanisms accounting for ZGA onset are multifaceted and their underpinnings remain unclear. As the N/C ratio increases, a maternal factor responsible for transcriptional quiescence could be diluted against the mass of DNA, loosening the efficiency of the repression. Studies in *Xenopus* have identified maternal histone as putative transcriptional repressors [98]. In line with this finding, a recently identified linker H1 variant, dBigH1, has been associated to transcriptional quiescence. Indeed, dBigH1 is cleared out prior to the MBT and *big1* mutants show signs of disorganized chromatin and early zygotic transcription. Nevertheless, we know that the repressor titration model is not sufficient to account for the transcriptional silence of the zygotic genome in *Drosophila*. Indeed, most zygotic transcripts display similar expression kinetics in haploid and diploid embryos, which present very different N/C ratios [54]. This observation points to the molecular clock model, which proposes that egg activation sets a timer in motion to eventually trigger ZGA. One possible

interpretation is that maternally contributed transcripts encoding pioneer TFs such as ZLD require time to be translated and accumulate sufficiently before pioneer activity has reached a level amenable to widespread ZGA. Together, a molecular clock set at egg activation and the rapid increase in nucleocytoplasmic ratio likely converge to exert changes in the structure of zygotic chromatin. Specific loci may exhibit enhanced sensitivity to pioneer TF activity. In addition, titration of the maternal repressor may not occur at an homogeneous rate across the genome. Such effects could account for the gradual nature of ZGA and explain why only a subset of zygotic genes display N/C-dependent expression dynamics.

Independently of its underlying mechanisms, once it's been triggered, zygotic transcription contributes to shaping a complex genome topology before the MBT, which largely remains in place throughout the life of the fly. After NC10, the growing population of transcripts produced by cortical nuclei play key roles in driving the course of development through the MBT. The transcriptional process itself exposes ssDNA and triggers the activation of the DNA replication checkpoint, possibly through the formation of stalled replication forks. Cellularization onset is linked to ZGA timing and it coincides with the loss of mitotic synchrony and the expression of additional zygotic genes. Real-time imaging has revealed that zygotic transcription proceeds as bursts, with the frequency of bursting events linked to its total RNA output. An example of transcriptional memory in *Drosophila*, the preferential post-mitotic reactivation of loci transcribed at NC13 has been demonstrated through this approach. In addition, embryonic Pol II elongation rates have been revised via the development of a dual fluorescence system. The deployment of real-time, *in vivo* RNA labelling to study transcription in *Drosophila* is still very recent. Future applications will likely contribute to clarify how ZGA is triggered. Indeed, important insights could be revealed by monitoring transcription in mutants of specific chromatin components, such as *bigH1*, or pioneer TFs, such as *zld*. This technology can notably reveal transcriptional dynamics at the single cell level, a sizeable advantage when investigating heterogeneous and multifaceted responses such as ZGA.

Figure Legends

Figure 3.1. Overview of morphologic and transcriptomic features of early embryonic development

Comparative timescale of Bownes' stages, nuclear divisions and histological organization as of function of time after egg activation. The nucleocytoplasmic (N:C) ratio increases rapidly during the cleavage cycles, driven by fast-paced mitotic cycles in absence of cytokinesis and growth of the syncytium. The transcriptomic landscape that prevails during the first embryonic cycles reflects maternally-deposited transcripts encoded in ovarian nurse cells and deposited in the oocyte before egg activation. Zygotic genome activation (ZGA) begins in late preblastoderm embryos and leads to the progressive accumulation of zygotic transcripts. Concomitantly, a large fraction of maternally deposited RNAs undergo targeted degradation.

Figure 3.2. Alternative degradation profiles of maternally-deposited transcripts

The clearance of maternally-deposited transcripts can proceed through a strictly maternal pathway, a strictly zygotic pathway or a combination of both maternal and zygotic effectors. Transcripts strictly targeted by the maternal pathway, such as *nanos* (*nos*) display identical dynamics in fertilized and activated eggs. The maternal RNA-binding proteins Smaug, Pumilio, Brat and/or ME31B selectively interact with these RNAs through a consensus motif and recruit the CCR4/POP2/NOT deadenylase complex to initiate their degradation. Transcripts targeted through the zygotic pathway include *bicoid* (*bcd*) and their degradation depends on the ZGA. Transcripts targeted through both maternal and zygotic degradation pathways include *Hsp83* and their clearance relies on the activity of both maternally deposited and zygotically encoded factors.

Figure 3.3. Developmental regulation of the PNG kinase coupled to cell cycle progression

In mature oocytes arrested at metaphase of meiosis I, cyclinB-CDK1 dependent phospohorylation of GNU exerts an inhibition of PNG complex assembly and activation. The completion of meiosis that follows egg activation results in CDK1 inactivation, prompting the dephosphorylation of GNU. In meiosis II, the accumulation of dephosphorylated GNU proteins leads to the spontaneous assembly of an active PNG kinase complex, consisting of GNU, PNG

and PLU. The PNG kinase regulates the translation of hundreds of maternal mRNAs, including *cycB* and and *smg*. In addition, GNU protein degradation is promoted by PNG activity, enacting a negative feedback loop that restricts the activity of the complex to the temporal context of early embryonic development.

Figure 3.4. Models of zygotic genome activation

Several models have been proposed to contribute to ZGA. Each is supported by a set of empirical evidence (outlined under each cartoon) and these different propositions likely contribute synergistically to the emergence of zygotic transcription. Model I posits that early zygotic transcription is restricted due to the short duration of interphases during early embryogenesis, effectively preventing the complete transcription of long genes. Model II stipulates that zygotic transcription is prohibited during early embryogenesis due to the abundance of a maternally-inherited transcriptional repressor. The titration of this repressor against the increasing mass of zygotic nuclei would progressively lead to transcriptional competence. Model III proposes that egg activation sets in a molecular clock, which times key events of the MZT, including ZGA. Model IV postulates that chromatin is kept is a state that precludes transcription during early embryogenesis and is progressively remodeled through active changes in its composition to promote gene expression.

Figure 3.5. Developmental regulation of chromatin landscapes and genomic architecture

The histone H1 variant BigH1 is a constitutive chromatin component in the germline and in fertilized embryos until the MBT. The acetylation marks H4K8ac, H3K18ac and H3K27ac appear at the ZGA and scale up with the prevalence of zygotic transcription in syncytial embryos. By contrast, H3K9ac and H3K4me1/3 emerge around the MBT. Nucleosome-free regions (NFR) are found upstream of maternally-deposited genes throughout embryonic development but their appearance upstream of zygotic genes is concomitant with their transcription. Endowed with pioneer transcription factor activity, the binding of Zelda (ZLD) to its consensus sequence leads to local nucleosome depletion around NC10, exposing surrounding enhancers to promote the recruitment of patterning transcription factors by NC14. Hi-C data shows that chromatin is poorly organized prior to NC10. The emergence of intricate long-range interactions emerges after ZGA, by NC14.

Figure 3.6. Zygotic genome activation times the onset of the DNA replication checkpoint, maternal clearance and cellularization

X161 mutants display premature zygotic transcription due to a point mutation in the *RNPII215* gene, which encodes a subunit of Pol II. This disruption leads to premature cellularization, which is rescued by altering transcription in X161 *zld* double mutants. Precocious transcription also leads to an early deployment of maternal clearance, exemplified by the premature degradation of the maternal transcripts *string*, *twine* and *smaug*. In addition, early ZGA leads to a premature activation of the replication checkpoint, as inferred from a precipitate requirement for Chk1/grp activity. None of these phenotypes is rescued in haploid X161 embryos, which exhibit a decreased N/C ratio, meaning that the N/C ratio does not act upstream of transcription activation in the regulation of cellularization, maternal clearance and checkpoint activation.

Figure 3.7. Real-time imaging of zygotic transcription enables the determination of Pol II elongation rate and demonstrates transcriptional memory at NC14

(A) *Fukaya et al.* used a dual fluorescence approach involving the MS2 system to measure Pol II elongation rates at NC13. They integrated a construct encompassing a *lacZ* reporter flanked by 24 *MS2* repeats at its 5'-end and 24 *PP7* repeats at the 3'end. Its expression in conjunction with the MCP-GFP and mCherry-PP7 coat proteins leads to the emission of dual green and red fluorescence. By measuring the delay between the emission of the green and red signals at the single molecule level, the time required to transcribe the intervening *lacZ* sequence can be determined. This value is then used to calculate the elongation rate of Pol II. **(B)** *Ferraro et al.* provided evidence of transcriptional memory by monitoring post-mitotic reactivation. Sensitized transgenes were used to obtain sporadic expression of the *yellow* reporter downstream of 24 *MS2* repeats. They tracked expression during NC13 and NC14 in a lineage-specific manner. After mitosis, the authors found that the daughter of nuclei having expressed the reporter during NC13 where four times more likely to re-express it rapidly. This result indicates that transcription prior to mitosis increases the chance and rapidity of re-expression across cell generations, a phenomenon termed transcriptional memory.

Acknowledgments

This work was supported by grants from the Canadian Institutes of Health Research (CIHR; MOP-111161) and Natural Sciences and Engineering Research Council of Canada (NSERC; Discovery Grant #386644), as well as a Junior 2 research scholar of the Fonds de Recherche du Québec - Santé (FRQS) to E.L. F.A.L. is funded by a Frederick Banting and Charles Best Canada Graduate Scholarship from CIHR.

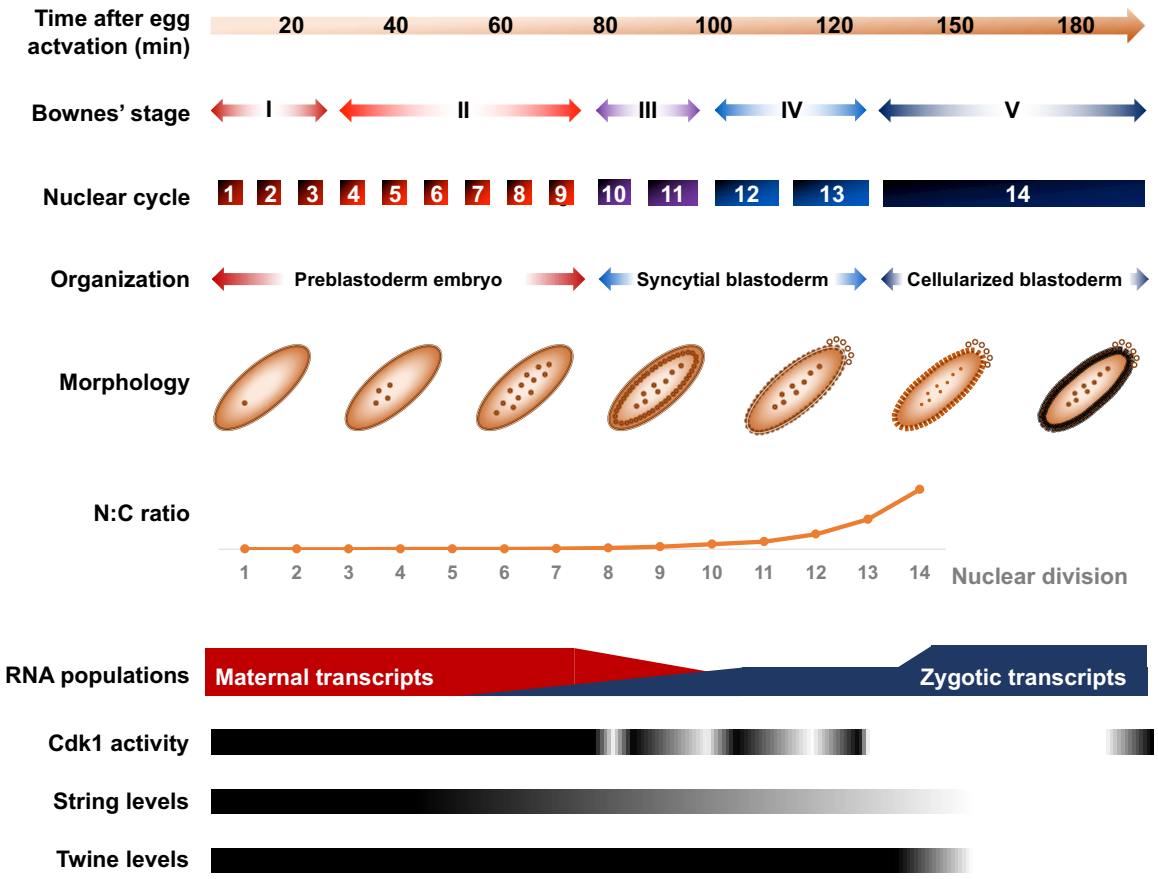


Figure 3. 1. Overview of morphologic and transcriptomic features of early embryonic development

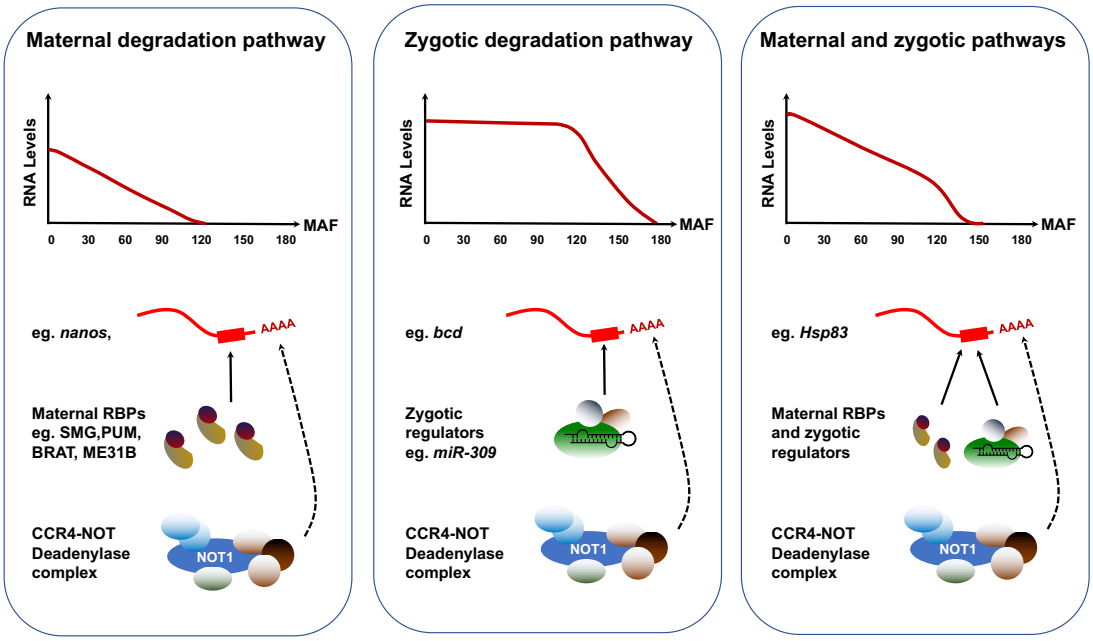


Figure 3. 2. Alternative degradation profiles of maternally-deposited transcripts

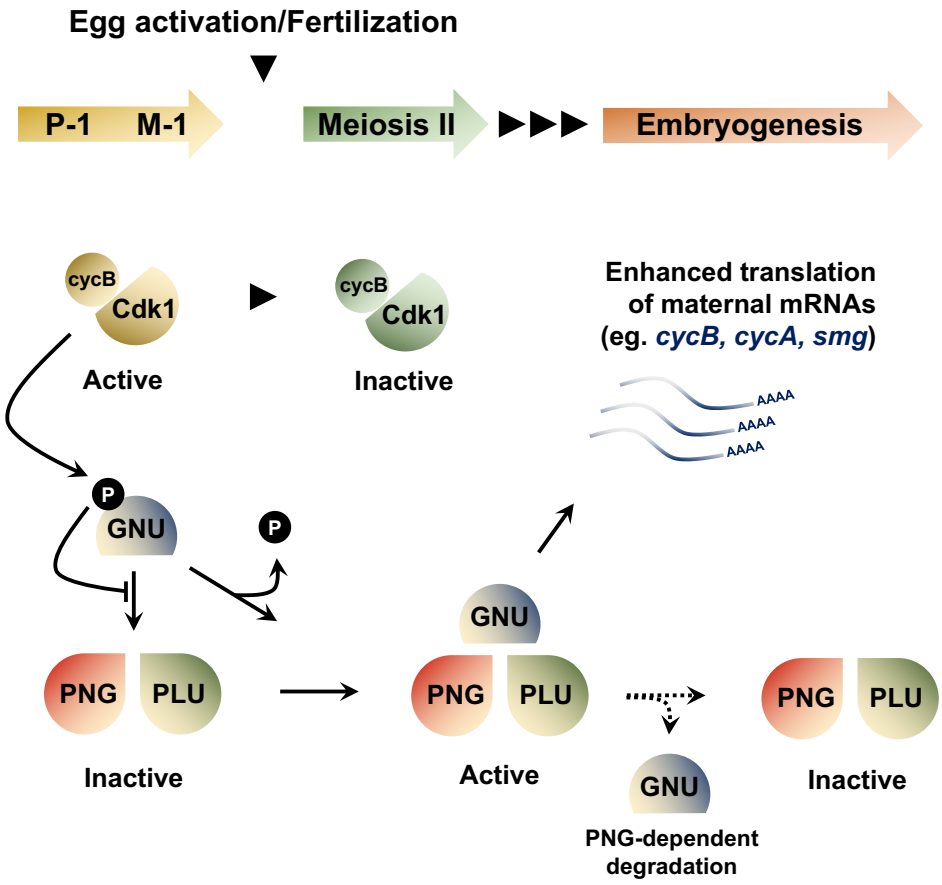


Figure 3. 3. Developmental regulation of the PNG kinase coupled to cell cycle progression

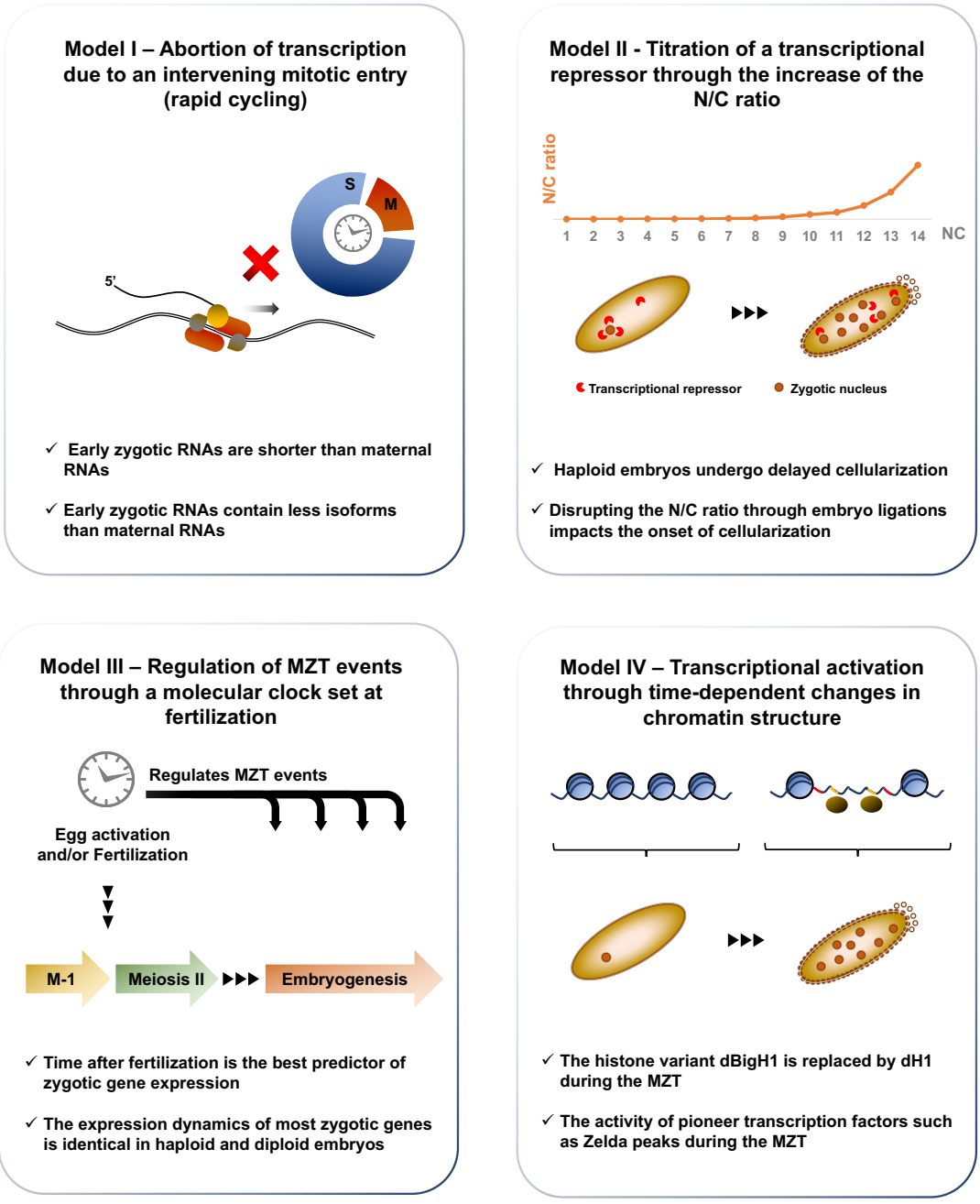


Figure 3. 4. Models of zygotic genome activation

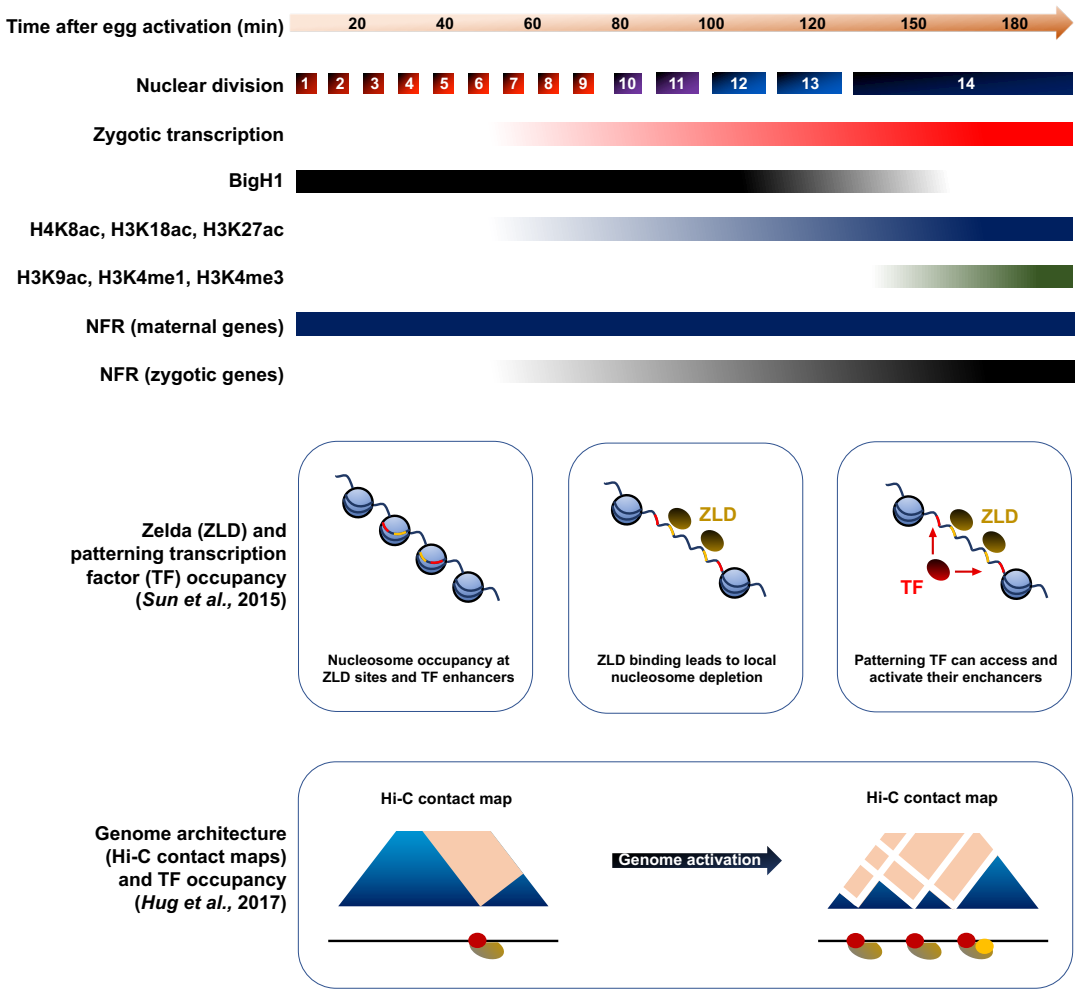


Figure 3. 5. Developmental regulation of chromatin landscapes and genomic architecture

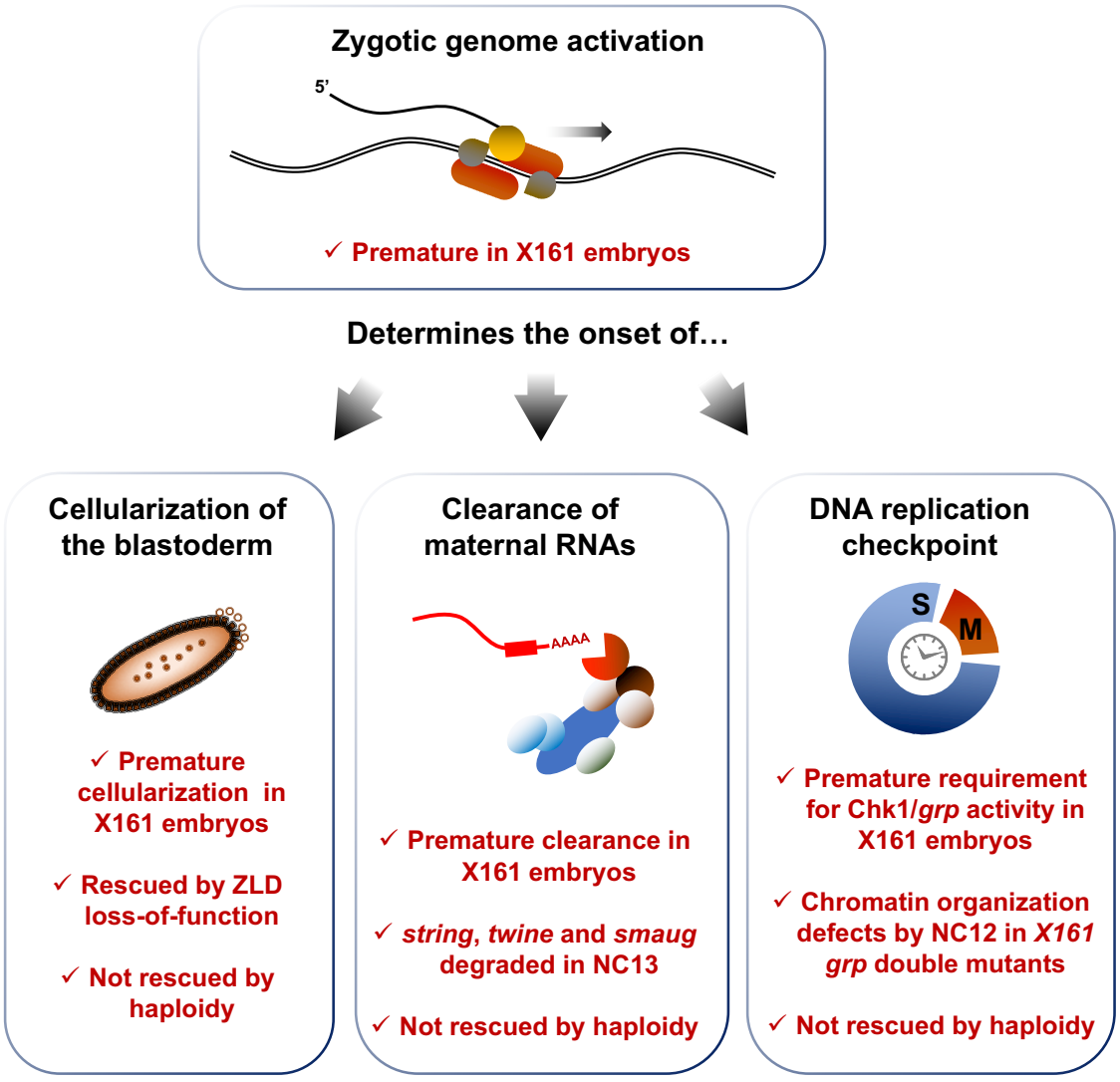


Figure 3. 6. Zygotic genome activation times the onset of the DNA replication checkpoint, maternal clearance and cellularization

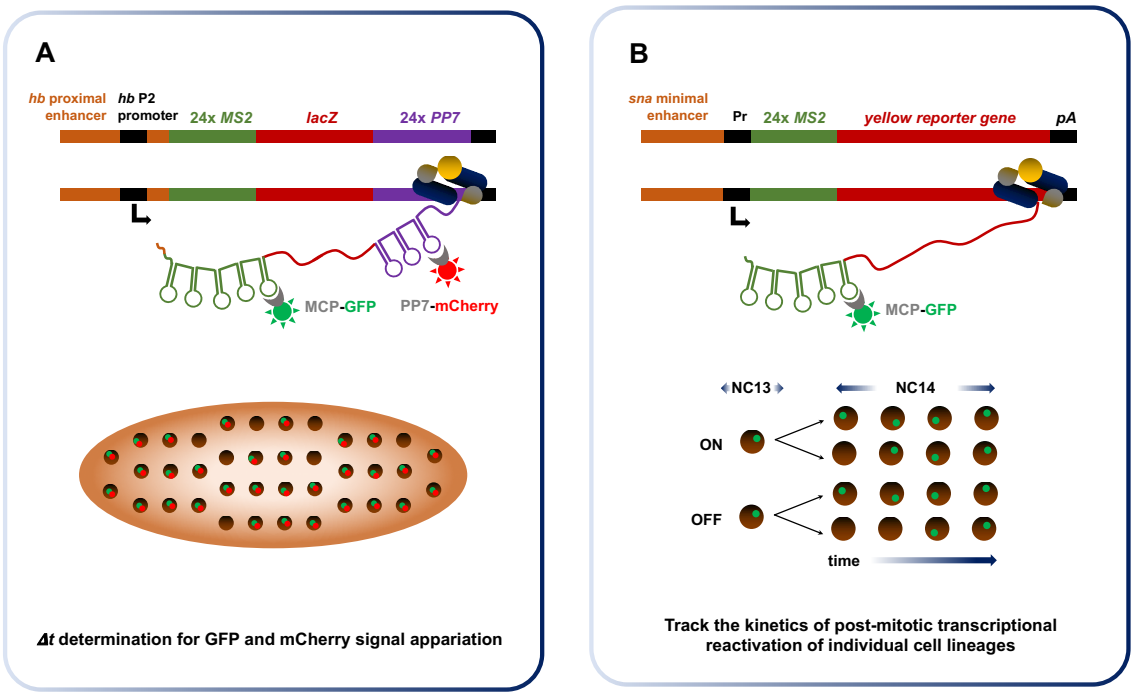


Figure 3. 7. Real-time imaging of zygotic transcription enables the determination of Pol II elongation rate and demonstrates transcriptional memory at NC14

References

[1] Von Stetina JR, Orr-Weaver TL. Developmental control of oocyte maturation and egg activation in metazoan models. Cold Spring Harb Perspect Biol. 2011;3:a005553.
[2] Farrell JA, O'Farrell PH. From egg to gastrula: how the cell cycle is remodeled during the Drosophila mid-blastula transition. Annu Rev Genet. 2014;48:269-94.
[3] Shermoen AW, McCleland ML, O'Farrell PH. Developmental control of late replication and S phase length. Curr Biol. 2010;20:2067-77.
[4] Pritchard DK, Schubiger G. Activation of transcription in Drosophila embryos is a gradual process mediated by the nucleocytoplasmic ratio. Genes Dev. 1996;10:1131-42.
[5] Heyn P, Kircher M, Dahl A, Kelso J, Tomancak P, Kalinka AT, et al. The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell Rep. 2014;6:285-92.
[6] Heyn P, Neugebauer KM. Purification of Zygotically Transcribed RNA through Metabolic Labeling of Early Zebrafish Embryos. Methods Mol Biol. 2017;1605:121-31.
[7] Palfy M, Joseph SR, Vastenhouw NL. The timing of zygotic genome activation. Curr Opin Genet Dev. 2017;43:53-60.
[8] Bashirullah A, Halsell SR, Cooperstock RL, Kloc M, Karaiskakis A, Fisher WW, et al. Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in Drosophila melanogaster. EMBO J. 1999;18:2610-20.
[9] Tadros W, Lipshitz HD. The maternal-to-zygotic transition: a play in two acts. Development. 2009;136:3033-42.
[10] Riparbelli MG, Gottardo M, Callaini G. Parthenogenesis in Insects: The Centriole Renaissance. Results Probl Cell Differ. 2017;63:435-79.
[11] van der Kooi CJ, Schwander T. Parthenogenesis: birth of a new lineage or reproductive accident? Curr Biol. 2015;25:R659-61.
[12] Rabinowitz M. Studies on the cytology and early embryology of the egg of Drosophila melanogaster. Philadelphia,: Press of the Wistar institute of anatomy and biology; 1941.
[13] Deneke VE, Melbinger A, Vergassola M, Di Talia S. Waves of Cdk1 Activity in S Phase Synchronize the Cell Cycle in Drosophila Embryos. Dev Cell. 2016;38:399-412.

[14] Shermoen AW, O'Farrell PH. Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. Cell. 1991;67:303-10.
[15] Langley AR, Smith JC, Stemple DL, Harvey SA. New insights into the maternal to zygotic transition. Development. 2014;141:3834-41.
[16] Fogarty P, Campbell SD, Abu-Shumays R, Phalle BS, Yu KR, Uy GL, et al. The Drosophila grapes gene is related to checkpoint gene chk1/rad27 and is required for late syncytial division fidelity. Curr Biol. 1997;7:418-26.
[17] Sibon OC, Stevenson VA, Theurkauf WE. DNA-replication checkpoint control at the Drosophila midblastula transition. Nature. 1997;388:93-7.
[18] Iampietro C, Bergalet J, Wang X, Cody NA, Chin A, Lefebvre FA, et al. Developmentally regulated elimination of damaged nuclei involves a Chk2-dependent mechanism of mRNA nuclear retention. Dev Cell. 2014;29:468-81.
[19] O'Farrell PH, Stumpff J, Su TT. Embryonic cleavage cycles: how is a mouse like a fly? Curr Biol. 2004;14:R35-45.
[20] Stack JH, Newport JW. Developmentally regulated activation of apoptosis early in Xenopus gastrulation results in cyclin A degradation during interphase of the cell cycle. Development. 1997;124:3185-95.
[21] Marlow FL. Maternal Control of Development in Vertebrates: My Mother Made Me Do It! San Rafael (CA)2010.
[22] Tadros W, Lipshitz HD. Setting the stage for development: mRNA translation and stability during oocyte maturation and egg activation in Drosophila. Dev Dyn. 2005;232:593-608.
[23] Ding D, Lipshitz HD. A molecular screen for polar-localised maternal RNAs in the early embryo of Drosophila. Zygote. 1993;1:257-71.
[24] Lecuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, et al. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell. 2007;131:174-87.
[25] Lefebvre FA, Benoit Bouvrette LP, Bergalet J, Lecuyer E. Biochemical Fractionation of Time-Resolved Drosophila Embryos Reveals Similar Transcriptomic Alterations in Replication Checkpoint and Histone mRNA Processing Mutants. J Mol Biol. 2017.

[26] Lefebvre FA, Bouvrette LPB, Bergalet J, Lecuyer E. Data for the generation of RNA spatiotemporal distributions and interpretation of Chk1 and SLBP protein depletion phenotypes during Drosophila embryogenesis. Data Brief. 2017;13:28-31.
[27] Johnstone O, Lasko P. Translational regulation and RNA localization in Drosophila oocytes and embryos. Annu Rev Genet. 2001;35:365-406.
[28] Martin KC, Ephrussi A. mRNA localization: gene expression in the spatial dimension. Cell. 2009;136:719-30.
[29] Roth S, Lynch JA. Symmetry breaking during Drosophila oogenesis. Cold Spring Harb Perspect Biol. 2009;1:a001891.
[30] Jambor H, Surendranath V, Kalinka AT, Mejstrik P, Saalfeld S, Tomancak P. Systematic imaging reveals features and changing localization of mRNAs in Drosophila development. Elife. 2015;4.
[31] De Renzis S, Elemento O, Tavazoie S, Wieschaus EF. Unmasking Activation of the Zygotic Genome Using Chromosomal Deletions in the Drosophila Embryo. PLoS Biol. 2007;5:e117.
[32] Semotok JL, Cooperstock RL, Pinder BD, Vari HK, Lipshitz HD, Smibert CA. Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early Drosophila embryo. Curr Biol. 2005;15:284-94.
[33] Benoit B, He CH, Zhang F, Votruba SM, Tadros W, Westwood JT, et al. An essential role for the RNA-binding protein Smaug during the Drosophila maternal-to-zygotic transition. Development. 2009;136:923-32.
[34] Semotok JL, Luo H, Cooperstock RL, Karaiskakis A, Vari HK, Smibert CA, et al. Drosophila maternal Hsp83 mRNA destabilization is directed by multiple SMAUG recognition elements in the open reading frame. Mol Cell Biol. 2008;28:6757-72.
[35] Laver JD, Li X, Ray D, Cook KB, Hahn NA, Nabeel-Shah S, et al. Brain tumor is a sequence-specific RNA-binding protein that directs maternal mRNA clearance during the Drosophila maternal-to-zygotic transition. Genome Biol. 2015;16:94.
[36] Wang M, Ly M, Lugowski A, Laver JD, Lipshitz HD, Smibert CA, et al. ME31B globally represses maternal mRNAs by two distinct mechanisms during the Drosophila maternal-to-zygotic transition. Elife. 2017;6.
[37] Giraldez AJ. microRNAs, the cell's Nepenthe: clearing the past during the maternal-to-zygotic transition and cellular reprogramming. Curr Opin Genet Dev. 2010;20:369-75.

[38] Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, et al. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science. 2006;312:75-9.
[39] Luo H, Li X, Claycomb JM, Lipshitz HD. The Smaug RNA-Binding Protein Is Essential for microRNA Synthesis During the Drosophila Maternal-to-zygotic Transition. G3 (Bethesda). 2016.
[40] Bushati N, Stark A, Brennecke J, Cohen SM. Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in Drosophila. Curr Biol. 2008;18:501-6.
[41] Tadros W, Goldman AL, Babak T, Menzies F, Vardy L, Orr-Weaver T, et al. SMAUG is a major regulator of maternal mRNA destabilization in Drosophila and its translation is activated by the PAN GU kinase. Dev Cell. 2007;12:143-55.
[42] Tay J, Hodgman R, Richter JD. The control of cyclin B1 mRNA translation during mouse oocyte maturation. Dev Biol. 2000;221:1-9.
[43] Vardy L, Orr-Weaver TL. The Drosophila PNG kinase complex regulates the translation of cyclin B. Dev Cell. 2007;12:157-66.
[44] Hara M, Petrova B, Orr-Weaver TL. Control of PNG kinase, a key regulator of mRNA translation, is coupled to meiosis completion at egg activation. Elife. 2017;6.
[45] Kronja I, Yuan B, Eichhorn SW, Dzeyk K, Krijgsveld J, Bartel DP, et al. Widespread changes in the posttranscriptional landscape at the Drosophila oocyte-to-embryo transition. Cell Rep. 2014;7:1495-508.
[46] Almouzni G, Wolffe AP. Constraints on transcriptional activator function contribute to transcriptional quiescence during early Xenopus embryogenesis. EMBO J. 1995;14:1752-65.
[47] Gilbert SF. Developmental biology. 10th ed. Sunderland, Mass.: Sinauer Associates; 2014.
[48] Ali-Murthy Z, Lott SE, Eisen MB, Kornberg TB. An essential role for zygotic expression in the pre-cellular Drosophila embryo. PLoS Genet. 2013;9:e1003428.
[49] Kimelman D, Kirschner M, Scherson T. The events of the midblastula transition in Xenopus are regulated by changes in the cell cycle. Cell. 1987;48:399-407.
[50] Veenstra GJ, Destree OH, Wolffe AP. Translation of maternal TATA-binding protein mRNA potentiates basal but not activated transcription in Xenopus embryos at the midblastula transition. Mol Cell Biol. 1999;19:7972-82.
[51] Collart C, Allen GE, Bradshaw CR, Smith JC, Zegerman P. Titration of four replication factors is essential for the Xenopus laevis midblastula transition. Science. 2013;341:893-6.

[52] Newport J, Kirschner M. A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell. 1982;30:675-86.
[53] Edgar BA, Kiehle CP, Schubiger G. Cell cycle control by the nucleo-cytoplasmic ratio in early Drosophila development. Cell. 1986;44:365-72.
[54] Lu X, Li JM, Elemento O, Tavazoie S, Wieschaus EF. Coupling of zygotic transcription to mitotic control at the Drosophila mid-blastula transition. Development. 2009;136:2101-10.
[55] Clarke HJ, McLay DW, Mohamed OA. Linker histone transitions during mammalian oogenesis and embryogenesis. Dev Genet. 1998;22:17-30.
[56] Saeki H, Ohsumi K, Aihara H, Ito T, Hirose S, Ura K, et al. Linker histone variants control chromatin dynamics during early embryogenesis. Proc Natl Acad Sci U S A. 2005;102:5697-702.
[57] Perez-Montero S, Carbonell A, Moran T, Vaquero A, Azorin F. The embryonic linker histone H1 variant of Drosophila, dBigH1, regulates zygotic genome activation. Dev Cell. 2013;26:578-90.
[58] Jenuwein T, Allis CD. Translating the histone code. Science. 2001;293:1074-80.
[59] Li XY, Harrison MM, Villalta JE, Kaplan T, Eisen MB. Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. Elife. 2014;3.
[60] Gaertner B, Johnston J, Chen K, Wallaschek N, Paulson A, Garruss AS, et al. Poised RNA polymerase II changes over developmental time and prepares genes for future expression. Cell Rep. 2012;2:1670-83.
[61] Nien CY, Liang HL, Butcher S, Sun Y, Fu S, Gocha T, et al. Temporal coordination of gene networks by Zelda in the early Drosophila embryo. PLoS Genet. 2011;7:e1002339.
[62] Harrison MM, Li XY, Kaplan T, Botchan MR, Eisen MB. Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. PLoS Genet. 2011;7:e1002266.
[63] Liang HL, Nien CY, Liu HY, Metzstein MM, Kirov N, Rushlow C. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature. 2008;456:400-3.
[64] Sun Y, Nien CY, Chen K, Liu HY, Johnston J, Zeitlinger J, et al. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. Genome Res. 2015;25:1703-14.

[65] Lee MT, Bonneau AR, Takacs CM, Bazzini AA, DiVito KR, Fleming ES, et al. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature. 2013;503:360-4.
[66] Leichsenring M, Maes J, Mossner R, Driever W, Onichtchouk D. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. Science. 2013;341:1005-9.
[67] Iwafuchi-Doi M, Zaret KS. Cell fate control by pioneer transcription factors. Development. 2016;143:1833-7.
[68] Zaret KS, Mango SE. Pioneer transcription factors, chromatin dynamics, and cell fate control. Curr Opin Genet Dev. 2016;37:76-81.
[69] Misteli T. Beyond the sequence: cellular organization of genome function. Cell. 2007;128:787-800.
[70] Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485:376-80.
[71] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326:289-93.
[72] van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, et al. Hi-C: a method to study the three-dimensional architecture of genomes. J Vis Exp. 2010.
[73] Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, et al. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell. 2012;148:458-72.
[74] Tolhuis B, Blom M, van Lohuizen M. Chromosome conformation capture on chip in single Drosophila melanogaster tissues. Methods. 2012;58:231-42.
[75] Li HB. Chromosome Conformation Capture in Drosophila. Methods Mol Biol. 2016;1480:207-12.
[76] Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell. 2017;169:216-28 e19.
[77] Dai Q, Ren A, Westholm JO, Duan H, Patel DJ, Lai EC. Common and distinct DNA-binding and regulatory activities of the BEN-solo transcription factor family. Genes Dev. 2015;29:48-62.

[78] Sung HW, Spangenberg S, Vogt N, Grosshans J. Number of nuclear divisions in the Drosophila blastoderm controlled by onset of zygotic transcription. Curr Biol. 2013;23:133-8.
[79] Fogarty P, Kalpin RF, Sullivan W. The Drosophila maternal-effect mutation grapes causes a metaphase arrest at nuclear cycle 13. Development. 1994;120:2131-42.
[80] Blythe SA, Wieschaus EF. Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. Cell. 2015;160:1169-81.
[81] Laver JD, Lipshitz HD. Transcription gets to the checkpoint. Cell. 2015;160:1043-4.
[82] Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. Localization of ASH1 mRNA particles in living yeast. Mol Cell. 1998;2:437-45.
[83] Weil TT, Parton RM, Davis I. Making the message clear: visualizing mRNA localization. Trends Cell Biol. 2010;20:380-90.
[84] Forrest KM, Gavis ER. Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in Drosophila. Curr Biol. 2003;13:1159-68.
[85] Belaya K, St Johnston D. Using the mRNA-MS2/MS2CP-FP system to study mRNA transport during Drosophila oogenesis. Methods Mol Biol. 2011;714:265-83.
[86] Trovisco V, Belaya K, Nashchekin D, Irion U, Sirinakis G, Butler R, et al. bicoid mRNA localises to the Drosophila oocyte anterior by random Dynein-mediated transport and anchoring. Elife. 2016;5.
[87] Halstead JM, Lionnet T, Wilbertz JH, Wippich F, Ephrussi A, Singer RH, et al. Translation. An RNA biosensor for imaging the first round of translation from single cells to living animals. Science. 2015;347:1367-671.
[88] Halstead JM, Wilbertz JH, Wippich F, Lionnet T, Ephrussi A, Chao JA. TRICK: A Single-Molecule Method for Imaging the First Round of Translation in Living Cells and Animals. Methods Enzymol. 2016;572:123-57.
[89] Bothma JP, Garcia HG, Ng S, Perry MW, Gregor T, Levine M. Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. Elife. 2015;4.
[90] Sanchez A, Golding I. Genetic determinants and cellular constraints in noisy gene expression. Science. 2013;342:1188-93.
[91] Chong S, Chen C, Ge H, Xie XS. Mechanism of transcriptional bursting in bacteria. Cell. 2014;158:314-26.

[92] Fukaya T, Lim B, Levine M. Enhancer Control of Transcriptional Bursting. Cell. 2016;166:358-68.
[93] O'Brien T, Lis JT. Rapid changes in Drosophila transcription after an instantaneous heat shock. Mol Cell Biol. 1993;13:3456-63.
[94] Fukaya T, Lim B, Levine M. Rapid Rates of Pol II Elongation in the Drosophila Embryo. Curr Biol. 2017;27:1387-91.
[95] Muramoto T, Muller I, Thomas G, Melvin A, Chubb JR. Methylation of H3K4 Is required for inheritance of active transcriptional states. Curr Biol. 2010;20:397-406.
[96] Zhao R, Nakamura T, Fu Y, Lazar Z, Spector DL. Gene bookmarking accelerates the kinetics of post-mitotic transcriptional re-activation. Nat Cell Biol. 2011;13:1295-304.
[97] Ferraro T, Esposito E, Mancini L, Ng S, Lucas T, Coppey M, et al. Transcriptional Memory in the Drosophila Embryo. Curr Biol. 2016;26:212-8.
[98] Amodeo AA, Jukam D, Straight AF, Skotheim JM. Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition. Proc Natl Acad Sci U S A. 2015;112:E1086-95.

Chapitre 4 – Hypothèses, objectifs de travail et structure de la thèse

4.1. Objectifs et hypothèses de travail

4.1.1. Hypothèse globale de la thèse

Cette thèse fait état du développement d'une approche biochimique couplée à des analyses transcriptomiques dans l'étude systématique de la localisation subcellulaire et extracellulaire de l'ARN. Cette approche, qualifiée de CeFra-seq (Cell Fractionation, RNA-seq), permet d'estimer chez les lignées cellulaires l'abondance relative des ARNs présents au sein de diverses fractions subcellulaires et extracellulaires, soient la fraction cytosolique, la fraction cytoplasmique insoluble, la fraction des endomembranes cytoplasmiques, la fraction nucléaire et la fraction extracellulaire [1]. Une adaptation de cette approche à un système dynamique, celui de l'embryogenèse de la Drosophile, révèle les ARNs contribués maternellement (fraction syncytiale précoce), les ARNs nucléaires (fraction nucléaire du blastoderme) et les ARNs cytoplasmiques (fraction cytoplasmique du blastoderme) [2, 3]. À ce titre, l'hypothèse centrale de cette thèse pourrait être formulée comme suit : « L'approche de CeFra-seq permet de documenter la distribution spatiale des ARNs de manière systématique dans divers systèmes». Cette hypothèse est explorée par des efforts de validation de la méthode, qui démontrent qu'elle permet de récapituler les signatures asymétriques préalablement caractérisée pour certains transcrits chez l'humain (chapitre 4) et chez la Drosophile (chapitre 7). La méthode CeFra-seq d'étude de la distribution spatiale et spatiotemporelle des ARNs est ensuite appliquée à divers systèmes biologiques afin d'éclairer une série de questions fonctionnelles.

4.1.2. Hypothèse I : La localisation subcellulaire des ARNs est couplée à leur ciblage extracellulaire

En premier lieu, la notion de ciblage des ARNs aux vésicules extracellulaires (VEs) est explorée. Les VEs contiennent un assortiment spécifique d'ARNs et de protéines qui diffèrent considérablement de la population cellulaire. Cette observation, étayée par de nombreuses études, implique la prévalence de mécanismes biologiques associés au ciblage extracellulaire [4, 5] [6, 7]. Ces mécanismes demeurent mal compris. J'ai donc émis l'hypothèse que les propriétés de localisation subcellulaire des ARNs contribuent à leur ciblage extracellulaire. L'approche de CeFra-seq a été utilisée pour recenser les ARNs enrichis aux VEs et aux quatre compartiments subcellulaires, soient le cytosol, la fraction insoluble cytoplasmique, les endomembranes

cytoplasmique et le noyau. Une comparaison détaillée de ces répertoires asymétriques a ensuite été effectuée pour comparer les propriétés des ARNs enrichis aux VEs aux propriétés des ARNs enrichis aux fractions subcellulaires. Les résultats de cette analyse sont discutés dans le chapitre 5.

4.1.3. Hypothèse II : Le contenu transcriptomique des vésicules extracellulaires est conservé de la Drosophile à l'humain

Les VEs sont surtout été étudiées chez les mammifères. Pourtant, une caractérisation de leur contenu transcriptomique chez des espèces plus éloignées pourrait permettre d'établir les propriétés conservées dans le ciblage extracellulaire, contribuant à mieux définir ces mécanismes. Dans le chapitre suivant, j'émets l'hypothèse que les propriétés morphologiques et le profil transcriptomique des VEs est conservé de la Drosophile à l'humain. Afin d'évaluer cette hypothèse, je procède à une analyse de la morphologie et du contenu transcriptomique des VEs provenant de quatre modèles cellulaires, incluant deux modèles issus de la Drosophile et deux modèles issus de l'humain. Les résultats de cette analyse, qui reposent sur des résultats de microscopie électronique, de *nanoparticle tracking analysis* (NTA) et de séquençage à haut débit, confirment un niveau de conservation élevé et sont discutés dans le chapitre 6.

4.14. Hypothèse III : La déplétion de la protéine SLBP compromet l'activation transcriptionnelle du génome zygotique

Dans les deux chapitres suivants, l'approche de CeFra-seq est étendue à un nouveau système biologique, l'embryogenèse de la Drosophile, afin d'éclairer deux nouvelles questions fonctionnelles. Chez les métazoaires, le développement précoce est le théâtre d'une transformation profonde et rapide du paysage transcriptomique. En effet, les populations d'ARN exprimées par le génome maternel et déposées dans l'oocyte avant la fertilisation font l'objet d'une dégradation hautement régulée. En parallèle, le génome zygotique, issu de la fusion des pro-noyaux mâle et femelle, acquiert progressivement une capacité transcritpionnelle et génère graduellement de nouvelles populations d'ARN [8-10]. Dans ce contexte, j'ai utilisé une adaptation de l'approche CeFra-seq afin d'obtenir un portrait spatiotemporel du transcriptome pendant le développement embryonnaire. Les répertoires de transcrits asymétriques sont ensuite analysés pour évaluer une hypothèse fonctionnelle liée à l'expression des gènes d'histones. Ces

gènes hautement répétés, qui encodent les protéines situées au cœur du nucléosome, font l'objet d'une contribution maternelle et sont également rapidement exprimés par le génome zygotique [11, 12]. L'expression zygotique des gènes d'histone dépend de diverses protéines, notamment la Stem loop binding protein (SLBP), un facteur qui orchestre leur maturation, leur export nucléaire et leur traduction. Dans ce contexte, j'ai émis l'hypothèse qu'une déplétion du facteur SLBP et la perte d'expression des histones zygotiques qui lui est associée conduit à un blocage dans la progression du développement embryonnaire, se traduisant par une perte généralisée de l'activation du génome zygotique. Ainsi, les changements d'expression des divers répertoires de transcrits asymétriques identifiés par CeFra-seq ont été comparés chez les mutants SLBP. Ces résultats, qui confirment une inhibition sélective de l'activation du génome zygotique, sont présentés et discutés dans le chapitre 7.

4.1.5. Hypothèse IV : La transcription antisens des gènes d'histone contribue à la dégradation des ARNm maternels via Argonaute-2

L'étude de l'expression des gènes d'histone au moyen de l'approche CeFra-seq se poursuit dans le chapitre 8. Ici, j'identifie chez l'embryon de Drosophile des populations de transcrits non-codants exprimés dans l'orientation antisens au sein du corps génique des cinq histones canoniques. Ces transcrits font l'objet d'une caractérisation détaillée par hybridation *in situ*, buvardage de Northern et séquençage à haut débit. Un traitement à l'α-amanitine, inhibiteur de la Polymérase II, révèle que ces transcrits sont strictement produits par le génome zygotique, résultat confirmé par le séquençage d'extraits d'ovaires, qui montre que ce tissu maternel n'exprime pas les ARNs antisens d'histone. Ici l'approche CeFra-seq révèle que les ARNs antisens d'histone co-ségréguent avec leurs ARNm complémentaires chez l'embryon, mais pas chez les cellules larvaires, où ces ARNs antisens sont plutôt localisés au noyau. Cette observation suggère que les ARNs antisens pourraient former des populations d'ARN double-brin par hybridation à leur messagers complémentaires. Cette observation débouche sur l'hypothèse centrale du chapitre 8 : étant donné que les ARNs double-brin peuvent s'associer à Dicer pour générer de petits ARNs interférents, je propose que l'expression antisens des gènes d'histone est impliquée dans la dégradation régulée des ARNm d'histones déposés maternellement. Des expériences d'immunoprécipitation d'Argonaute-2, facteur important des voies de l'interférence aux ARNs, révèlent l'association de petits ARNs issus du locus des

histones chez les cellules embryonnaires S2 [13]. De plus, la déplétion du facteur Argonaute-2 chez l'embryon conduit à une dérepression des ARNm d'histones, résultat qui supporte le modèle d'une contribution de la transcription antisens dans le processus de dégradation des transcrits maternels.

Ainsi, cette thèse présente d'abord l'approche CeFra-seq, une méthode alternative aux outils d'imagerie dans l'étude systématique de la localisation subcellulaire et extracellulaire des ARNs. Cette thèse se divise ensuite en quatre chapitres qui explorent les applications de cette méthode pour répondre à quatre questions fonctionnelles chez divers systèmes biologiques.

4.2. Structure détaillée de la thèse

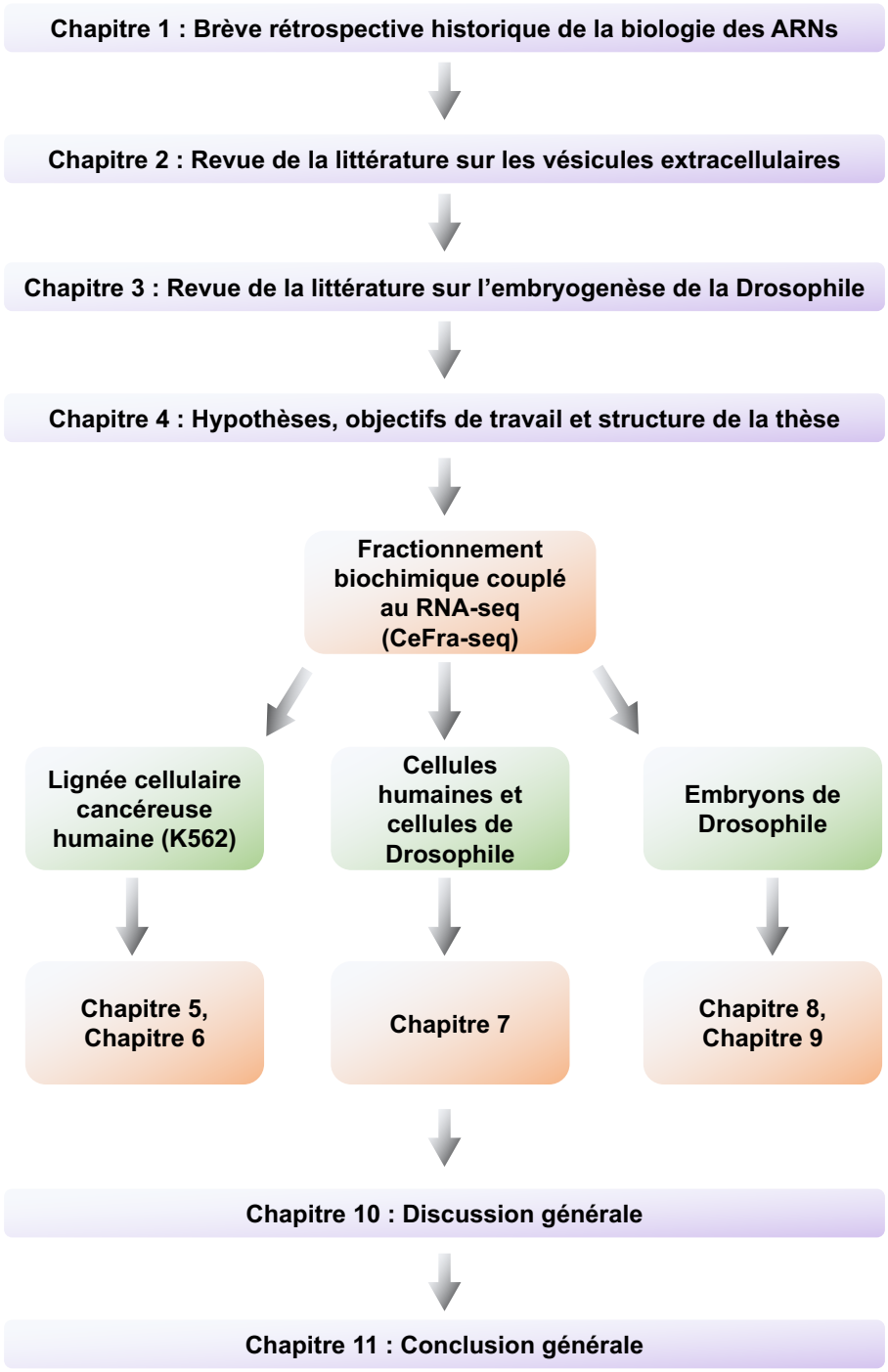


Figure 4.1 Représentation graphique de la structure détaillée de cette thèse

4.2.1. Chapitre 1

Le présent ouvrage est constitué de onze chapitres. Le premier chapitre propose une brève rétrospective historique de la biologie de l'ARN. Il retrace la découverte des acides nucléiques, l'émergence du dogme central et le dévoilement du code génétique. Il mentionne les découvertes clés de la biologie de l'ARN lors de la seconde moitié du XX^{e} siècle, notamment l'identification de la rétrotranscription, l'identification des trois domaines du Vivant à partir d'une analyse phylogénétique de séquences d'ARN ribosomal, la télomérase, les *riboswitches*, puis l'interférence aux ARNs et les long ARNs non-codants. Il présente ensuite un survol des phénomènes de localisation des ARNs, en retraçant quelques exemples classiques d'ARN localisés, les mécanismes impliqués dans le processus ainsi que les technologies permettant d'étudier ces phénomènes.

4.2.2. Chapitre 2

Le deuxième chapitre est présenté sous la forme d'un article de revue de la littérature publié dans le journal *Frontiers in Microbiology*, intitulé « *Small Luggage for a Long Journey : Transfer of Vesicle-Enclosed Small RNA in Interspecies Communication* ». Il permet d'introduire la notion de vésicule extracellulaire et le ciblage des ARNs à ces structures. Il explore une notion émergente en écologie : le transfert de petits ARNs entre espèces biologiques symbiotiques. À ce titre, la première partie de la revue propose un recensement des cas de transferts inter-espèces documentés, notamment entre les plantes et leurs symbiotes bactériens, fongiques, arthropodes et nématodes. Les cas de transferts d'ARNs dans le tractus intestinal des mammifères, entre les cellules épithéliales et le microbiome, sont ensuite abordés. Ces exemples débouchent finalement sur la proposition d'un rôle pour les vésicules extracellulaires dans le transfert des ARNs entre espèces. Cette question est étayée par une revue détaillée des propriétés de ces vésicules et des mécanismes impliqués dans le ciblage des ARNs.

4.2.3. Chapitre 3

Le troisième chapitre, présenté sous la forme d'un article de revue de la littérature publié dans le *Journal of Developmental Biology*, est intitulé « *Flying the RNA Nest :* Drosophila

Reveals Novel Insights into the Transcriptome Dynamics of Early Development ». Il permet d'introduire le développement embryonnaire de la Drosophile et traite des changements dans le paysage transcriptomique qui se déroulent pendant cette période dynamique. Il porte une attention particulière aux avancées récentes dans le domaine, survenues grâce à l'établissement de nouvelles méthodes, notamment le Hi-C, le profilage des ribosomes et le système MS2. Il retrace les notions de localisation de l'ARN, de régulation de la traduction et de dégradation sélective des ARNs maternels. Il se penche ensuite sur les divers mécanismes contribuant à l'activation transcriptionnelle graduelle du génome zygotique et explore les changements qui surviennent dans le paysage chromatinien au fil de l'embryogenèse. Il traite finalement de la contribution de la transcription zygotique au déroulement de l'embryogenèse, ainsi que des propriétés émergentes de ces transcrits.

4.2.4. Chapitre 4

Le quatrième et présent chapitre propose un survol des hypothèses et objectifs de travail ainsi qu'un aperçu des différents chapitres de la thèse.

4.2.5. Chapitre 5

Le cinquième chapitre est constitué d'un article de méthode décrivant l'approche CeFra-seq, publié dans le journal *Methods* et intitulé « *CeFra-seq : Systematic Mapping of RNA Subcellular Distribution Properties through Cell Fractionation coupled to Deep-Sequencing* ». Son introduction récapitule brièvement les approches établies de détection des ARNs, notamment les outils d'imagerie, et plaide en faveur du développement d'approches systématiques pour étudier intégralement la localisation du transcriptome. Le protocole détaillé de l'approche CeFra-seq permettant de documenter la distribution spatiale du transcriptome chez les cellules en culture est ensuite fourni. Une validation de l'approche est proposée chez le modèle cellulaire leucémique humain K562 et un guide détaillé d'étapes de transformation de données est fourni afin de représenter la distribution spatiale des ARNs.

4.2.6. Chapitre 6

Le sixième chapitre est présenté sous la forme d'un article de recherche non-publié destiné au journal *RNA*. Il s'intitule « *Transcriptomic Profiling of Extracellular Vesicles and Subcellular Fractionation Reveals Strong Associations with the Cytosolic Repertoire* ». Il fait suite au développement de l'approche CeFra-seq chez les cellules K562 et vise à mieux comprendre les déterminants du ciblage des ARNs aux VEs. Les propriétés des ARNs enrichis dans chaque fraction subcellulaire sont contrastées et comparées dans le détail à celles des ARNs ciblés aux VEs. Ce travail révèle un lien entre l'enrichissement cytosolique et le ciblage au VEs des ARNs courts, surtout des produits de la Polymérase III. De plus, il met en évidence des éléments de séquence qui corrèlent avec l'abondance aux VEs chez les miARNs.

4.2.7. Chapitre 7

Le septième chapitre est présenté sous la forme d'un article de recherche publié dans le journal *Scientific Reports* et intitulé « *Comparative Transcriptomic Analysis of Human and* Drosophila *Extracellular Vesicles* ». Il propose une analyse de la morphologie et du contenu transcriptomique des VEs provenant de quatre modèles cellulaires, incluant deux modèles issus de la Drosophile et deux modèles issus de l'humain. Les résultats de cette analyse, qui reposent sur des analyses de microscopie électronique, de *nanoparticle tracking analysis* (NTA) et de séquençage à haut débit, permettent d'identifier les propriétés conservées ainsi que les points de divergence qui prévalent dans le ciblage extracellulaire des ARNs à travers l'évolution.

4.2.8. Chapitre 8

Le huitième chapitre est constitué d'un article publié dans le *Journal of Molecular Biology*, intitulé « *Biochemical Fractionation of Time-Resolved Drosophila Embryos Reveals Similar Transcriptomic Alterations in Replication Checkpoint and Histone mRNA Processing Mutants* ». Il traite de l'application du protocole CeFra-seq à un système dynamique, l'embryogenèse de la Drosophile. Cette adaptation permet d'identifier les transcrits présentant une forte asymétrie d'expression dans le temps et dans l'espace. Parmi ces répertoires asymétriques, j'ai notamment pu identifier les gènes d'expression strictement zygotiques, qui sont enrichis dans le noyau des embryons au stade blastoderme. En retour, un profilage

transcriptomique de mutants SLBP et Chk1 révèlent que l'expression des gènes zygotiques identifiés par CeFra-seq est sélectivement compromise chez ces mutants. Ce portrait suggère que la perte de SLBP et de Chk1 conduit à un arrêt dans le développement, qui rend caduque l'activation du génome zygotique.

4.2.9. Chapitre 9

Le neuvième chapitre se présente comme un article non-publié intitulé « *SLBP-dependent Antisense Transcription of Histone Genes is Developmentally Regulated and Modulates Histone Expression through Argonaute-2* » destiné au journal *Development*. Il repose sur l'identification chez l'embryon de Drosophile de populations d'ARNs non-codants provenant des cinq gènes d'histone canoniques. L'article en propose une caractérisation détaillée par diverses approches, notamment l'hybridation *in situ*, le buvardage de Northern et le séquençage à haut débit. Un traitement à l'α-amanitine, inhibiteur de la Polymérase II, indique que ces transcrits sont strictement issus du génome zygotique, résultat confirmé par le séquençage d'extraits d'ovaires, qui montre que ce tissu maternel n'exprime pas les ARNs antisens d'histone. Des études de perte de fonction et de transcriptomique comparative montrent ensuite que ces ARNs sont SLBP-dépendants et dépourvus d'une queue polyA. De plus, l'application de l'approche CeFra-seq révèle que les ARNs antisens d'histone co-ségréguent avec leurs ARNm complémentaires chez l'embryon, mais pas chez les cellules larvaires D17, où ces ARNs antisens sont plutôt localalisés au noyau. Ce résultat suggère que les ARNs antisens pourraient former des populations d'ARN double-brin par hybridation à leur messagers complémentaires. Puisque les ARNs double-brin peuvent s'associer à Dicer pour générer de petits ARN interférents, l'article investigue l'idée que ces transcrits antisens seraient impliqués dans la dégradation régulée des ARNm d'histones déposés maternellement. Une contribution des voies des petits ARNs dans la régulation de l'expression des histones est supportée par l'observation d'une association entre Argonaute-2, facteur clé des voies de l'interférence aux ARNs, et divers fragments des ARNs d'histone. Finalement, la déplétion du facteur Argonaute-2 chez l'embryon se traduit par une dérépression des ARNm d'histones. Ensemble, ces résultats supportent un modèle selon lequel une transcription antisens zygotique précoce des gènes d'histone contribue à la dégradation des ARNm maternels via les voies de petits ARNs.

4.2.10. Chapitre 10

Le dixième chapitre propose une discussion et une mise en perspective des résultats exposés dans cette thèse. Il revisite l'approche CeFra-seq en contrastant ses avantages, ses inconvénients ainsi que les limites inhérentes à l'approche. Il propose ensuite une mise en perspective des principales contributions des quatre études présentées en offrant des pistes pour poursuivre ces recherches.

4.2.11. Chapitre 11

Le onzième et dernier chapitre propose une brève conclusion de cette thèse.

Références

1. Lefebvre FA, Cody N, Bouvrette LPB, Bergalet J, Wang X, Lecuyer E: **CeFra-seq: Systematic mapping of RNA subcellular distribution properties through cell fractionation coupled to deep-sequencing.** *Methods* 2017.
2. Lefebvre FA, Benoit Bouvrette LP, Bergalet J, Lecuyer E: **Biochemical Fractionation of Time-Resolved Drosophila Embryos Reveals Similar Transcriptomic Alterations in Replication Checkpoint and Histone mRNA Processing Mutants.** *J Mol Biol* 2017.
3. Lefebvre FA, Bouvrette LPB, Bergalet J, Lecuyer E: **Data for the generation of RNA spatiotemporal distributions and interpretation of Chk1 and SLBP protein depletion phenotypes during Drosophila embryogenesis.** *Data Brief* 2017, **13:**28-31.
4. Villarroya-Beltri C, Gutierrez-Vazquez C, Sanchez-Cabo F, Perez-Hernandez D, Vazquez J, Martin-Cofreces N, Martinez-Herrera DJ, Pascual-Montano A, Mittelbrunn M, Sanchez-Madrid F: **Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs.** *Nature Communications* 2013, **4**.
5. Chen T, Xi QY, Ye RS, Cheng X, Qi QE, Wang SB, Shu G, Wang LN, Zhu XT, Jiang QY, Zhang YL: **Exploration of microRNAs in porcine milk exosomes.** *Bmc Genomics* 2014, **15**.
6. Villarroya-Beltri C, Baixauli F, Gutierrez-Vazquez C, Sanchez-Madrid F, Mittelbrunn M: **Sorting it out: regulation of exosome loading.** *Semin Cancer Biol* 2014, **28:**3-13.
7. Cha DJ, Franklin JL, Dou Y, Liu Q, Higginbotham JN, Demory Beckler M, Weaver AM, Vickers K, Prasad N, Levy S, et al: **KRAS-dependent sorting of miRNA to exosomes.** *Elife* 2015, **4:**e07197.
8. Tadros W, Lipshitz HD: **The maternal-to-zygotic transition: a play in two acts.** *Development* 2009, **136:**3033-3042.
9. Langley AR, Smith JC, Stemple DL, Harvey SA: **New insights into the maternal to zygotic transition.** *Development* 2014, **141:**3834-3841.
10. Harrison MM, Li XY, Kaplan T, Botchan MR, Eisen MB: **Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition.** *PLoS Genet* 2011, **7:**e1002266.

11. Marzluff WF, Wagner EJ, Duronio RJ: **Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail.** *Nat Rev Genet* 2008, **9:**843-854.
12. Iampietro C, Bergalet J, Wang X, Cody NA, Chin A, Lefebvre FA, Douziech M, Krause HM, Lecuyer E: **Developmentally regulated elimination of damaged nuclei involves a Chk2-dependent mechanism of mRNA nuclear retention.** *Dev Cell* 2014, **29:**468-481.
13. Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H: **Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells.** *Nature* 2008, **453:**793-797.

Préface au chapitre 5

Ce chapitre est présenté sous la forme d'un article de méthode publié dans le journal *Methods* (Référence : Lefebvre, F. A., Cody, N. A., Bouvrette, L. P. B., Bergalet, J., Wang, X., & Lécuyer, E. (2017). CeFra-seq: Systematic mapping of RNA subcellular distribution properties through cell fractionation coupled to deep-sequencing. Methods, 126, 138-148.)

L'article relate le dévelopmment d'une approche méthodologie nommée CeFra-seq, qui permet d'estimer la distribution subcellulaire des éléments du transcriptome à partir d'un fractionnement subcellulaire couplé à des analyses de séquençage à haut débit. Le protocole détaillé de CeFra-seq est fourni, puis une démonstration basée sur l'exemple de lignée humaine leucémique K562 est développé. Des approches de transformation de données permettant de représenter la distribution subcellulaire des ARNs sont ensuite discutées.

Dans le contexte de la présente thèse, les notions exposées dans cet article de méthode permettent de contextualiser l'ensemble de mes travaux recourant à l'approche CeFra-seq (Chapitres 6, 7, 8 et 9).

J'ai écrit l'intégralité du manuscrit, qui a fait l'objet d'une relecture et d'une série de corrections proposées par mon directeur de thèse, Dr Éric Lécuyer, dans le contexte de sa soumission à *Methods*. J'ai effectué les analyses liées aux vésicules extracellulaires ainsi que la préparation des figures présentées dans cette publication. Neal Cody, un ancien stagiaire postdocotoral du laboratoire Lécuyer, a fortement contribué au développement de l'approche CeFra-seq et à son application sur les lignée K562. Louis Philip Benoit Bouvrette, étudiant au Ph.D. au laboratoire, a contribué au développement des approches bio-informatique associées à la méthode CeFra-seq, notamment les étapes de transformation de données liées aux applications graphiques. Julie Bergalet, associée de recherche du laboratoire, a effectué les analyses de spectrométrie de masse sur les extraits subcellulaires. Xiaofeng (Andy) Wang, assistant de recherche du laboratoire, a effectué les immunobuvardages de Western et a contribué à la préparation des analytes pour la spectrométrie de masse. Dr Éric Lécuyer, mon superviseur, a contribué à l'étude en amassant les financements requis et en relisant le texte pour y suggérer diverses améliorations.

Je tiens à insister sur le fait que le développement de l'approche CeFra-seq est essentiellement le fruit du travail de Neal Cody. Ma contribution à cet égard est minime. J'ai pris le relai suite à son départ du laboratoire et j'ai contribué, dans ce contexte, aux analyses transcriptomiques. Les expériences de fractionnement subcellulaire ont donc été effectuées en grande majorité par Neal. Les expériences liées aux VEs sont mon travail. J'ai écrit l'intégralité du manuscrit.

Chapitre 5 : CeFra-seq: Systematic mapping of RNA subcellular distribution properties through cell fractionation coupled to deep-sequencing (Article #3)

Chapitre 5 : Article #3

CeFra-seq: Systematic mapping of RNA subcellular distribution properties through cell fractionation coupled to deep-sequencing

Fabio Alexis Lefebvre[1,2,*], Neal Cody[1,*],
Louis Philip Benoit Bouvrette[1,2], Julie Bergalet[1],
Xiaofeng Wang[1], Eric Lécuyer[1,2,3,4]

1- Institut de Recherches Cliniques de Montréal (IRCM)
Montréal, Québec, Canada

2- Département de Biochimie
Université de Montréal, Montréal, Québec, Canada

3- Division of Experimental Medicine
McGill University, Montréal, Québec, Canada

4- Address correspondences to: Dr. Eric Lécuyer
IRCM, RNA Biology Laboratory
110 Avenue des Pins, Ouest
Montréal, Québec, Canada
H2W 1R7
Tel: 514-987-5646, Fax: 514-987-5752
Email: Eric.Lecuyer@ircm.qc.ca

*These authors contributed equally to this manuscript

The datasets described here can be accessed through the ENCODE project portal under the accession ENCSR885DVH and at GEO under the accession numbers GSE88473, GSE87979, GSE88602, GSE87982, GSE88622 and GSE78625

Abstract

The subcellular trafficking of RNA molecules is a conserved feature of eukaryotic cells and plays key functions in diverse processes implicating polarised cellular activities. Large-scale imaging and subcellular transcriptomic studies suggest that regulated RNA localization is a highly prevalent process that appears to be disrupted in several neuromuscular disorders. These features underline the importance and usefulness of implementing procedures to assess global transcriptome subcellular distribution properties. Here, we present a method combining biochemical fractionation of cells and high-throughput RNA sequencing (CeFra-seq) that enables rapid and efficient systematic mapping of RNA cytotopic distributions in cells. The described procedure involves biochemical fractionation to derive extracts of nuclear, cytosolic, endomembrane, cytoplasmic insoluble and extracellular material from cell culture lines. The RNA content of each fraction can then be profiled by deep-sequencing, revealing global subcellular signatures. We provide a detailed protocol for the CeFra-seq procedure along with relevant validation steps and data analysis guidelines to graphically represent RNA spatial distribution features. As a complement to imaging approaches, CeFra-seq represents a powerful and scalable tool to investigate global alterations in RNA trafficking.

Introduction

1.1 Examples and functions of localized RNA transcripts

Over the last three decades, mRNA localization has emerged as a prevalent and conserved post-transcriptional mechanism to modulate local protein expression [1-3]. Specific examples of localized RNA transcripts have been described in various eukaryotic models, ranging from yeast to mammals. Notably, during development in *Drosophila*, localization of maternally provided transcripts such as *bicoid*, *nanos* and *oskar* plays pivotal roles in cell fate and body patterning commitments [4-7]. The recruitment of *ASH1* mRNA to the bud tip in *S. cerevisiae* regulates mating type switching [8], while the targeting of *β–actin* mRNA to the leading edge of migrating cells modulates chemotactic dynamics [9, 10]. In mammalian neurons, dendritic targeting and localized translation of transcripts encoding diverse cytoskeleton-associated proteins, kinases and subunits of glycinergic and cholinergic receptors can promote synaptogenesis and axon guidance [11]. In addition, aberrant sequestration of RNA in nuclear foci is an emerging feature of diverse human diseases related to nucleotide repeat expansions, such as Huntington's disease, myotonic dystrophy or spinocerebellar ataxia [12, 13].

1.2 Regulatory principles of RNA localization

While RNA length, nucleotide composition and transcriptional features likely impact diffusion kinetics and dynamics [14, 15], the paradigm of RNA trafficking revolves around *cis*-acting elements [16]. These functional motifs found within the RNA sequence can establish specific interactions with *trans*–acting factors, typically RNA-binding proteins (RBPs) [16, 17]. RNA elements can be recognized as primary sequence motifs and/or secondary structures by *trans*-acting factors [2, 17, 18]. They are frequently repeated within the sequence, functionally redundant, and found within the 3'-UTR of localized mRNAs. At least three principles through which *cis*- and *trans*-elements mediate RNA localization have been described: (1) RBPs bound to the RNA can establish interactions with cytoskeletal proteins for active and directed transport; (2) RNAs can be selectively protected from degradation by interacting with spatially-constrained RBPs; and (3) RBPs can restrict the diffusion of RNAs through local entrapment [1, 19].

1.3 Secretion of extracellular RNAs

Beyond subcellular targeting, specific RNAs can also be selectively secreted in the environment, notably within extracellular vesicles (EVs). Cells release diverse membranous structures, including larger membrane-shed ectosomes (100-500 nm) and smaller vesicles that arise from multivesicular endosomes (30-100 nm), also termed exosomes [20]. EV membranes provide topological protection to the intraluminal RNA cargo, which can enhance the stability of RNAs circulating in biological fluids. EVs have been described as vehicles of intercellular signalling in diverse systems and enable the transfer of functional miRNAs and mRNAs across tissues [21, 22]. Intercellular communication through EVs has notably been linked to tumour progression, immune surveillance, tissue regeneration and early embryogenesis [23-28]. RNA elements can modulate transcript abundance in EVs, yet the targeting mechanisms involved remain poorly defined [29]. At least two RBPs, hnRNPA2B1 [30] and Annexin A2 [31], have been linked to miRNA sorting to EVs.

1.4 Analysis of RNA localization via biochemical fractionation of cells combined with gene expression profiling

The development of ultracentrifugation techniques enabled the first biochemical separations of mammalian organelles via cell fractionation in the 1940s [32]. The introduction of velocity sedimentation using density gradients of sucrose or cesium chloride refined the technique, yielding finer, more resolved separations. In RNA biology, polysome fractionation can be used to infer differential translational outcomes by revealing ribosomal occupancy and partitioning of a given mRNA across cytosolic and reticulum-bound pools of ribosomes [33]. In addition, cell fractionation combined with high-throughput gene expression profiling has offered an insightful strategy to characterize asymmetrically localized RNA populations in eukaryotic cells [34]. Such approaches have been employed to identify mRNA and non-coding RNA populations targeted to different organelles and subcellular compartments, including the nucleus [35-38], cytoplasm [35, 37, 39], cytosol [36, 40-42], membrane and insoluble fractions [41], polysomes [42-44], the endoplasmic reticulum [45-49], mitochondria [50, 51], microtubules [52, 53], pseudopodia [54], and neuronal projections [11, 55-62]. In addition to revealing the enrichment of functionally

coherent RNAs in specific intracellular compartments, these approaches can also provide mechanistic insights into transcriptome regulatory dynamics when applied to cellular samples exposed to specific environmental stimuli [35], or when combined with loss-of-function approaches to deplete specific RBPs [38, 41].

1.5 Introduction to CeFra-seq

Here, we detail an optimized method to study RNA subcellular distribution properties in cultured cells via biochemical cellular fractionation combined with RNA-seq (CeFra-seq). CeFra-seq relies on a cell fractionation approach adapted from previous protocols to yield extracts of nuclear, cytosolic, endomembrane, cytoplasmic insoluble and extracellular material from culture human or *Drosophila* cell lines [33, 41, 63-65]. These extracts can then be subjected to in depth transcriptomic characterization, revealing global subcellular distribution patterns. The key advantage of CeFra-seq lies in its systematic rendition of transcriptome-wide subcellular distributions, which may yield insights into RNA localization mechanisms when combined with functional dissection strategies. For example, this approach was previously used to investigate alterations in RNA trafficking upon depletion of splicing regulators of the Muscleblind family [41] and the histone mRNA processing factor SLBP in *Drosophila* embryos [38] [66]. Here, we provide a detailed CeFra-seq protocol along with guidelines to validate the subcellular preparations. In addition, we describe data transformation steps that can be implemented to visualize global RNA cytotopic distribution signatures using simplex plot graphs. As an illustration, we provide an overview of CeFra-seq results in the human myelogenous leukemia cell line K562, which our laboratory contributed to the Encyclopedia of DNA Elements (ENCODE) consortium (accessible at encodeproject.org, under the accession number ENCSR885DVH).

2 CeFra-seq Protocol

2.1 Outline of the procedure

Starting with a population of freshly growing cells, EVs (E) were collected over a window of 48h from conditioned cell media using a sequential ultracentrifugation approach. Cells were subsequently harvested and subjected to subcellular fractionation to yield four intracellular extracts, the nuclear (N), cytosolic (C), endomembrane (M) and insoluble (I) fractions (**Figure 5.1**). Briefly, cells were subjected to a mild hypotonic lysis involving Dounce homogenization. Nuclei were recovered by centrifugation and purified through sucrose cushion ultracentrifugation. The cytoplasmic supernatant was further processed to derive a highly soluble extract, the cytosolic fraction. The resulting pellet was solubilized with Triton X-100 (1%) and further processed by ultracentrifugation, yielding a re-solubilized part, the endomembrane fraction, and a cytoplasmic insoluble pellet. Total RNA and protein can be recovered through guanidinium thiocyanate-phenol-chloroform extraction using TRIzol (Thermo Fisher Scientific) [67, 68].

2.2 Cell number and culture media

For subcellular fractionation, the number of cells may vary depending on histological type: 2.5×10^7 is optimal for K562, whereas 3.5×10^7 cells is recommended for smaller *Drosophila* cells, such as Dm-D17-c3 and Dm-S2R+ cells. For EV purification, a minimum of 80 mL of cultured cell supernatant corresponding to cultures of approximately 10^8 K562 cells is indicated. Conditioned media must be used to prevent contamination in the EV fraction with bovine samples present in fetal bovine serum (FBS, Wisent). FBS conditioning involves bovine EV depletion, achieved by ultracentrifugation at 110,000g for 18 h at 4°C. FBS supernatant is pipetted and filtered under sterile conditions prior to culture media preparation.

2.3 Ultracentrifugation

The subcellular fractionation protocol uses small volumes for hypotonic lysis, warranting the use of a micro-ultracentrifuge, such as the RCM-100 centrifuge (Sorvall). The protocol is best executed with a pair of small rotors, such as the RP100-AT4 angled rotor (Sorvall) and the RPS55 swinging bucket rotor (Sorvall). To process the larger volumes of conditioned media involved in EV purification, we have used the SW-32-Ti rotor (Beckman) with the XPN Optima centrifuge (Beckman). All three recommended rotors have space for 6 individual tubes. It is crucial to balance the tubes exactly using a sensitive scale. Rotors and ultracentrifuge chambers should be cooled to 4°C before use.

2.4 Reagents/solutions

- **Hypotonic lysis buffer**

20 mM Tris HCl (pH=7.5),10 mM KCl**,** 1.5 mM $MgCl_2$, 5 mM EGTA, 1 mM EDTA, 1 mM DTT, 1 mM PMSF, 0.15 U/mL Aprotinin, 20μM Leupeptin, 40 U/mL RNase Out (Thermo Fisher Scientific). Note: Hypotonic buffer containing 1% Triton X-100 is used for steps 26-32.

- **Sucrose Buffer [0.32 M]**

0.32 M Sucrose, 3 mM $CaCl_2$, 2 mM MgAc, 0.1 mM EDTA, 10 mM Tris-HCl (pH=8).
Immediately before use, add 1 mM DTT, 0.5% v/v NP-40, protease and RNase out.

- **Sucrose buffer** [**2.0 M**]

2 M Sucrose, 5 mM MgOAc, 0.1 mM EDTA, 10 mM Tris-HCl (pH=8).
Before use, add 1 mM DTT, protease and RNase out.

- **Guanidinium thiocyanate-phenol-chloroform extraction**

We recommend the use of Trizol reagent for RNA and protein extraction. Trizol LS (liquid sample) was used for the endomembrane and cytosolic fractions according to the manufacturer's instructions.

2.5 Detailed Protocol

In the following section, we provide a detailed description of the steps involved in the purification of extracellular and subcellular fractions. The steps 1 to 9 relate to EV purification, the steps 12 to 26 enable nuclear and cytosolic purification while the steps 27 to 34 lead to the recovery of the endomembrane and cytoplasmic insoluble extracts.

- **EV fraction**

1) Collect 80 mL of conditioned cell culture media and pellet floating cells at 400g for 10 min at room temperature (RT).
2) Transfer the supernatants to 50 mL conical tubes and centrifuge at 3,000g for 30 min at 4°C. Discard pellet.
3) Transfer the supernatants to 39 mL polyallomer tubes (Beckman) and ultracentrifuge at 10,000g for 30 min at 4°C. Discard the pellets.
4) Transfer the supernatants to new 39 mL tubes and ultracentrifuge at 110,000g for 60 min at 4°C.
5) Aspirate and discard the supernatant, leaving about 1 mL not to disrupt the EV pellet.
6) Vigorously pipet up and down the remaining 1 mL of supernatant to resuspend the EV pellet and transfer to a 1.5 mL microtube.
7) Balance the microtubes with cold PBS and proceed to spin at 110,000g for 60 min at 4°C using the RP100-AT4 rotor and the RCM-100 micro-ultracentrifuge.
8) Discard supernatant without disrupting the pellets and add 1 mL of cold PBS. Repeat the spin at 110,000g for 60 min at 4°C to wash the EV pellet.
9) For RNA extraction, save the EV pellet in 500 µL of Trizol, then freeze the tube at -80°C. Alternatively, EV preparations can be fixed in glutaraldehyde for electron microscopy validation or readily be resuspended in 1.0 mL of PBS for nanoparticle tracking analysis.

- **Total cell extract**

10) Pellet cells at 400g for 5 min at RT and wash twice with PBS.
11) Resuspend the pellet with 5 mL of PBS. Aliquot 500 µL of the cell suspension in a 1.5 mL microtube and centrifuge at 1,000g for 5 min at RT. Upon centrifugation, remove the

supernatant and add 1 mL of Trizol reagent. Pipet vigorously, let stand at RT for 10 min, then freeze the total cell extract at -80°C.

- **Nuclear fraction**

12) Spin the remaining 4.5 mL of cell suspension to obtain a cell pellet and discard the PBS supernatant.
13) Resuspend the cell pellet in 1.5 mL of cold hypotonic lysis buffer containing fresh protein and RNA inhibitors.
14) Keep on ice for a total of 15 min to swell the cells. Gently swirl the suspension if clumps appear at the bottom of the tube.
15) Pour the swelled cells into a homogenizer chamber and dounce gently for 5 strokes (up and down = one stroke) using a ball-bearing homogenizer with tight fitting Type B Pestle.
16) Pour the homogenized cells into a 1.5 mL microtube (labeled 1st spin) and clear the cell debris and nuclei by spinning at 1,000g for 10 min at 4°C.
17) Transfer the supernatant to a new microtube; take care to leave about 50 µL to prevent disrupting the pellet. Spin the supernatant again for 1,000g for 10 min at 4°C to clear any further cell debris. Keep the first tube containing pelleted cell debris and nuclei aside on ice.
18) After the 2nd spin, remove the supernatant and leave about 75 µL away from any remaining debris, place this volume in a new tube: "S-100" and keep on ice.
19) During this 10 minute period, take the cell debris/nuclei pellet and discard the remaining 75 µL of supernatant away from the pellet obtained from the 1st spin. Gently rinse the nuclear pellet with 100 µL of hypotonic lysis buffer.
20) Resuspend the nuclear/debris pellet in 0.5 mL of 0.32 M sucrose buffer containing 0.5% v/v NP-40, gently pour into the Dounce homogenizer and homogenize with Type B pestle using 5 strokes.
21) Pour the 0.32 M sucrose homogenate into a fresh tube and add an equal volume of 2.0 M sucrose buffer. Mix by inversion.
22) Place 0.5 mL of 2.0 M sucrose buffer into a fresh 2 mL polyallomer tube and carefully place the sucrose homogenate on top of this 2.0 M sucrose cushion.

23) Use the Sorval RPS55 swinging bucket rotor and the RCM-100 ultracentrifuge to separate the cell debris from the nuclei. Carefully place the polyallomer tube into the appropriate chamber of the RPS55 rotor (hook into place) and spin at 30,000g for 30 min at 4°C.

24) Discard the supernatant (containing cell debris). Rinse the nuclear pellet 2-3 times in 0.32M sucrose (with freshly added 0.5% v/v NP-40); and immediately harvest the nuclei with 100 µL of Trizol reagent. Transfer the nuclear Trizol suspension to a new 1.5 mL microtube, bring the volume of Trizol to 1 mL and dissociate the nuclei by pipetting thoroughly; then freeze the tube at -80°C.

- **Cytosolic fraction**

25) Centrifuge the "S-100" tube at 100,000g for 60 min at 4°C using the RP100-AT4 micro-ultracentrifuge rotor.

26) The supernatant corresponds to the cytosolic fraction. Gently aliquot 250 µL of the supernatant into 3-4 tubes microtubes. Immediately place 750 µL of Trizol LS into each tube. Let stand 10 min at RT, then store at -80°C.

- **Endomembrane fraction**

27) Gently rinse the pellet from "S-100" tube twice with 100 µL of hypotonic buffer.

28) Resuspend the pellet in 100 µL hypotonic lysis buffer containing 1% Triton X-100, protein and RNase inhibitors. Transfer this suspension into a cleaned Dounce homogenizer chamber and add 900 µL of cold hypotonic lysis buffer containing 1% Triton X-100, protein and RNase inhibitors.

29) Dounce homogenize for 40 strokes, let stand on ice for 45 min and transfer the homogenate into a new tube labelled "P-100".

30) The "P-100" tube will undergo the second round of ultracentrifugation to obtain the endomembrane fraction (soluble in Triton X-100) in the supernatant, and the cytoplasmic insoluble fraction in the pellet (insoluble in Triton X-100). Spin at 100,000g for 30 min at 4°C.

31) The supernatant corresponds to the endomembrane fraction**.** Upon ultracentrifugation, gently aliquot 250 µL of the supernatant from the "P-100" tube into 3-4 microtubes while leaving about 75 µL not to disrupt the pellet. Immediately add 750 µL of Trizol LS into each of the endomembrane sample tubes. Let stand 10 mins at RT, and save at -80°C.

- **Cytoplasmic Insoluble fractions**

32) Gently rinse the pellet from the "P-100" tube with 100 µL of hypotonic Triton X-100 1% buffer. Repeat once.
33) This pellet corresponds to the cytoplasmic insoluble fraction. Resuspend the pellet in 50 µL of buffer, transfer to a new microtube and add 1 mL of Trizol reagent. Pipette vigorously, let stand 10 mins RT, and store at -80°C.
34) Extract RNA/protein extraction from Trizol and Trizol LS samples using standard procedures, as recommended by the manufacturer.

In conclusion, the steps detailed above provide an efficient approach to recover RNA extracts targeted to subcellular and extracellular compartments, which can then be mapped systematically through deep-sequencing.

3. CeFra-seq Validation in K562 cells

3.1 EV validation through morphological characterization

The integrity of EV preparations should be validated before proceeding to transcriptomic analyses [69]. We opted for a standard characterization involving nanoparticle tracking analysis (NTA) and transmission electron microscopy (TEM) of K562 EVs. NTA conducted with the Nanosight LM10 (Malvern) device identified approximately 25 billion particles and revealed heterogeneous populations with an average diameter (μ_{NTA}) of 179 nm and a standard deviation (SD_{NTA}) of 104 nm (**Figure 5.2A**). In addition, EV pellets were fixed in glutaraldehyde, stained with uranyl acetate and observed by TEM using the Technai 12 120kV microscope as previously described [27, 64]. Visual examination of EV preparations confirmed the heterogeneity determined by NTA, revealing small cup-shaped exosome-like vesicles next to large, leaflet-like structures likely shed by the plasma membrane (**Figure 5.2B**). TEM micrographs were quantified by individually assessing the observed diameter of EVs, which broadly confirmed NTA measurements (μ_{TEM}=139 nm, SD_{TEM}=104 nm, n=181). Such size distributions are consistent with heterogeneous preparations, likely consisting of both small secreted exosomes and larger membrane-shed ectosomes. In addition to morphological analyses, EV preparations can be validated through the identification of various protein and RNA markers, which have been curated in databases such as Vesiclepedia [70] and ExoCarta [71].

3.2 Validation of protein subcellular distributions by Western blotting

Before submitting subcellular fractions to high throughput analyses, we recommend investigating the distributions of a few protein markers to confirm their integrity. For this purpose, 50 μg of protein extracts derived from total K562 cells and subcellular fractions were submitted to SDS-PAGE and Western blotting (**Figure 5.2C**). Diverse protein targets can be used in this validation effort; we selected four candidates associated to strong spatial asymmetry that were also employed in previous fractionation studies [33, 72]. For example, the nucleosomal factor histone H3 was strongly enriched in the nuclear fraction, while monomeric α-tubulin was overrepresented in the cytosolic fraction. ER-resident proteins bearing the tetrapeptide motif KDEL [73] were enriched in the endomembrane fraction, while the centrosomal protein Ninein

involved in microtubule minus-end anchoring [74] was strongly enriched in the cytoplasmic insoluble fraction.

3.3. Validation of RNA subcellular distributions by RT-qPCR

In addition to protein profiling, we recommend documenting the distribution of a few RNA transcripts through the fractions prior to deep-sequencing. For this purpose, aliquots (1 μg) of DNase-treated RNA extracts from K562 cells and fractions were submitted to a reverse-transcription reaction using Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT, Promega). Since the analysis of subcellular fractions is not amenable to housekeeping gene normalization, it is crucial to work with aliquots of strictly equal RNA amounts. We performed quantitative PCR to investigate the distribution of four RNA transcripts across fractions. For normalization, we used the $\Delta\Delta CT$ approach [75] to perform cross-comparisons of subcellular fractions with the total cell extracts and derive cumulative percentage values (**Figure 5.2D**). The long non-coding transcript *XIST*, which coats the silent female X chromosome [76], was exclusively found in the nuclear fraction. This transcript appears to be an optimal nuclear marker in female somatic lineages. For cytoplasmic fractions, *Histone H3* mRNA was most prevalent in the cytosolic fraction, while *Anillin* mRNA, which encodes a regulator of cytoskeletal dynamics [77], is enriched in the insoluble fraction.

3.4 Validation of Bioanalzyer electrophoretic profiles

We recommend profiling the RNA extracts with Bioanalyzer electropherograms to validate the samples and confirm cross-replicate reproducibility. We profiled biological duplicates of 5 ng aliquots from extracellular, subcellular and total K562 samples using the Agilent 2100 Bioanalyzer device and the RNA 6000 Pico Chip (Agilent) (**Figure 5.3**). The characteristic 18S and 28S ribosomal RNA (rRNA) signatures were scarce in the EV and cytosolic fractions but abundant in the insoluble, membrane and nuclear extracts. Similarly, low rRNA frequencies have been found in EVs purified from diverse models [78, 79].

4. Data Normalization, transformation and simplex plot representation

4.1 Transcriptomic analyses

Deep-sequencing of K562 fractions was performed using the Illumina HiSeq2000 sequencer following rRNA depletion with the RiboZero kit (Illumina). Between 15 and 40 million reads were mapped in each sample (**Table 5.1**), a sequencing depth deemed sufficient for gene-centric analyses [80]. FastQC was used to perform quality control on the Fastq files and low quality reads were discarded (base quality <20). If necessary, we recommend discarding the first nucleotides from the reads 5' end using Cutadapt [81]. Genome alignment of the trimmed reads was conducted with TopHat (with options --no-novel-juncs --no-discordant --library-type fr-firststrand) [82]. The raw alignment count was calculated using featureCounts (options: -p -B -C -s 2) [83]. The GTF annotation file was used for the genome alignment and the raw alignment count was obtained from ENSEMBL (GRCh37.75). Differential gene expression was calculated with DEseq2 using raw alignment counts from featureCount as inputs [80]. To facilitate the viewing of the alignments in the UCSC genome browser, bigWig files were generated from the BAM file using bedtools genomecov (options –split -bg) and the bedGraphToBigWig executable obtained from UCSC [84]. We used the bigWig files to generate UCSC browser screenshots of selected transcripts associated with strong subcellular asymmetry (**Figure 5.4**). This data confirmed that RNA-seq analysis efficiently recapitulates gene-specific asymmetries inferred from qRT-PCR. Indeed, the canonical histone gene *HIST1H2AG* was strongly enriched in the cytosolic fraction, whereas the lncRNA *XIST* displayed a substantial nuclear enrichment. RNA-seq also revealed the cytoplasmic insoluble enrichment of *NUP160*, which encodes a nucleoporin, and the endomembrane enrichment of *ELK1*, a transcriptional activator.

Transcript expression was quantified using the FPKM metric (Fragments Per Kilobase of exon per million fragments Mapped) [82]. Mapped transcripts were ranked as a function of cumulative read counts and the 5,000 most abundant transcripts were considered. FPKM values were then used to derive cytotopic ratios, which correspond to the proportion of total cumulative abundance and reflect the distribution of a specific transcript through the five fractions. For example, the nuclear cytotopic ratio for a given transcript is defined as:

$$R_{nuc} = NUC_{FPKM}/(EV_{FPKM} + CYT_{FPKM} + INS_{FPKM} + MEM_{FPKM} + NUC_{FPKM})$$

4.2 Simplex plot representations

In this section, we describe a set of trigonometric transformations to represent cytotopic ratios within simplex plots. Simplex plots are typically used to represent a set of three variables, forming an equilateral triangle (2-simplex) [85]. In such systems, every corner corresponds to a subcellular fraction and every point represents a different RNA. For each point, fraction-specific enrichment decreases as its distance increases from the considered corner. Simplex representations are obtained by computing the vectors of cytotopic ratios (R) of a given factor for each fraction, which are then projected towards each compartment represented in the simplex. When 3 fractions are to be represented (**Figure 5.5A**), the resulting 2-simplex centered at the origin has coordinates V_1=[0,1], V_2=[$\sqrt{3}/2$, -1/2], V_3=[-$\sqrt{3}/2$, -1/2] in Cartesian space. The coordinates of a point representing a transcript are obtained from cytotopic ratios (R_x) and defined as:

(i) $$x = R_{ins} \times \cos\frac{11\pi}{2} + R_{cyt} \times \cos\frac{\pi}{2} + R_{mem} \times \cos\frac{7\pi}{6};$$
$$y = R_{ins} \times \sin\frac{11\pi}{2} + R_{cyt} \times \sin\frac{\pi}{2} + R_{mem} \times \sin\frac{7\pi}{6}$$

Accurate representations of cytotopic ratios in four different fractions involve a tridimensional projection in the form of a tetrahedron (**Figure 5.5B**). Assuming a 3-simplex centered at the origin with coordinates V_1=(-1/3,-$\sqrt{2}/3$, -$\sqrt{\frac{2}{3}}$), V_2=(-1/3,-$\sqrt{2}/3$, $\sqrt{\frac{2}{3}}$), V_3=(-1/3, $\sqrt{8}/3$,0), V_4=(1,0,0), the coordinates of a point representing a transcript are computed as:

(ii) $$x = R_{cyt} \times \left(-\frac{1}{3}\right) + R_{mem} \times \left(-\frac{1}{3}\right) + R_{ins} \times \left(-\frac{1}{3}\right) + R_{nuc}$$
$$y = R_{cyt} \times \left(-\frac{\sqrt{2}}{3}\right) + R_{mem} \times \left(-\frac{\sqrt{2}}{3}\right) + R_{ins} \times \left(-\frac{\sqrt{8}}{3}\right)$$
$$z = R_{cyt} \times \left(-\sqrt{\left(\frac{2}{3}\right)}\right) + R_{mem} \times \left(\sqrt{\left(\frac{2}{3}\right)}\right)$$

We extended the triangular and tetrahedral projections to represent cytotopic distributions inferred from 4 and 5 fractions with a 2-simplex rendition. The resulting diamond and pentagonal plots enable the simultaneous representation of four and five fractions, respectively (**Figure 5.5C-D**). These approaches can be useful to assess global distributions and visualize the enrichment of factors associated with a strong spatial asymmetry. However, the concomitant representation of more than three fractions in a two-dimensional plane impair the linear relationship between distance and enrichment, which can only be accurately achieved using triangular and tetrahedral projections. Therefore, rectangular and pentagonal plots should be interpreted with caution. 2-simplex rectangular (iii) and pentagonal (iv) coordinates are derived as follows:

(iii) $x = R_{mem} - R_{nuc}$

$y = R_{cyt} - R_{ins}$

(iv) $x = R_{ev} \times \cos(0.2618) + R_{cyt} \times \cos(1.571) + R_{ins} \times \cos(2.879) + R_{mem} \times \cos(4.101) + R_{nuc} \times \cos(5.323)$

$y = R_{ev} \times \sin(0.2618) + R_{cyt} \times \sin(1.571) + R_{ins} \times \sin(2.879) + R_{mem} \times \sin(4.101) + R_{nuc} \times \sin(5.323)$

Concluding remarks

We described a stepwise introduction to CeFra-seq, a biochemical approach that enables efficient, scalable transcriptome-wide mapping of subcellular spatial distributions. We envision CeFra-seq as a powerful tool to monitor global changes in RNA distributions associated with specific disease states or because of precise genetic or environmental alterations, such as targeted depletions of RBPs, cell differentiation events or in response to extracellular stimuli. While CeFra-seq provides rapid access to global spatial distributions, we recommend adopting protein and RNA imaging techniques such as immunofluorescence and *in situ* hybridization to validate and refine CeFra-seq findings on specific targets, as detailed in the FISH protocol provided by our laboratory for this issue of *Methods* [86]. We anticipate that these complementary approaches have the potential to further our understanding of RNA localization in a wide range of eukaryotic models.

Acknowledgments

We thank Christopher Burge (MIT), Eric T. Wang (University of Florida), Brent Graveley Michael Duff, Xintao Wei (University of Connecticut Health Center) and Atefeh Mohajeri (McGill University) for insightful discussions. We thank Janusz Rak and Delphine Garnier (McGill University) for their help with EV purification and characterization. We are grateful to Alexis Blanchet-Cohen and Odile Neyret (IRCM) for their help with transcriptomic analyses. We thank Cédric Diot for his contributions to the discussions regarding CeFra-seq. We are grateful to Nicole Francis (IRCM) for lending us the Sw-32-Ti rotor. F.A.L. is supported by a Canadian Institute of Health Research (CIHR) scholarship, J.B. by a Fonds de Recherche Québec Santé (FRQS) postdoctoral scholarship, while E.L. is a Junior 2 scholar of the FRQS. This work was supported by grants to E.L. from CIHR (MOP137096), FRQS and the Cancer Research Society (grant #20227).

Figure legends

Figure 5.1. CeFra-seq experimental workflow

Schematic view of the CeFra-seq experimental workflow. Cell lysates and culture supernatants are processed to derive 5 extracts: the extracellular (E), cytosolic (C), cytoplasmic insoluble (I), endomembrane (M) and nuclear (N) fractions. P=pellet S=supernatant.

Figure 5.2. Validation of the CeFra-seq approach

(A) Nanoparticle tracking analysis (NTA) of K562 EV preparations showing particle count (n), size distribution (red), average size (μ) and standard deviation (SD). **(B)** Transmission electron micrographs of K562 EVs preparations. Scale bar=100 nm. **(C)** Western blotting targeting Histone H3, α-tubulin, Ninein and the KDEL tetrapeptide was performed on equal amounts of protein extracted from total K562 cells (T) and nuclear (N), cytosolic (D), endomembrane (M) and cytoplasmic insoluble (I) fractions. **(D)** Subcellular cytotopic ratios inferred from qRT-PCR for the lncRNA *XIST* and the mRNAs encoding histone H3, Anillin and an Adiponectin receptor.

Figure 5.3. Fraction-specific RNA electrophoretic length distributions

Biological duplicates of Bioanalyzer electrophoretic profiles of extracellular, subcellular, and total K562 RNA extracts. 18S and 28S rRNA peaks are labeled "R".

Figure 5.4. Read coverage of selected asymmetrically distributed transcripts

UCSC screenshots of fraction-specific read coverage mapped at six loci encoding asymmetrically distributed RNAs: *VTRNA1-1*, *HIST1H2AG*, *NUP160*, *ELK1*, *SCARNA13* and *XIST*. Locus architecture and chromosomal coordinates are displayed above the tracks; dark boxes indicate exons and lines represent intronic regions. The *y*-axis refers to read coverage and the *x*-axis corresponds to base positions.

Figure 5.5. Simplex plot representations of cytotopic ratio distributions

Simplex plots of K562 total RNA cytotopic distributions (black) overlaid with histone mRNA distributions (red). Triangular (**a**), tetrahedral (**b**), rectangular (**c**) and pentagonal (**d**) distributions were generated based on the number of subcellular fractions considered.

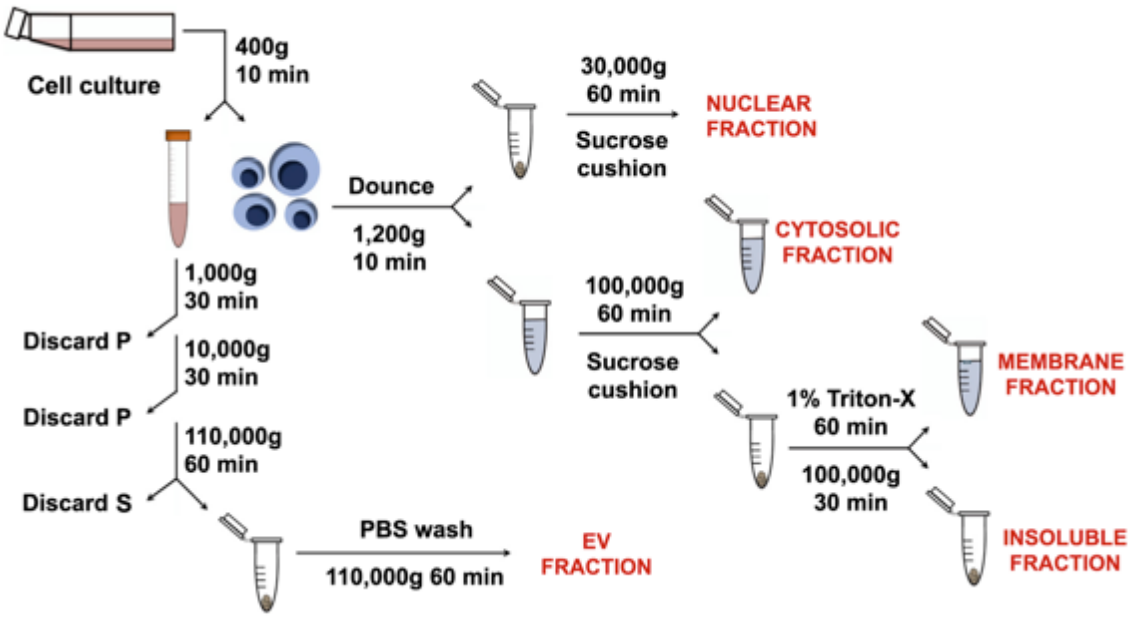


Figure 5. 1. CeFra-seq experimental workflow

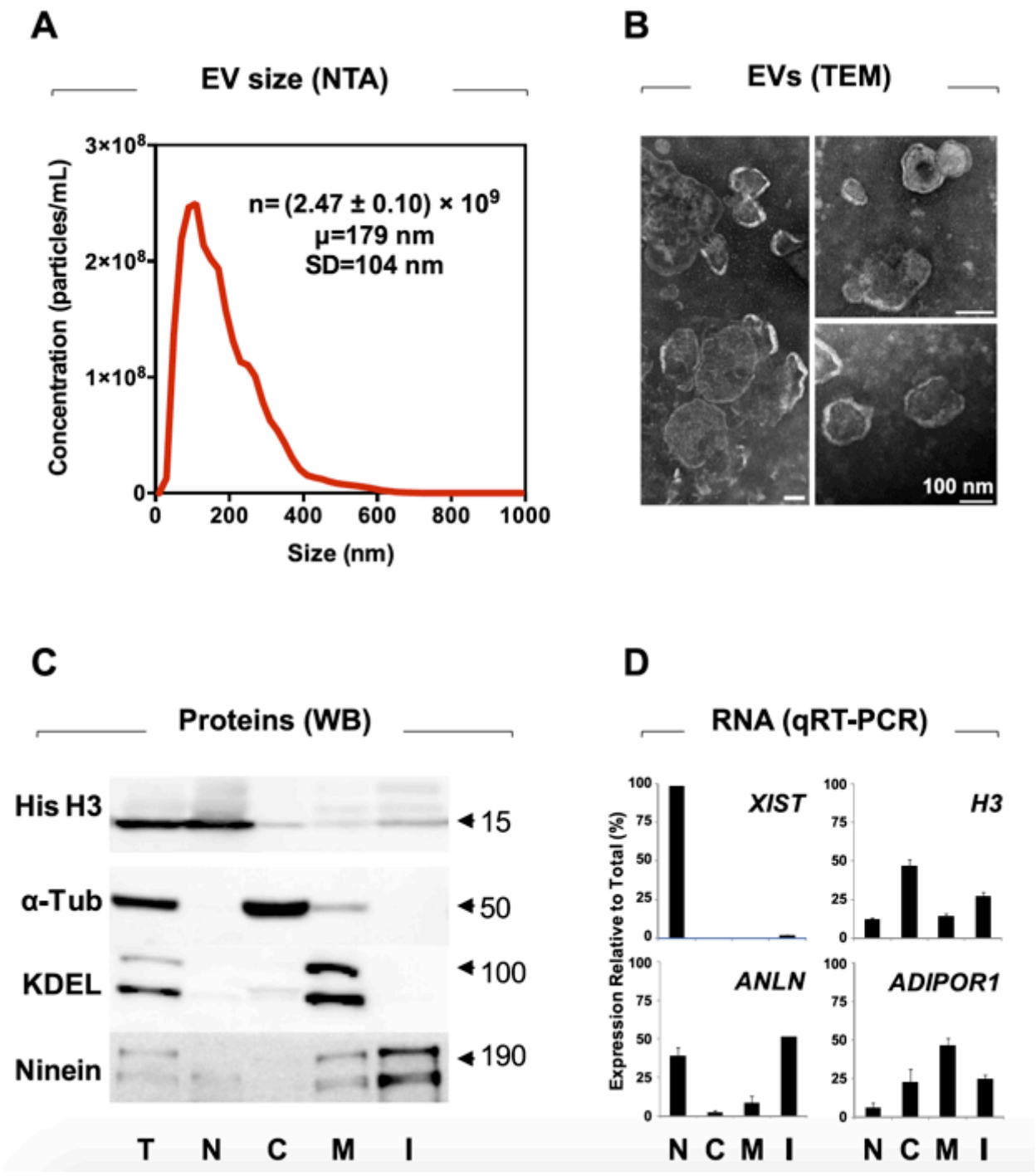


Figure 5. 2. Validation of the CeFra-seq approach

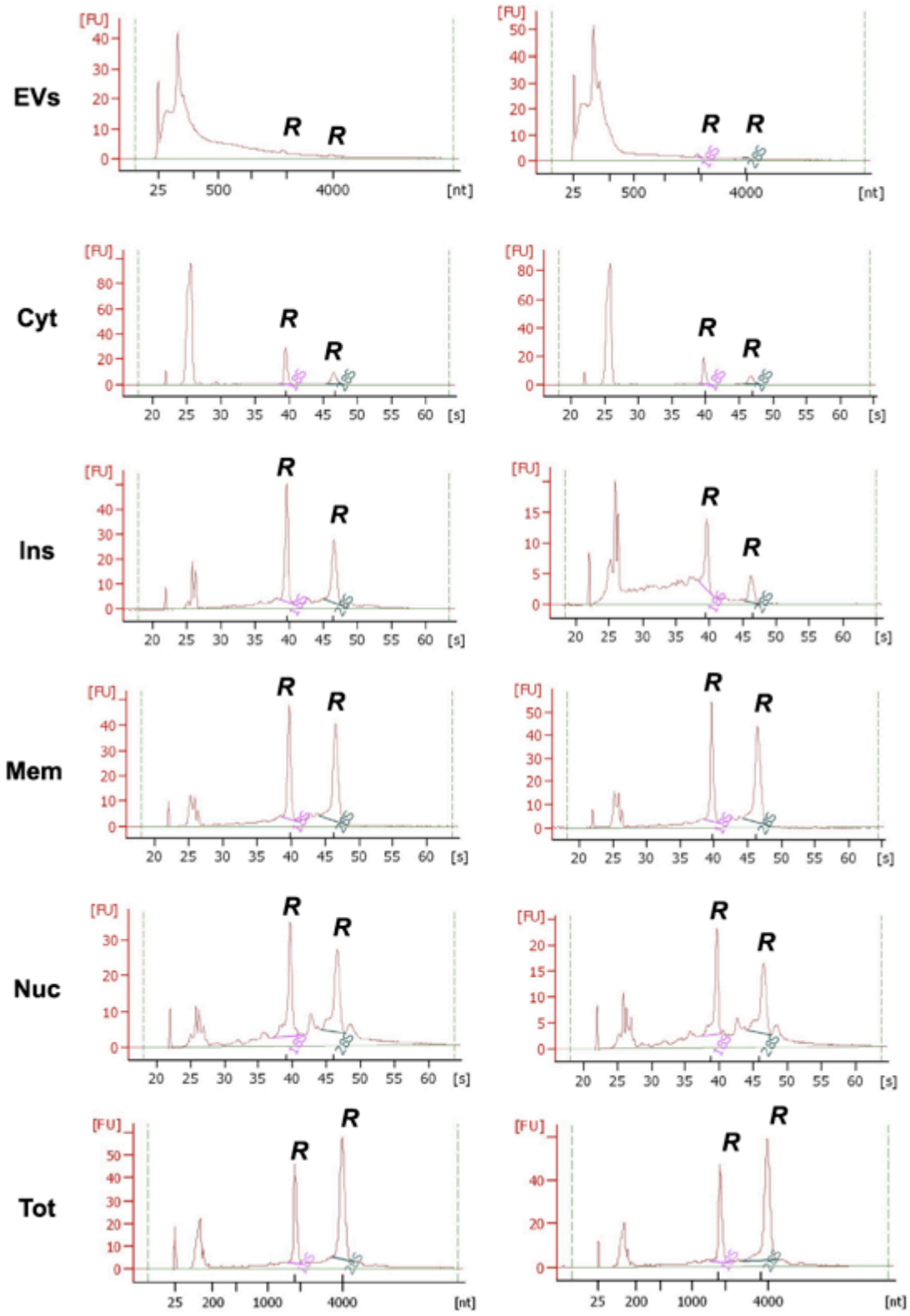


Figure 5. 3. Fraction-specific RNA electrophoretic length distributions

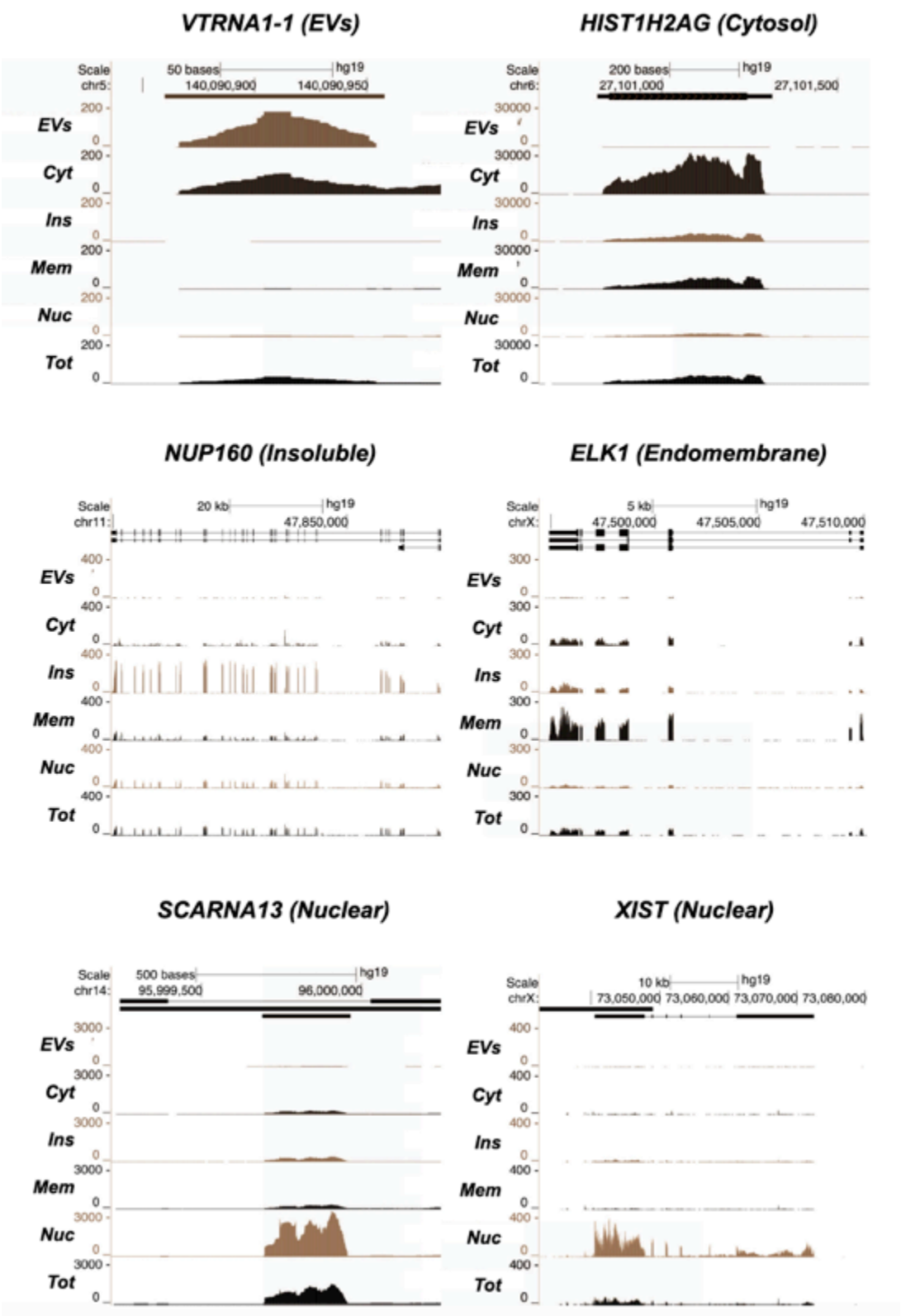


Figure 5. 4. Read coverage of selected asymmetrically distributed transcripts

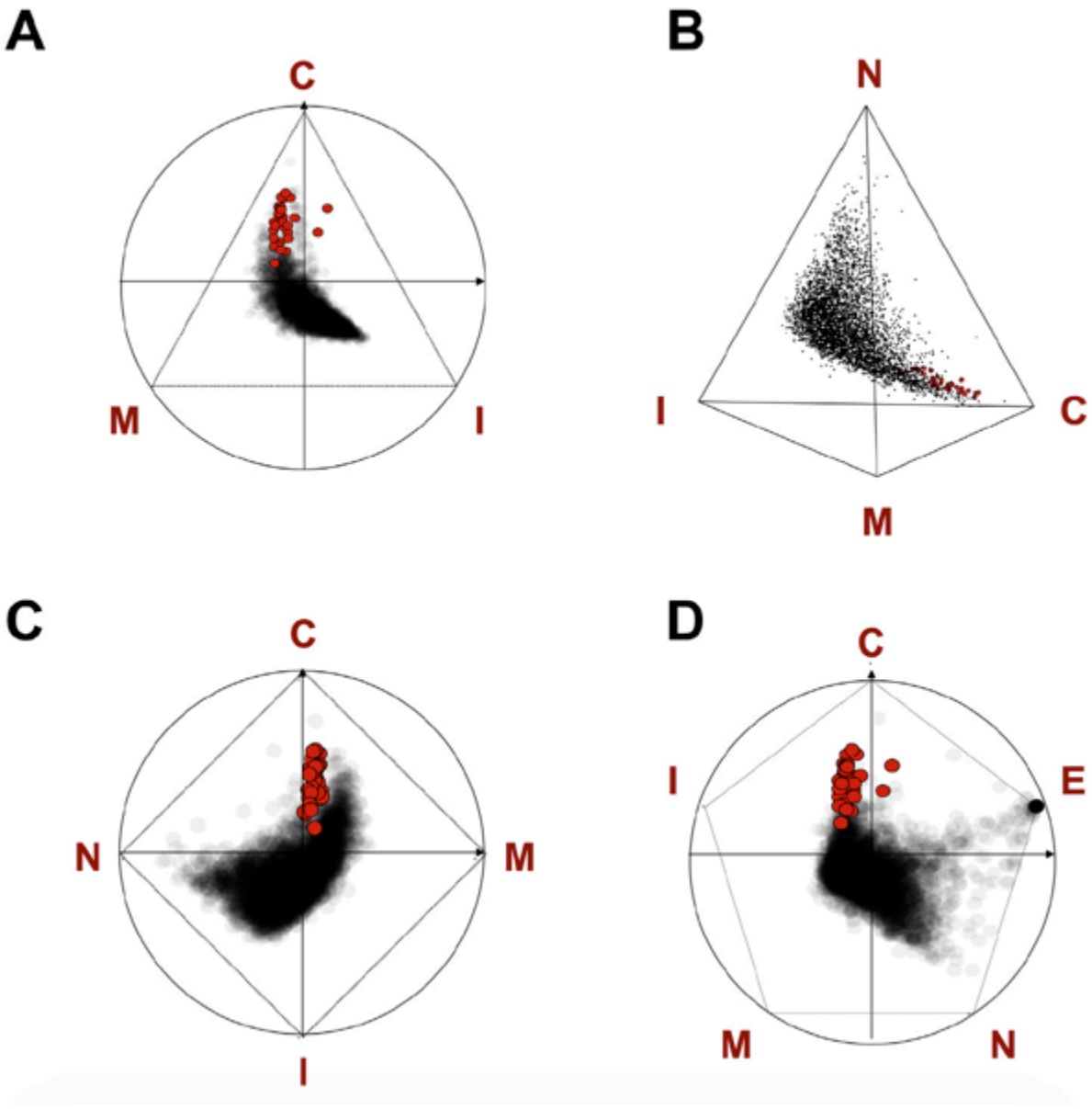


Figure 5. 5. **Simplex plot representations of cytotopic ratio distributions**

Table 5. I. **CeFra-seq read metrics in K562**

Name	Number of Reads	% Duplicate
EVs	16,740,794	85.25
EVs	21,303,078	80.43
Cytosol	29,811,993	57.06
Cytosol	22,700,706	50.75
Insoluble	27,018,740	14.88
Insoluble	27,922,514	13.24
Endomembrane	26,716,019	28.18
Endomembrane	29,873,818	34.65
Nuclear	24,811,186	10.48
Nuclear	24,756,790	10.00
Total	39,095,879	21.38
Total	36,989,957	20.31

References

1. Martin, K.C. and A. Ephrussi, *mRNA localization: gene expression in the spatial dimension.* Cell, 2009. **136**(4): p. 719-30.
2. Medioni, C., K. Mowry, and F. Besse, *Principles and roles of mRNA localization in animal development.* Development, 2012. **139**(18): p. 3263-76.
3. Cody, N.A., C. Iampietro, and E. Lécuyer, *The many functions of mRNA localization during normal development and disease: From pillar to post.* WIREs Dev Biol, 2013.
4. Berleth, T., et al., *The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo.* EMBO J, 1988. **7**(6): p. 1749-56.
5. Ephrussi, A., L.K. Dickinson, and R. Lehmann, *Oskar organizes the germ plasm and directs localization of the posterior determinant nanos.* Cell, 1991. **66**(1): p. 37-50.
6. St Johnston, D. and C. Nusslein-Volhard, *The origin of pattern and polarity in the Drosophila embryo.* Cell, 1992. **68**(2): p. 201-19.
7. Gavis, E.R. and R. Lehmann, *Localization of nanos RNA controls embryonic polarity.* Cell, 1992. **71**(2): p. 301-13.
8. Long, R.M., et al., *Mating type switching in yeast controlled by asymmetric localization of ASH1 mRNA.* Science, 1997. **277**(5324): p. 383-7.
9. Latham, V.M., Jr., et al., *Beta-actin mRNA localization is regulated by signal transduction mechanisms.* J Cell Biol, 1994. **126**(5): p. 1211-9.
10. Kislauskis, E.H., X. Zhu, and R.H. Singer, *beta-Actin messenger RNA localization and protein synthesis augment cell motility.* J Cell Biol, 1997. **136**(6): p. 1263-70.
11. Eberwine, J., et al., *Local translation of classes of mRNAs that are targeted to neuronal dendrites.* Proc Natl Acad Sci U S A, 2001. **98**(13): p. 7080-5.
12. Krzyzosiak, W.J., et al., *Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target.* Nucleic Acids Res, 2012. **40**(1): p. 11-26.
13. La Spada, A.R. and J.P. Taylor, *Repeat expansion disease: progress and puzzles in disease pathogenesis.* Nat Rev Genet, 2010. **11**(4): p. 247-58.
14. Prasanth, K.V., et al., *Regulating gene expression through RNA nuclear retention.* Cell, 2005. **123**(2): p. 249-63.

15. Hung, M.E. and J.N. Leonard, *A platform for actively loading cargo RNA to elucidate limiting steps in EV-mediated delivery.* Journal of Extracellular Vesicles, 2016. **5**.
16. Jambhekar, A., et al., *Unbiased selection of localization elements reveals cis-acting determinants of mRNA bud localization in Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 2005. **102**(50): p. 18005-10.
17. Tekotte, H. and I. Davis, *Intracellular mRNA localization: motors move messages.* Trends Genet, 2002. **18**(12): p. 636-42.
18. Bergalet, J. and E. Lecuyer, *The functions and regulatory principles of mRNA intracellular trafficking.* Adv Exp Med Biol, 2014. **825**: p. 57-96.
19. Lipshitz, H.D. and C.A. Smibert, *Mechanisms of RNA localization and translational regulation.* Curr Opin Genet Dev, 2000. **10**(5): p. 476-88.
20. Stoorvogel, W., et al., *The biogenesis and functions of exosomes.* Traffic, 2002. **3**(5): p. 321-30.
21. Montecalvo, A., et al., *Mechanism of transfer of functional microRNAs between mouse dendritic cells via exosomes.* Blood, 2012. **119**(3): p. 756-66.
22. Izumi, H., et al., *Bovine milk exosomes contain microRNA and mRNA and are taken up by human macrophages.* Journal of Dairy Science, 2015. **98**(5): p. 2920-2933.
23. Gould, S.J., A.M. Booth, and J.E. Hildreth, *The Trojan exosome hypothesis.* Proc Natl Acad Sci U S A, 2003. **100**(19): p. 10592-7.
24. Al-Nedawi, K., et al., *Intercellular transfer of the oncogenic receptor EGFRvIII by microvesicles derived from tumour cells.* Nat Cell Biol, 2008. **10**(5): p. 619-24.
25. Lakkaraju, A. and E. Rodriguez-Boulan, *Itinerant exosomes: emerging roles in cell and tissue polarity.* Trends Cell Biol, 2008. **18**(5): p. 199-209.
26. Taylor, D.D. and C. Gercel-Taylor, *Exosomes/microvesicles: mediators of cancer-associated immunosuppressive microenvironments.* Semin Immunopathol, 2011. **33**(5): p. 441-54.
27. Lefebvre, F.A., et al., *Comparative transcriptomic analysis of human and Drosophila extracellular vesicles.* Sci Rep, 2016. **6**: p. 27680.
28. Lefebvre, F.A. and E. Lecuyer, *Small Luggage for a Long Journey: Transfer of Vesicle-Enclosed Small RNA in Interspecies Communication.* Front Microbiol, 2017. **8**: p. 377.

29. Squadrito, M.L., et al., *Endogenous RNAs Modulate MicroRNA Sorting to Exosomes and Transfer to Acceptor Cells.* Cell Reports, 2014. **8**(5): p. 1432-1446.
30. Villarroya-Beltri, C., et al., *Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs.* Nature Communications, 2013. **4**.
31. Hagiwara, K., et al., *Commitment of Annexin A2 in recruitment of microRNAs into extracellular vesicles.* Febs Letters, 2015. **589**(24): p. 4071-4078.
32. Claude, A., *Fractionation of mammalian liver cells by differential centrifugation; experimental procedures and results.* J Exp Med, 1946. **84**: p. 61-89.
33. Jagannathan, S., C. Nwosu, and C.V. Nicchitta, *Analyzing mRNA localization to the endoplasmic reticulum via cell fractionation.* Methods Mol Biol, 2011. **714**: p. 301-21.
34. Lecuyer, E., H. Yoshida, and H.M. Krause, *Global implications of mRNA localization pathways in cellular organization.* Curr Opin Cell Biol, 2009. **21**(3): p. 409-15.
35. Bhatt, D.M., et al., *Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions.* Cell, 2012. **150**(2): p. 279-90.
36. Tilgner, H., et al., *Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs.* Genome Res, 2012. **22**(9): p. 1616-25.
37. Bahar Halpern, K., et al., *Nuclear Retention of mRNA in Mammalian Tissues.* Cell Rep, 2015. **13**(12): p. 2653-62.
38. Lefebvre, F.A., et al., *Biochemical Fractionation of Time-Resolved Drosophila Embryos Reveals Similar Transcriptomic Alterations in Replication Checkpoint and Histone mRNA Processing Mutants.* J Mol Biol, 2017.
39. Carlevaro-Fita, J., et al., *Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells.* RNA, 2016. **22**(6): p. 867-82.
40. Chen, Q., et al., *Hierarchical regulation of mRNA partitioning between the cytoplasm and the endoplasmic reticulum of mammalian cells.* Mol Biol Cell, 2011. **22**(14): p. 2646-58.
41. Wang, E.T., et al., *Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins.* Cell, 2012. **150**(4): p. 710-24.
42. van Heesch, S., et al., *Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes.* Genome Biol, 2014. **15**(1): p. R6.

43. Sterne-Weiler, T., et al., *Frac-seq reveals isoform-specific recruitment to polyribosomes.* Genome Res, 2013. **23**(10): p. 1615-23.
44. Martinez-Nunez, R.T. and J.R. Sanford, *Studying Isoform-Specific mRNA Recruitment to Polyribosomes with Frac-seq.* Methods Mol Biol, 2016. **1358**: p. 99-108.
45. Kopczynski, C.C., et al., *A high throughput screen to identify secreted and transmembrane proteins involved in Drosophila embryogenesis.* Proc Natl Acad Sci U S A, 1998. **95**(17): p. 9973-8.
46. Diehn, M., et al., *Large-scale identification of secreted and membrane-associated gene products using DNA microarrays.* Nat Genet, 2000. **25**(1): p. 58-62.
47. Diehn, M., et al., *Genome-scale identification of membrane-associated human mRNAs.* PLoS Genet, 2006. **2**(1): p. e11.
48. de Jong, M., et al., *Membrane-associated transcripts in Arabidopsis; their isolation and characterization by DNA microarray analysis and bioinformatics.* Plant J, 2006. **46**(4): p. 708-21.
49. Lerner, R.S., et al., *Partitioning and translation of mRNAs encoding soluble proteins on membrane-bound ribosomes.* Rna, 2003. **9**(9): p. 1123-37.
50. Marc, P., et al., *Genome-wide analysis of mRNAs targeted to yeast mitochondria.* EMBO Rep, 2002. **3**(2): p. 159-64.
51. Mercer, T.R., et al., *The human mitochondrial transcriptome.* Cell, 2011. **146**(4): p. 645-58.
52. Blower, M.D., et al., *Genome-wide analysis demonstrates conserved localization of messenger RNAs to mitotic microtubules.* J Cell Biol, 2007. **179**(7): p. 1365-73.
53. Sharp, J.A., et al., *Functional analysis of the microtubule-interacting transcriptome.* Mol Biol Cell, 2011. **22**(22): p. 4312-23.
54. Mili, S., K. Moissoglu, and I.G. Macara, *Genome-wide screen reveals APC-associated RNAs enriched in cell protrusions.* Nature, 2008. **453**(7191): p. 115-9.
55. Job, C. and J. Eberwine, *Localization and translation of mRNA in dendrites and axons.* Nat Rev Neurosci, 2001. **2**(12): p. 889-98.
56. Moccia, R., et al., *An unbiased cDNA library prepared from isolated Aplysia sensory neuron processes is enriched for cytoskeletal and translational mRNAs.* J Neurosci, 2003. **23**(28): p. 9409-17.

57. Poon, M.M., et al., *Identification of process-localized mRNAs from cultured rodent hippocampal neurons.* J Neurosci, 2006. **26**(51): p. 13390-9.
58. Zivraj, K.H., et al., *Subcellular profiling reveals distinct and developmentally regulated repertoire of growth cone mRNAs.* J Neurosci, 2010. **30**(46): p. 15464-78.
59. Gumy, L.F., et al., *Transcriptome analysis of embryonic and adult sensory axons reveals changes in mRNA repertoire localization.* RNA, 2011. **17**(1): p. 85-98.
60. Khaladkar, M., et al., *Subcellular RNA sequencing reveals broad presence of cytoplasmic intron-sequence retaining transcripts in mouse and rat neurons.* PLoS One, 2013. **8**(10): p. e76194.
61. Sasaki, Y., et al., *Identification of axon-enriched microRNAs localized to growth cones of cortical neurons.* Dev Neurobiol, 2014. **74**(3): p. 397-406.
62. Taliaferro, J.M., et al., *Distal Alternative Last Exons Localize mRNAs to Neural Projections.* Mol Cell, 2016. **61**(6): p. 821-33.
63. Dignam, J.D., R.M. Lebovitz, and R.G. Roeder, *Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei.* Nucleic Acids Res, 1983. **11**(5): p. 1475-89.
64. Thery, C., et al., *Isolation and characterization of exosomes from cell culture supernatants and biological fluids.* Curr Protoc Cell Biol, 2006. **Chapter 3**: p. Unit 3 22.
65. Vedeler, A., I.F. Pryme, and J.E. Hesketh, *The characterization of free, cytoskeletal and membrane-bound polysomes in Krebs II ascites and 3T3 cells.* Mol Cell Biochem, 1991. **100**(2): p. 183-93.
66. Fabio Alexis Lefebvre, L.P.B.B., Julie Bergalet, Eric Lécuyer, *Data for the Generation of RNA Spatiotemporal Distributions and Interpretation of Chk1 and SLBP Protein Depletion Phenotypes during Drosophila Embryogenesis.* Data in Brief, 2017.
67. Chomczynski, P. and N. Sacchi, *Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction.* Anal Biochem, 1987. **162**(1): p. 156-9.
68. Simoes, A.E., et al., *Efficient recovery of proteins from multiple source samples after TRIzol((R)) or TRIzol((R))LS RNA extraction and long-term storage.* BMC Genomics, 2013. **14**: p. 181.

69. Hill, A.F., et al., *ISEV position paper: extracellular vesicle RNA analysis and bioinformatics.* J Extracell Vesicles, 2013. **2**.
70. Kalra, H., et al., *Vesiclepedia: a compendium for extracellular vesicles with continuous community annotation.* PLoS Biol, 2012. **10**(12): p. e1001450.
71. Keerthikumar, S., et al., *ExoCarta: A Web-Based Compendium of Exosomal Cargo.* J Mol Biol, 2016. **428**(4): p. 688-92.
72. Uhlen, M., et al., *Proteomics. Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.
73. Townsley, F.M., D.W. Wilson, and H.R. Pelham, *Mutational analysis of the human KDEL receptor: distinct structural requirements for Golgi retention, ligand binding and retrograde transport.* EMBO J, 1993. **12**(7): p. 2821-9.
74. Mogensen, M.M., et al., *Microtubule minus-end anchorage at centrosomal and non-centrosomal sites: the role of ninein.* J Cell Sci, 2000. **113 (Pt 17)**: p. 3013-23.
75. Schmittgen, T.D. and K.J. Livak, *Analyzing real-time PCR data by the comparative C(T) method.* Nat Protoc, 2008. **3**(6): p. 1101-8.
76. Rastan, S., *X chromosome inactivation and the Xist gene.* Curr Opin Genet Dev, 1994. **4**(2): p. 292-7.
77. Piekny, A.J. and A.S. Maddox, *The myriad roles of Anillin during cytokinesis.* Semin Cell Dev Biol, 2010. **21**(9): p. 881-91.
78. Valadi, H., et al., *Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells.* Nat Cell Biol, 2007. **9**(6): p. 654-9.
79. Balaj, L., et al., *Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences.* Nat Commun, 2011. **2**: p. 180.
80. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**(10): p. R106.
81. Criscuolo, A. and S. Brisse, *AlienTrimmer removes adapter oligonucleotides with high sensitivity in short-insert paired-end reads. Commentary on Turner (2014) Assessment of insert sizes and adapter content in FASTQ data from NexteraXT libraries.* Front Genet, 2014. **5**: p. 130.
82. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.* Nat Protoc, 2012. **7**(3): p. 562-78.

83. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-30.
84. Rosenbloom, K.R., et al., *The UCSC Genome Browser database: 2015 update.* Nucleic Acids Res, 2015. **43**(Database issue): p. D670-81.
85. Graffelman, J. and J.M. Camarena, *Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot.* Hum Hered, 2008. **65**(2): p. 77-84.
86. Cédric Diot, A.C., Éric Lécuyer, *Visualizing RNA localization dynamics in Drosophila embryos, tissues and cultured cells.* Methods, 2017.

Préface au chapitre 6

Ce chapitre est présenté sous la forme d'un article de recherche non-publié, destiné au journal *RNA*. La soumission est prévue après le dépôt de la présente thèse.

L'article détaille l'application de l'approche CeFra-seq pour tenter de comprendre les liens entre la localisation subcellulaire et le ciblage des ARNs aux vésicules extracellulaires issues de la lignée K562. Cette approche permet de mettre en évidence diverses propriétés qui corrèlent avec le ciblage extracellulaire, notamment l'accessibilité cytosolique. Nous avons également identifié des déterminants de séquence enrichis au sein des miRNAs prédomiants des vésicules.

Dans le contexte de la présente thèse, les résultats exposées dans cet article font suite au développement de l'approche CeFra-seq (Chapitre 5).

J'ai écrit l'intégralité du manuscrit. J'ai effectué les analyses liées aux vésicules extracellulaires ainsi que la préparation des figures présentées dans cette publication. Neal Cody, un ancien stagiaire postdocotoral du laboratoire Lécuyer, a fortement contribué au développement de l'approche CeFra-seq et à son application sur les lignée K562. Louis Philip Benoit Bouvrette, étudiant au PhD au laboratoire, a contribué au développement des approches bio-informatique associées à la méthode CeFra-seq, notamment les étapes de transformation de données liées aux applications graphiques. Dr Éric Lécuyer, mon superviseur, a contribué à l'étude en amassant les financements requis et en relisant le texte pour y suggérer diverses améliorations.

Chapitre 6 : Transcriptomic Profiling of Extracellular Vesicles and Subcellular Fractionation Reveals Strong Associations with the Cytosolic Repertoire (Article #4)

Chapitre 6 : Article #4

Transcriptomic Profiling of Extracellular Vesicles and Subcellular Fractionation Reveals Strong Associations with the Cytosolic Repertoire

Fabio Alexis Lefebvre[1,2], Neal Cody[1], Louis Philip Benoit Bouvrette[1,2], Éric Lécuyer[1,2,3,4]

1- Institut de Recherches Cliniques de Montréal (IRCM), Montréal, QC H2W 1R7, Canada
2- Département de Biochimie, Université de Montréal, Montréal, QC H3T 1J4, Canada
3- Division of Experimental Medicine, McGill University, Montréal, QC H3A 1A3, Canada

4-Corresponding author:

Éric Lécuyer (eric.lecuyer@ircm.qc.ca)

Abstract

Cells release membrane-bound extracellular vesicles (EVs) enclosing complex and distinctive protein and RNA repertoires. However, the mechanisms involved in sorting cellular RNAs to EVs remain poorly defined. To glean insights on these processes, we systematically characterized the molecular cargo of EVs released by K562 cells through protein mass spectrometry and deep sequencing of size-selected RNA populations. We then comprehensively contrasted EV populations with subcellular repertoires of asymmetrically distributed proteins and transcripts inferred from biochemical cell fractionation (CeFra-seq). Our analyses show that EV RNA populations strongly relate to the subcellular cytosolic repertoire. Full-length and fragmented polymerase III transcripts, such as Y RNA and vault RNA, account for a sizeable proportion of the RNA population in EVs and are highly enriched in the cytosolic fraction. In addition, we uncovered a vast repertoire of full-length mRNA and miRNA signatures in EVs, collectively exhibiting low cumulative abundance. The most enriched mRNAs in EVs were overrepresented in nuclei, possibly underlying regulated EV targeting resulting in a short cytoplasmic half-life. The GGGUUG hexanucleotide was overrepresented among EV-enriched miRNAs, which were broadly distributed at the subcellular level. Altogether, our analyses argue for a semi-selective model of RNA incorporation in EVs. Cytosolic enrichment is a strong predictor of protein and small RNA incorporation in EVs, while sequence determinants and a short cytoplasmic half-life appear to promote the export of specific miRNA and mRNA.

Introduction

The localization of specific mRNAs to subcellular domains and compartments of eukaryotic cells represents a refined layer of post-transcriptional regulation involved in shaping the local proteome. Indeed, localization of mRNAs and spatially-resolved translation are instrumental in orchestrating a wide range of physiological events linked to local protein expression, including mating type determination in yeast [1] [2], chemotaxis in fibroblasts [3], oogenesis and embryogenesis in *Drosophila* and *Xenopus* [4], or synaptogenesis in mammalian neurons [5]. While seminal studies first revealed the crucial roles of a handful of mRNAs, such as *gurken*, *bicoid* and *oskar*, in the establishment of morphogen gradients [6] [7], more recent work has emphasized the contribution of RNA localization signals in directing the translation of membrane-associated and secreted factors near specific organelles, such as the endoplasmic reticulum [8] [9]. These observations suggest that the regulatory contributions of mRNA localization may be broader than first expected, a view supported by a few large high-throughput studies [4] [10]. In addition, recent studies have revealed that regulated subcellular localization also prevails among microRNAs (miRNA), a vast and conserved class of short transcripts (22 nt) that mediate gene expression regulation through RNA degradation and translational inhibition. Indeed, a short motif at the 3'end of miR-29b has been linked to the nuclear accumulation of the mature transcript and a similar localization signal may mediate the shuttling of miR-1 in mitochondria [11]. Similarly, long non-coding RNAs (lncRNA), which can assume a wide range of functions [12], have revealed a plethora of gene-specific localization patterns, including diverse subnuclear domains and ribosomal associations [13, 14] [15]. A series of meticulous mutagenesis assays have established the role of *cis*-acting nucleotide motifs, often found in the 3' untranslated region (3'UTR), in localizing specific mRNAs. Typically, such *cis*-acting elements specifically interact with RNA-binding proteins (RBPs), which behave as *trans*-acting factors. RBPs can modulate the diffusion of bound mRNAs, exhibit active subcellular motility through cytoskeletal association or selectively protect populations of mRNAs from degradation [16], resulting in the asymmetric spatial distribution of the associated transcripts.

Over the last decade, a regain of interest in profiling secreted material has extended the realm of RNA localization beyond the boundaries of the cell and into the surrounding extracellular space. Indeed, extracellular vesicles (EVs) have emerged as vehicles of intercellular communication and are envisioned as promising disease biomarkers that could facilitate the

accession of key diagnostic and prognostic signatures [17] [18] [19]. Extensive profiling efforts have revealed that EV populations, which encompass secreted exosomes, membrane-shed microvesicles and apoptotic bodies, enclose complex repertoires of bioactive gene products, including nucleic acids and proteins. These populations distinctively differ from their cellular counterparts [20] and can undergo profound alterations in response to cell fate commitment or disruptions in environmental status and stimuli, such as hypoxia or osmotic stress. The specificity and inherent plasticity of EV RNA repertoires elicits the prevalence of sequence- or structure-specific determinants in the EV targeting processes[21, 22]. Yet, deciphering the regulatory mechanisms that account for this specificity have remained challenging and multiple determinants appear to drive the incorporation of RNAs in EVs. In lymphoblasts, sumoylation of a nuclear RBP, hnRNPA2B1, was shown to promote its targeting to exosomes, in conjunction with its associated miRNA population, which bears the *cis*-acting motif GGAG. Similarly, the calcium-binding, membrane-associated protein Annexin A2 can bind miRNAs and has been envisioned as a key *trans*-acting factor in EV loading. One study identified a 25 nt motif within the 3'UTR of mRNAs overrepresented in EVs derived from glioblastoma and melanoma cell lines[23]. The functionality of the sequence was later confirmed through mutagenesis coupled to reporter assays. By contrast, a second study [24] attempting to define *cis*-acting motifs concluded that multiple alignments and position-specific scoring approaches failed to identify shared signatures among EV RNAs, concluding that short half-lives were a stronger predictor of EV targeting than sequence identity. Similarly, RNA length was shown to strongly modulate targeting efficiency through the development of an ingenious tailored approach based on the MS2-GFP system, which confirmed that long sequences (≥1.5 kb) are poorly loaded into EVs, independently of their nucleotide composition [135].

Since the advent of *in situ* hybridization in 1968 [25], multiple microscopy approaches have been developed to image RNA transcripts in fixed tissues and in living cells. In addition, cell fractionation combined with high-throughput gene expression profiling has offered an insightful alternative to map asymmetrically distributed RNA populations [26]. Such approaches have been used to identify transcripts targeted to different organelles and subcellular compartments, including the nucleus [27-30], cytoplasm [27, 29, 31], cytosol [15, 28, 32, 33], polysomes [15, 34, 35], the endoplasmic reticulum [36-40], the membrane and insoluble fractions [33], mitochondria [41, 42], pseudopodia [43], microtubules [44, 45], and neuronal projections [46-54]. In addition, biochemical purification coupled to high throughput analyses such as RNA microarray or deep-sequencing have remained the main options to systematically

investigate the transcriptomic composition of EVs. Here, we aimed to glean insights into the routes governing RNA loading to EVs through a comparative analysis of subcellular and extracellular RNA repertoires. We first purified EVs from K562 cells, assessed their structural properties through nanoparticle-tracking analysis (NTA) and transmission electron microscopy (TEM) and systematically profiled their molecular content through proteomic and transcriptomic analyses of size-selected small (<40 nt) and long (>40 nt) RNA populations. We next applied a subcellular fractionation scheme termed CeFra-seq to define the composition of four subcellular extracts consisting of asymmetrically distributed material: the cytosolic, cytoplasmic insoluble, endomembrane and nuclear fractions [55]. Datasets of subcellular and extracellular RNAs were then extensively contrasted through comparative, correlative and co-clustering approaches. Our correlative analyses strongly emphasize the cytosolic identity of proteins and small RNAs enriched in EVs. We show that full-length and 3' end fragments of Y RNAs and vault RNAs, which figure among the most enriched EV transcripts, are nearly exclusive to the cytosolic fraction. While protein-coding were collectively rare in EVs, we identified a handful of full-length mRNAs enriched in EVs, which predominantly displayed a nuclear localization in cells. Finally, we show that the EV miRNA repertoire is varied, but collectively accounts for less than 5% of total sequencing reads in the small RNA-seq library derived from EVs. The most abundant EV miRNAs showed ubiquitous cellular localization. Independently of localization, hierarchical clustering broadly recapitulated EV targeting and identified the hexanucleotidic sequence GGGUUG as enriched in miRNAs targeted to EVs using two independent algorithms. Our work shows that regulated sorting, stochastic dynamics and cytosolic accessibility account for to EV encapsulation in the K562 model, suggesting that RNA export is a complex and multifactorial process.

Results

CeFra-seq enables the purification of extracellular and subcellular material

We sought to investigate the subcellular distribution properties of RNA transcripts targeted to extracellular vesicles. To systematically map transcriptome-wide subcellular and extracellular distributions, we devised a biochemical subcellular fractionation adapted from previous protocols and submitted each extract to proteomic and transcriptomic analysis. Our approach, termed CeFra-seq [33, 56-59], has been described in detail before with relevant validation steps and high throughput data transformation guidelines (**Figure 6.1A**). Briefly, EVs released in conditioned cell media over a timeframe of 48h are collected through a sequential ultracentrifugation approach. Then, cells undergo a mild hypotonic lysis involving Dounce homogenization and nuclei are recovered by centrifugation and purification through a sucrose cushion ultracentrifugation. The resulting cytoplasmic supernatant is further processed to yield a highly soluble extract, the cytosolic fraction. The pellet is then solubilized using a detergent and further processed by ultracentrifugation to yield a re-solubilized part, the endomembrane fraction, and a cytoplasmic insoluble pellet. Total RNA and protein are ultimately recovered through a guanidium thiocyanate-phenol-chloroform extraction [60, 61].

EV preparations were first characterized by transmission electron microscopy (TEM) and nanoparticle tracking analysis (NTA). TEM revealed a predominance of cup-shaped membranous structures with heterogeneous size ranging from 30 to 200 nm, which were quantified through direct measurements of the micrographs (**Figure 6.1B,C**). EV pellets were also suspended in PBS and analyzed by NTA along with the appropriate controls, providing both the diameter and the number of particles in solution (**Figure 6.1D**). The size distributions inferred from TEM and NTA largely overlapped, although the mean diameter measured by NTA (μ=206.5 ± 2.2 nm) was higher than by TEM (μ=139.1 ± 7.7 nm), consistent with reports of EV collapsing upon uranyl acetate staining [57]. We then extracted total protein and RNA of standardized EV collections and evaluated the amounts retrieved through spectrophotometric quantifications, both at the protein (μ=161.4 ± 20.2 µg) and RNA (170.4 ± 40.6 ng) levels (**Figure 6.1E**). In addition, we took advantage of the EV counts inferred from NTA data to derive quantitative estimates of average molecular cargos counts per individual EV. NTA identified (4.38 ± 0.10) × 10^{11} EVs per preparation, which corresponds to a release rate of (5.48

± 0.98) × 10^3 EVs per cell over the course of 48h when postulating homogenous release, ideal yields and absence of dynamic intercellular EV exchange or degradation. Our calculations show that a single, idealized EV contains (7.80 ± 1.06) × 10^{-17} g of proteins, which broadly corresponds to 1,250 ± 170 single peptides of 37.5 kDa, the median mass of the human proteome [62], or 470 ± 64 peptides of 100 kDa. At the RNA level, we calculated a release rate of (1.66 ± 0.57) × 10^{-15} g of EV-associated RNA per cell over 48h, pointing to an average RNA mass of (3.02 ± 0.68) × 10^{-19} g per EV, equivalent to 6 ± 1 molecules of a 100 nt-long RNA, or 11 ± 3 molecules of a 50 nt-long RNA. While these estimates should be considered with caution due to the inherent limitations linked to both NTA and spectrophotometric quantifications of macromolecules, our results suggest that EVs contain approximately 100 times more individual proteins than RNA transcripts.

The integrity of subcellular extracts corresponding to nuclear (N), cytosolic (C), cytoplasmic insoluble (I) and endomembrane (M) material was validated by profiling the abundance of specific proteins and RNA transcripts which have previously been reported to display asymmetric spatial distributions. Protein extracts were submitted to SDS-PAGE followed by western blotting (**Figure 6.S1**). This analysis showed that histone H3 is strongly enriched in the nuclear fraction, whereas monomeric α-tubulin was most abundant in the cytosolic fraction. In addition, factors bearing the tetrapeptide motif KDEL, which is linked to ER targeting, [72] were enriched in the endomembrane fraction and Ninein, a centrosomal regulator of microtubule minus-end anchoring [73] was enriched in the cytoplasmic insoluble fraction. At the RNA level, qRT-PCR confirmed the strong nuclear predominance of the lncRNA *XIST*, involved in chromosome X silencing, while histone H3 mRNA was significantly enriched in the cytosolic fraction. By contrast, (**Figure 6.S2**).

Proteomic analyses link EVs to the cytosolic fraction

The electrophoretic profiles of total protein extracts obtained from integral cell lysates and from EVs were visualized through an SDS-PAGE followed by silver staining, which revealed that the content of EVs deviates markedly from its parental cell counterpart (**Figure 6.2A**). We conducted a liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis to compare the abundance distributions (spectral counts) of proteins identified in EV and in total cell lysates (CL). We found weak positive linear (Pearson) and monotonic (Spearman)

correlations (Pearson's r=0.46 $P<10^{-4}$; Spearman's ρ=0.61 $P<10^{-4}$; **Figure 6.2B-C**), supporting the idea that the EV proteome differs strongly from its cellular counterpart.

β-actin (*ACTB*) was the most abundant protein in EVs (**Table 6.1**), while the most enriched was Coronin 1A (*COR1A*), one of several EV-associated actin-binding factors. Diverse established EV markers displayed robust enrichments, notably Flotilin-1 (*FLOT1*) and small GTPases encoded by *RAB10*, *RAB1A* and *RAP1B* [63]. Consistent with previous reports [22], metabolic enzymes such as lactate dehydrogenases (*LDHA*, *LDHB*) or glutathione S-transferases (*GSTA1*, *GSTA5*, *GSTP1*) were present in EVs, along with heat shock proteins (*HSP7C*, *HSP90A*, *HSP90B*) and elongation factors (*EF1A1*, *EF2*). The tetraspanin family member *TSPAN14* classically associated with exosomes was traced in EVs but undetected in cells. The endoplasmic reticulum marker Calnexin or the Golgi apparatus marker GM130 were not found in EVs, consistent with the absence of subcellular contamination [63]. Gene ontology (GO) enrichment analyses confirmed the global profile of EV-associated proteins, retrieving terms such as "membrane-bounded vesicles", "endocytic vesicles", "extracellular region part", "GTPase activity" or "actin skeleton organization" (**Figure 6.S3**).

Our LC-MS analysis of subcellular extracts recovered over 1,000 high confidence protein signatures in each fraction. The ten most abundant proteins of each fraction were markedly different, confirming the proficiency of our protocol at recovering asymmetric preparations (**Table 6.S1**). The cytosolic fraction was dominated by cytoskeletal monomers (e.g. *ACTB*, *TUBB*, *TUBB4B*), along with metabolic factors (e.g. *ALDOA*, *GAPDH*, *ENO1*) and heat shock proteins (e.g. *HSP90AB1*, *HSP90AA1*). The insoluble fraction exhibited a strong ribosomal signature, as 15 of its 20 most abundant factors were 40S and 60S components (**Table 6.S1**). Initiation and elongation factors (e.g. *EIF5A*, *EEF1G*, *EIF2S1*) and several RNA-binding proteins, (eg. *SERBP1*, *PTBP1*, *PABPC1*) were also abundant in the cytoplasmic insoluble fraction. The endomembrane fraction largely consisted of proteasomal components (e.g. *PSME2*, *PSMC2*, *PSMA1*), markers of the endoplasmic reticulum (e.g. *CALR*, *VCP*, *CANX*) and 40S ribosomal proteins (e.g. *RPS19*, *RPS16*). Histones were most abundant in the nuclear fraction, along with Prelamin-A and diverse splicing factors (e.g. *SF3B5*, *SFPQ*, *SPF27*). To contrast the functional signature of each sample, we retrieved enriched gene ontology (GO) terms associated to factors found exclusively in each fraction (**Table 6.S2**).

Approximately half of the EV repertoire (125 proteins) could not be traced in any subcellular fraction, although higher coverage was obtained for all fractions. The cytosolic fraction had 114 proteins in common with EVs (41% of the EV repertoire), while 102 EV proteins (36%) were traced in the membrane fraction, 75 (27%) were present in the insoluble fraction and 60 (21%) in the nuclear fraction (**Figure 6.2D).** We then determined global correlations of protein distributions in EVs and subcellular fractions. Pair-wise correlation analyses we used to contrast protein abundance, expressed as spectral counts (**Figure 6.2F**), and protein-specific enrichment ratios (**Figure 6.2G**), corresponding to the fraction of cumulative abundance associated to a given fraction, for example $Ratio_C=C_{Counts}/(E_{Counts}+C_{Counts}+I_{Counts}+M_{Counts}+N_{Counts})$. The cytosolic fraction showed the strongest association with EVs, as quantified by both Pearson's ($r=0.64$ $P<10^{-4}$) and Spearman's ($\rho=0.42$ $P<10^{-4}$) coefficients. The membrane fraction, which largely corresponds to ER material, showed a modest linear correlation ($r=0.33$ $P<10^{-4}$) but no rank correlation (0.05 *NS*), a metric described as more robust and less sensitive to outliers than Pearson's correlation[64]. Similarly, the cytoplasmic insoluble ($r=0.08$ *NS* $\rho=0.17$ *NS*) and nuclear ($r=0.20$ $P=0.002$; $\rho=0.37$ $P=0.002$) repertoires related poorly to the EV protein distribution.

Transcriptomic analyses link EVs to the cytosolic fraction

The size distribution profiles of total RNA extracted from EVs, integral cell lysates and subcellular fractions were determined by bioanalyzer capillary electrophoresis (**Figure 6.3A)**. In accordance with previous reports [20, 21], EVs contained very low peaks of the full-length *18S* and *28S* rRNA signatures and were enriched in short sequences (20-250 nt). The cytosolic distribution resembled that of EVs, containing low amounts of mature *18S* and *28S* rRNAs and a strong enrichment in short transcripts, whereas all the other subcellular profiles were dominated by rRNA peaks. To capture this diversity, we chose to profile size-selected populations and sequenced small (<40 nt) and standard (>40 nt) RNA libraries obtained from biological duplicates of EV, total cellular and subcellular material, and mapped 12 to 20 million reads per sample (**Table 6.2).** Read length distributions in the small RNA library were markedly different across each fraction (**Figure 6.3B**). While the total cell distribution displayed two local maxima at 22 and 24 nt, 23 nt-long sequences were most abundant in EVs. In addition, a group of longer reads centered at 32 nt was observed in EVs but undetected in cells. A peak corresponding to

short sequences centered at 18 nt dominated cytosolic distributions, along with an additional peak at 22 nt, which was also prevalent in EVs. In addition, a scarce population was present at 31 nt in both EV and cytosolic profiles.

We first considered the biotypes of RNA species mapped in each fraction to broadly contrast their content (**Figure 6.3C**). Subcellular fractions exhibited highly divergent small RNA biotype profiles, suggesting that short transcripts display a strong spatial asymmetry. Fragments of miscelleanous RNAs, which include RNA polymerase III products such as Y RNAs and vault RNAs, were dominant in the EV and cytosolic fractions. Meanwhile, miRNAs were the most abundant class in the insoluble (40.1%) and endomembrane (41.5%) extracts but were underrepresented in EVs (4.6%), the cytosolic (6.2%) and the nuclear (3.7%) populations. Fragments of snRNAs accounted for most reads in the nuclear fraction (78.3%). By contrast, the populations mapped in the standard RNA-seq library were all dominated by mRNA reads, especially the insoluble (97.5%), endomembrane (94.5%) and nuclear (92.0%) fractions. The cytosolic population showed lower levels of mRNAs (75.2%) and harbored abundant non-coding species (21.7%), reflecting high levels of infrastructural RNAs such as *RPPH1* and *RN7SL*. This profile is reminiscent of the long RNA population found in EVs (**Figure 6.3D**).

Correlative analyses showed that small RNA populations of EVs and total cell extracts differ considerably, as attested by low correlation coefficients (**Figure 6.3D**; r=0.24 $P<10^{-4}$; ρ=0.58 $P<10^{-4}$). However, the EV long RNA population was a closer reflection of cellular pools (**Figure 6.3E**; r=0.75 $P<10^{-4}$; ρ=0.68 $P<10^{-4}$). Pair-wise correlative analyses were used to contrast small RNA abundance, expressed as FPKM (**Figure 6.3E**), and transcript-specific cytotopic ratios (**Figure 6.3F**), defined as the fraction of cumulative abundance associated to a given fraction, for example $Ratio_C=C_{FPKM}/(E_{FPKM}+C_{FPKM}+I_{FPKM}+M_{FPKM}+N_{FPKM})$. The cytosolic distribution of small RNA showed the strongest link with EVs, exhibiting strong Spearman (r=0.73 $P<10^{-4}$) and Pearson correlations (**Figure 6.3F** ; ρ = 0.76 $P<10^{-4}$). This result was confirmed by the analysis of enrichment ratios : the cytosolic extract was the only fraction positively correlated with EV material (**Figure 6.3G** ; r=0.21 $P<10^{-4}$; ρ=0.05 $P<10^{-4}$). Among full-length transcripts mapped in the standard RNA-seq library, the cytoplasmic insoluble fraction displayed the strongest ordinal correlation with EVs (ρ=0.66 $P<10^{-4}$), while the correlations of enrichment ratios were negative for all comparisons involving EVs (**Figure 6.3H**). Collectively, RNA electrophoretic profiles, read length distributions, biotype composition

and correlative analyses suggest that the cytosolic material shares strong similarities with EVs, reflecting the enrichment of fragments of miscellaneous transcripts.

Y RNAs and vault RNAs are enriched in EVs and in the cytosolic fraction

We focussed our analyses on Y RNA sequences, the most abundant transcript in EVs, accounting for 84.4% of total reads mapped in the small RNA library (**Figure 6.3I**). Y RNAs form a group of short (88-110 nt), polymerase III-dependent transcripts known to bind the autoimmune antigen proteins Ro and La, forming ribonucleoproteic complexes thought to function in genomic DNA repair and replication. Exceptionally high EV and cytosolic coverage was noted for a fragment of 22 to 24 nt (CCCCCCACTGCTAAATTTGACTGG) mapped to the 3' end of the *RNY4* gene[65]. Similarly, full-length coverage of the *RNY4* locus on chromosome 7 displayed strong enrichments in EVs and in the cytosolic fraction but were scarce in every subcellular fraction (**Figure 6.4A,C**). A similar portrait prevailed with all four vault RNAs (vRNA) (**Figure 6.4B,D**). These RNA pol III-dependent transcripts (88-100 nt) are components of the large cytoplasmic barrel-shaped vault RNPs [66]. Vault RNAs have additionally been described as precursors of shorter miRNA-like species termed small vault RNAs (svRNAs), involved in the regulation of cytochrome factors [67]. Full-length *VTRNA1-1* and a 23 nt-long fragment mapped to the 3' end of the gene were abundant and enriched in the EV and cytosolic populations, while scarce in other fractions (**Figure 6.4B,D**). Transcriptome-wide MA plots comparing the repertoires of EVs and total cell extracts confirmed that Y RNAs and vRNAs are among the most enriched EV transcripts in both the small and standard RNA-seq libraries (**Figure 6.4E-F**). To visualize the distribution of these EV-enriched transcripts through the subcellular fractions, we generated histograms (**Figure 6.4G-H**) and simplex plots (**Figure 6.4I-J**) of their cytotopic distributions, which confirmed strong cytosolic enrichments.

EVs contain complex protein-coding repertoires enriched with nuclear mRNAs

The analysis of long RNA populations mapped in the standard RNA-seq library uncovered an abundant and diversified protein-coding repertoire in EVs (>5000 mRNAs), largely consisting of full-coverage, exonic signatures. Correlative analyses showed that EV mRNA distributions resembled the cytoplasmic insoluble (**Figure 6.5A**; $r=0.79$ $\rho=0.73$ $P<10^{-4}$),

the endomembrane (r=0.82 ρ=0.64 $P<10^{-4}$) and the nuclear identity (r=0.75 ρ=0.68 $P<10^{-4}$). The mRNA repertoire of the cytosolic fraction showed markedly weaker associations with EVs (r=0.65 ρ=0.51 $P<10^{-4}$). The most abundant mRNAs in EVs were highly expressed cellular transcripts, such as those encoding the elongation factors EEF2 and EEF1A1, β-actin or ferritin chains (FTL, FTH) (**Table 6.4**). These analyses suggest that the EV mRNA repertoire may largely reflect stochastic incorporation of abundant cellular products. Nevertheless, an MA plot analysis showed that a few specific mRNAs were highly enriched in EVs, consistent with regulated targeting (**Figure 6.5B**). Read coverage at the *PRKAB1* and *CIT* locus on chromosome 12 provides an example of selective mRNA targeting (**Figure 6.5C**): the *PRKAB1* signature was abundant in all subcellular fractions while the flanking *CIT* gene was scarce by comparison, especially in the cytosolic and membrane fractions. In EVs, *CIT* was markedly more abundant than *PRKAB1*, consistent with regulated sorting. We focussed on the 10 most enriched and the 10 most abundant EV mRNAs and considered their subcellular distribution using simplex plots and histograms (**Figure 6.5D-G**). Functionally, the proteins encoded by EV-enriched mRNAs show heterogeneous profiles: *TACC2* encodes a microtubule-interacting protein, while PCSK5 is a secreted factor resident in the trans-Golgi network and KDM6B is a lysine demethylase. Most of these mRNAs have previously been identified in EVs released by different human models [22]. Their nuclear enrichment could reflect a short cytoplasmic half-life consisting of active transport for EV incorporation following a lengthy nuclear processing phase.

EV miRNAs display sequence and localization similarities

Reads gathered in the small RNA libraries were mapped to the Ensembl gene library and the miRdeep repertoire. Over 900 signatures were recovered, although they collectively accounted for only 4.6% of total small RNA reads. Correlative analyses showed that miRNA abundance distributions are related in EVs and in all subcellular fractions, as determined by both monotonic ($0.78 \leq \rho \leq 0.82$ $P<10^{-4}$) and linear ($0.82 \leq r \leq 0.92$ $P<10^{-4}$) metrics (**Figure 6.6A**). Two proto-oncogenic transcripts abundant in K562 cells, miR-486-5p and miR-92a-3p, accounted together for 65% of the total EV miRNA read distribution (**Figure 6.6B-C**). Interestingly, the 10 most abundant EV miRNAs displayed highly similar subcellular distributions, and were traced in every fraction (**Figure 6.6D-F**). Despite strong global abundance correlations with subcellular material, several rare cellular miRNAs were highly

enriched in EVs, suggesting that sequence-specific targeting is involved in sorting at least a subset of EV miRNAs. We retrieved the nucleotide sequences of two groups of 40 miRNAs, respectively corresponding to the highest and lowest EV enrichment ratios ($E=EV_{CPM}/Cell_{CPM}$). Multiple sequence alignment broadly segregated EV-enriched and EV-depleted sequences in independent groups (**Figure 6.6G**). EV-enriched sequences were submitted to two independent *de novo* motif search algorithms [68, 69], each emphasizing a GGGUUG hexanucleotide (**Figure 6.6H**).

Discussion

Specific protein and RNA signatures are released in extracellular vesicles, providing promising circulating biomarkers and carrying instructive signals involved in cell-cell communication. However, the regulatory principles involved in sequence-specific targeting and subcellular transport of RNA to EVs remain elusive. To provide insights into the determinants of RNA targeting to EVs, we compared the protein and RNA content of EVs to four extracts of asymmetrically distributed subcellular material. We show that the EV transcriptome largely consists of short fragments derived from non-coding RNAs, mostly Y RNAs. The small RNA content of EVs strongly and selectively reflects the cytosolic fraction, characterized by short RNAs, mostly fragments of infrastructural, polymerase III transcripts. Complex repertoires of full-coverage mRNAs and miRNAs were also identified, some displaying high enrichments, although they collectively account for a marginal fraction of the EV RNA population. EV-enriched mRNAs were overrepresented in the nuclear fraction, possibly reflecting short cytoplasmic half-life tied to rapid EV targeting after their nuclear maturation.

Our quantitative assessments suggest that 5 to 14 copies of predominantly small (25-250 nt) RNAs are associated with an average EV, along with 400 to 1400 intraluminal and membrane-spanning proteins. EV quantifications provide a critical framework to consider putative contributions in intercellular communication. However, our assessments are approximate, since they rely on NTA data in conjunction with photometric quantifications of RNA and protein concentrations, methods that bear several inherent limitations. Emergent nucleic acid technologies such as digital PCR platforms[70] or the NanoString nCounter system[71] could provide a more resolved view of gene-specific EV RNA tolls. Despite these limitations, our RNA counts are consistent with the broad conclusions of a recent quantitative study, which profiled EVs from multiple sources and found a rate of 0.00825 ± 0.02 molecules/EV for the most abundant miRNA in each EV type [72]. Although numerous miRNAs were traced in our preparations, they collectively accounted for few sequencing reads. Based on this data, we derive a count of 0.089 ± 0.04 molecules/EV for the most abundant miRNA, miR-486-5p. We found that the sequence GGGUUG was overrepresented in EV-enriched miRNAs, consistent with previous reports of sequence preferences in EV targeting [73] [74].

While they clearly harbor highly enriched signatures, EV RNA populations are also dazzlingly diverse: more than 10,000 RNAs were traced in our population analyses, while only 5 to 14 RNA molecules are associated to a single EV. Although the design of this study doesn't enable absolute quantitative comparisons of small and long RNA abundance, electropherograms suggest that long RNAs (≥250 nt) are scarce in EVs. We can thus infer that most single vesicles contain fragmented and full-length infrastructural RNAs such as Y RNAs and rRNAs, while incorporation of miRNAs and especially mRNAs is a rare occurrence. At least two explanations could account for the high diversity observed in population analyses: stochastic dynamics are involved in RNA targeting, and/or our preparations contained subpopulations of EVs with highly heterogeneous RNA compositions. Correlative analyses suggest that stochastic incorporation influences the long RNA repertoire to a larger extent than small RNA signatures, but our EV preparations displayed morphologic heterogeneity and likely contained microvesicles shed by the plasma membrane in addition to exosomes released by multivesicular endosome exocytosis [75]. A recent study reported immuno-isolation of EV obtained by ultracentrifugation to capture subpopulations selectively expressing diverse tetraspanins[76]. Quantitative proteomics revealed sizeable heterogeneity across EV subtypes, suggesting that transcriptomic content may also differ. Extending our comparative analyses of subcellular material to diverse subpopulations of EVs obtained by immuno-capture or density-gradient ultracentrifugation may reveal further insights into RNA transport to EV subtypes.

A 23 nt fragment mapped to Ro-associated Y RNAs was consistently predominant in our preparations, accounting for 85% of total reads in the small library, or 1.7 ± 1.2 molecules/EV. Abundant EV-associated small RNAs (22-35 nt) derived from *RNY5* have recently been reported to elicit an apoptotic response in primary cells treated with K562 EVs [77]. While the authors mention that the fragments they've uncovered likely arise from processing events taking place within EVs, our subcellular evidence rather argues that EV-enriched Y RNA fragments are also abundant throughout the cell and highly prevalent in the cytosolic fraction. Analogously to the canonical role of the signal recognition particle in protein secretion [78], Y RNA fragments may serve as platforms to promote EV recruitment and extracellular release of their associated proteins. Interestingly, expression of *RNY3* is required for extracellular targeting of the RNA-binding protein Ro60 in fibroblasts entering apoptosis[79]. In addition, other non-coding fragments identified in EVs may function as miRNA-like regulators. Indeed, small vault RNAs

can downregulate CYP3A4, a major drug-metabolizing enzyme [67], while 5' end fragments of tRNAs are involved in paternal inheritance of epigenetic metabolic cues[80]. In light of this evidence, it is tempting to speculate that the abundant fragments of infrastructural RNAs we uncovered in EVs may serve as instructive signals in RNA-mediated intercellular signaling, possibly promoting hematopoietic niche remodeling during leukemic progression.

Attempts to define EV-targeting motifs have relied on *ab initio* computational screening of RNA sequences [24]. While these efforts have identified several motifs enriched in EVs, the main conclusion is that no single sequence can account for the complex EV signatures revealed by transcriptomic profiling. Interestingly, EV-enriched RNAs were found to display significantly shorter half-lives than cell-retained RNAs, suggesting that broad biophysical properties may influence EV targeting. Similarly, a recent study using the MS2 coat protein to engineer a targeted and modular RNA loading system showed that active targeting is more efficient in the case of short (≤0.5 kb) RNA molecules than for longer (≥1.5 kb) sequences [81]. Our subcellular analyses indicate that the cytosolic RNA pool is also enriched in short RNAs, suggesting that the size bias effect associated to EV loading is also involved in the accumulation of RNAs in the cytosolic soluble space. We propose that cytosolic availability of RNAs, perhaps underlined by low molecular weight, exerts a bottleneck selection over the cellular populations ultimately encapsulated in EVs. Along with sequence determinants such as the GGGUUG miRNA motif, broad properties such as cytoplasmic accessibility and low molecular weight appear to shape the complex composition of EV RNAs (**Figure 6.7**).

Conclusion

Our data is consistent with a model of semi-selective targeting of cellular RNAs to EVs (**Figure 6.7**). We show that the protein and small RNA composition of K562 EVs best reflects the cytosolic fraction identity, suggesting that cytosolic accessibility promotes EV incorporation. Full-length mRNA and miRNA signatures were numerous in EV preparations but displayed low cumulative counts and strongly reflected every subcellular abundance distribution. Despite evidence of broad stochastic incorporation, diverse mRNAs and miRNAs were highly enriched in EVs, suggesting that sequence determinants modulate the likelihood of EV localization. We found that EV-enriched mRNAs were mostly distributed in the nucleus, consistent with rapid cytoplasmic shuttling for regulated EV export. In addition, multiple sequence alignment segregated EV-enriched and EV-depleted miRNAs into separate groups, suggesting that sequence identity influences EV incorporation, with the sequence GGGUUG significantly linked to EV localization. Our study demonstrates the asymmetry of EV molecular imprints and establishes a strong link between selective cytosolic localization and EV enrichments of short transcripts. Identification of the *cis*- and *trans*-acting targeting determinants that mediate specific subcellular and extracellular localization of RNA molecules will be pivotal to our understanding of the functions of extracellular RNA. Determining the underpinnings of RNA targeting to EVs may pave the way to rationally re-engineer nucleic acids secretion, with promising therapeutic outcomes.

Material and method

Cell culture and EVs purification

K562 cells were maintained in Dulbecco's Modified Eagle Medium (RPMI) containing 1% of a 1:100 solution of penicillin and streptomycin (Wisent) and 10% fetal bovine serum (depleted FBS) (Wisent). FBS was depleted of bovine EVs via ultracentrifugation (110,000×g, 18h, 4°C) before use. Cells were cultured at 37 °C (5% CO_2) in T-175 flasks. We routinely monitored cell death by trypan blue staining (Sigma-Aldrich), which remained below 5%.

Electron microscopy

Steps were performed as previously described. In brief, 8 µL of fresh EVs diluted in PBS were loaded on formvar-coated copper grids and allowed to adhere for 20 min. The grids were then incubated in 2% uranyl acetate for 30 s and extensively washed with dH_2O. Samples were imaged on a Tecnai 12 120 kV transmission electron microscope. Contrast was enhanced with the software Photoshop (Adobe).

Nanoparticle Tracking Analysis

Biological triplicates of fresh EV preparations were diluted in PBS (1:1000) and analyzed by nanoparticle tracking analysis using an LM-10 machine (Nanosight) according to the manufacturer's instructions. Samples were submitted to 3 successive analyses of 30 s using the default settings. Four washes were performed with dH_2O between analyses. PBS and RPMI medium containing depleted FBS was used as controls.

Fluorescence imaging of EV intercellular transfers

EV preparations suspended in PBS were labeled with PKH67 () according to the manufacturer's instructions using a dilution factor of 1:2500 determined empirically. To remove unbound dye, labeled EVs and a PBS solution containing an equivalent amount of PKH67 were purified using the Exo-Quick on-column kit () according to the manufacturer's instructions. Labeled EVs and PBS were diluted 1:10 in DMEM and co-cultured for 3h at 37 °C (5% CO_2) with previously

seeded HeLa cells in 8-well slide chambers (). HeLa cells were washed with PBS, fixed (formaldehyde 3.7%, 10 min), washed and permeabilized (Triton 0.1%, 15 min) and stained with DAPI () and phalloidin-534 () for 20 min. Slides were mounted with DABCO and imaged using on a Leica DM5500B epifluorescence microscope and a Zeiss LSM700 confocal microscope.

Cell fractionation and EV purification

Cell fractionation and EV purification were performed as previously described [20, 55]. For EV isolation, low passage (P<10) K562 cells were cultured for 48h starting from ~10^7 cells.

Protein isolation and LC-MS/MS

Fresh preparations of EVs, subcellular fractions and total cells were sonicated in RIPA buffer containing protease inhibitors. SDS-PAGE and silver staining were performed as previously described, using 1 μg of proteins per sample.

Isolation and characterization of exRNA

EVs and corresponding cellular pellets were resuspended in 1 mL of TRIzol™ reagent (Ambion) and processed according to the manufacturer's instructions RNA extracts were purified with the RNA Clean & Concentrator™-5 system (Zymo Research). *In-column* DNase I (New England BioLabs) treatment, RNA washes and elution steps were performed according to the manufacturer's instructions. One additional centrifugation step was included (5 min at 16,000×g) to ensure complete removal of the washing buffer. RNA samples were eluded in 12 μL of RNAse-free water (Wisent). Absorbance distributions were immediately quantified using a NanoDrop 2000c spectrophotometer. RNA samples were pure ($A_{260}/A_{280} \geq 2.0$)($2.00 \leq A_{260}/A_{230} \leq 2.25$). Aliquots of ~5 ng were submitted to capillary electrophoresis on a Bioanalyzer 2100 machine (Agilent). RNAse protection assay was performed as previously described[3,15] to confirm the intraluminal topology of exRNA within EVs. Cellular RNA from D17 and HepG2 extracts were used as controls. RNAse A (Qiagen) was inactivated by heat (10 min at 65 ºC) and RNA was extracted as described above.

Library generation for RNA deep sequencing deep sequencing

Biological duplicates of sequencing libraries were prepared from high quality RNA extracts (50 ng exRNA and 500 ng cellular RNA) using the Illumina TruSeq Stranded RNA Kit according to the manufacturer's instructions. The TruSeq PE Clusterkit v3-cBot-HS was used on an Illumina HiSEq 2000 machine.

***In silico* analysis of RNA sequencing data**

Read quality was confirmed by examination of FASTQ files and no trimming was deemed necessary. Read alignment was performed using Tophat v2.0.10 on the human GRCh37/hg19 and the *Drosophila* BDGP5.78/dm3 genomes, respectively. Raw alignment counts were calculated using htseq-count v0.5.3. Alignment BAM files were used to generate bigWig files, which were submitted to the UCSC genome browser for read coverage viewing. Using BEDTools, BAM files were then aligned to the RepeatMasker 4.0.6 collection of repeated sequences coordinates. Using *Bowtie2*, BAM files were aligned to gene coordinates annotated as rRNA in Ensembl and mapped reads were subtracted. Adjusted libraries were aligned to reference genomes with Tophat and resulting BAM files were used for expression analyses with Cufflinks v2.2.1. FPKM count tracking were used with a threshold of 5 FPKM to derive relative abundance values, expressed as Transcript per Million mapped ($TPM_i=[FPKM_i/\Sigma_j FPKM_j]\times 10^6$). For biotype and correlation analyses, the 1,000 most abundant transcripts were considered, which accounted for over 90% of identified transcripts.

Acknowledgments

We are grateful to Christopher Burge (MIT), Eric Wang (University of Florida), Brent Graveley, Michael Duff and Xintao Wei (University of Connecticut Health Center) for fruitful discussions regarding transcriptomic analyses of subcellular fractionation. We thank Janusz Rak and Delphine Garnier (McGill University), who helped with EV purification and characterization. We thank Nicole Francis (IRCM) for lending us the Sw-32-Ti rotor and Célia Jéronimo (IRCM) who helped with the silver staining of SDS-PAGE gels. F.A.L. receives a scholarship from the Canadian Institute of Health Research (CIHR) scholarship and E.L. is a Junior 2 scholar of the FRQS. This work was supported by grants to E.L. from CIHR (MOP137096), FRQS and the Cancer Research Society (grant #20227).

Figure and Table Legends

Figure 6.1. **Experimental workflow and structural characterization of K562 EVs**
(A) Experimental workflow of the CeFra-seq approach. Cellular lysates and culture supernatants are processed to generate 5 extracts: the extracellular vesicle (E), cytosolic (C), cytoplasmic insoluble (I), endomembrane (M) and nuclear (N) fractions. P=pellet S=supernatant. Total protein and RNA extracts are extracted from biological duplicates of each preparation and processed for proteomic and transcriptomic analyses (**B**) Representative electron micrographs of K562 EVs obtained through uranyl acetate counterstaining. Scale bars are shown on the lower left corners. **(C,D**) Size distributions of EVs as determined by direct annotation of transmission electron micrographs (**c**) and by nanoparticle tracking analysis (NTA, **d**). Mean size, standard deviation and sample size are indicated on the upper right corner. (**E,F**) Amounts of total RNA and protein recovered per EV preparation (**e**) and per single average EV (**f**), as inferred from NTA counts and spectrophotometric measurements of protein and RNA concentrations.

Figure 6.2. **The EV protein profile is related to the cytosolic fraction**
(A) SDS-PAGE silver staining of total cellular (CL) and EV proteins. **(B**) Venn diagram of proteins identified in EVs and cells as determined by LC-MS/MS. Examples of abundant EV protein groups are listed and flanked by circles proportional to their relative abundance (percentage of total spectral counts) in EVs (left) and cells (right). (**C**) Scatter plot displaying the relative abundance (%spectral counts – logarithmic scale) of proteins identified in cells and EVs. Pearson's (r) and Spearman's (ρ) correlation coefficients are indicated with associated p-values. (**D**) Venn diagram of proteins identified in EVs and subcellular fractions. (**E**) Pie charts showing the proportion of fraction-specific proteins containing a transmembrane domain (TM) and a canonical signal peptide (SP). The EV- and membrane-specific repertoires contain a significantly higher proportion of factors associated with a signal peptide than the other fractions ($P<10^{-4}$, Khi-squared test). (**F**) Correlation heatmap of protein levels (spectral counts) in subcellular and EV fractions. Pearson's (r) and Spearman's (ρ) correlation coefficient are indicated with associated P-values designated by asterisks. Color intensity is proportional to r. (**G**) Correlation

heatmap of protein cytotopic enrichment ratios (%spectral counts) in subcellular and EV fractions.

Figure 6.3. Transcriptomic analyses link the cytosolic and EV repertoires of small RNA

(**A**) Representative electropherograms showing the electrophoretic mobility distributions of total RNA isolated from total cellular, subcellular and EV fractions. (**B**) Representative read length distributions of the small RNA populations isolated from total cellular, subcellular and EV fractions. (**C,D**) Scatter plots depicting relative levels (expressed as transcripts per million, TPM) of small (**c**) and long (**d**) RNA populations isolated from EVs (*x*-axis) and total cells (*y*-axis). Pearson's (r) and Spearman's (ρ) correlations are indicated with associated p-values. (**E,F,G,H**) Correlation heatmaps of RNA levels (TPM) in subcellular and EV fractions for size-selected small (**e**) and long (**g**) populations. Correlation heatmaps of RNA cytotopic enrichment ratios (%spectral counts) in subcellular and EV fractions for size-selected small (**f**) and long (**h**) populations. Pearson's (r) and Spearman's (ρ) correlation coefficient are indicated with associated P-values designated by asterisks. Color intensity is proportional to r. (**I**) Pie charts of RNA biotype distributions for the 5,000 most abundant RNAs identified in small and long RNA populations of EV and subcellular fractions. rRNA reads were excluded from the analysis.

Figure 6.4. 3'-end fragments and full-length transcripts of Y RNAs and Vault RNAs are enriched in the EV and cytosolic fractions

(A,B) *UCSC* genome browser view of *RNY4* (**a**; chr7:148,660,407-148,660,502) and *VTRNA1-1* (**b**; chr5:140,090,861-140,090,958) showing read coverage (*y*-axis) in small and long RNA libraries of EVs, total cells and subcellular fractions. (**C,D**) Predicted secondary structure of full-length *RNY4* (**c**) and *VTRNA1-1* (**d**). The sequences corresponding to abundant 3'-end fragments mapped in the small RNA libraries are displayed in grey. (**E,F**) *MA* plots of the small (**e**) and standard (**f**) RNA-seq libraries in EVs and cells. The *y*-axis reflects EV enrichment while the *x*-axis indicates average abundance of each RNA in the two samples. Fragments of EV-enriched Y RNAs and related pseudogenes (**e**) and full-length vault RNAs (**f**) are labeled. (**G**) Histogram of the subcellular cytotopic distribution ratios associated to the ten Y RNA fragments most enriched in EVs. **(H**) Histogram of the subcellular cytotopic distribution ratios associated to the ten Y RNA fragments most enriched in EVs. (**I**) Simplex graph displaying the subcellular distributions

of all three full-length vault RNAs identified in EVs. (**I, J**) Simplex graph displaying the subcellular distributions of RNY4 fragments (**i,** red) and full-length vault RNAs (**j**, red) overlaid on all non-coding RNAs identified in cells (standard RNA library). The p-values indicating cytosolic enrichments were obtained through a Mann-Whitney rank-sum test.

Figure 6.5. EV-enriched mRNAs are predominantly localized in the nucleus

(A) Correlative analyses of mRNA levels (TPM) in the subcellular and EV fractions. Pearson's (r) and Spearman's (ρ) coefficients are indicated on top of graphs with corresponding p-values. **(B)** *MA* plot of mRNAs in EVs and cells. (**C**) Example of differentially-targeted mRNAs: *UCSC* genome browser screenshot of read coverage (*y*-axis) for a locus (chr12:120,093,121-120,326,077) encompassing the protein-coding genes *PRKAB1* and *CIT*. Read coverage in the long RNA library is shown for biological duplicates of EVs, total cells and subcellular fractions. **(D, E)** Simplex graph displaying the subcellular distribution of the 10 mRNAs most enriched (**d**) and most abundant (**e**) RNAs in EVs (red) overlaid on all mRNAs traced in cells (grey). The p-value indicating nuclear enrichment was obtained through a Mann-Whitney rank-sum test. (**F,G**) Histogram of the subcellular cytotopic distribution ratios associated to the ten transcripts most enriched (**f**) and most abundant (**g**) in EVs.

Figure 6.6. Sequence determinants of miRNA enrichment in EVs

(A) Correlative analyses of miRNA levels (TPM) in the subcellular and EV fractions. Pearson's (r) and Spearman's (ρ) coefficients are indicated on top of graphs with corresponding p-values. **(B)** Bar graph depicting the total (left *y*-axis) and cumulative (right *y*-axis) levels of the 10 most abundant miRNAs traced in EVs. Error bars correspond to standard deviation. (**C**) *MA* plot of miRNAs in K562 EVs and total cells. EV-enriched miRNAs are labeled. **(D)** Simplex graph displaying the subcellular distribution of the 10 mRNAs most enriched (**left**) and most abundant (**right**) miRNAs in EVs (red) overlaid on all mRNAs traced in cells (grey). The p-value indicating nuclear enrichment was obtained through a Mann-Whitney rank-sum test. (**E,F**) Histogram of the subcellular cytotopic distribution ratios associated to the ten miRNAs most enriched (**e**) and most abundant (**f**) in EVs. **(G)** Cladogram depicting clustering of multiple sequence alignments for the 40 most enriched (red) and less enriched (black) EV miRNA. **(H)**

Most enriched motifs in EV-targeted miRNAs as determined by computational analysis using XXmotif (left) and HOMER (right).

Figure 6.7. Model of RNA localization to EVs

We identified five properties that of RNAs enriched in EVs. Indeed, our data suggests that RNA localization in EVs is linked to cytosolic enrichment, short RNA length and specific sequence determinants, such as those present in Y RNAs. In addition, enriched EV mRNAs showed strong nuclear enrichments and the sequence GGGUUG was overrepresented among miRNAs enriched in EVs. This set of properties suggest that RNA localization to EVs is a complex process that is shaped by both sequence-based determinants and stochasticity.

Figure 6.S1. Western blot validation of K562 subcellular fractionation

Western blotting targeting Histone H3, α-tubulin, Ninein and the KDEL tetrapeptide was performed on equal amounts of protein extracted from total K562 cells (T) and nuclear (N), cytosolic (D), endomembrane (M) and cytoplasmic insoluble (I) fractions.

Figure 6.S2. RT-qPCR validation of K562 subcellular fractionation

Subcellular cytotopic ratios inferred from qRT-PCR for the lncRNA *XIST* and the mRNAs encoding histone H3, Anillin and an Adiponectin receptor.

Figure 6.S3. Gene ontology analysis of EV-enriched proteins

P-value and associated false discovery rate (FDR) for enriched gene ontology (GO) terms associated to protein factors identified in EVs, as inferred from LC-MS experiments. The dot colours reflect the number of proteins corresponding to each term.

Table 5.1. Most abundant proteins identified in EVs by LC-MS/MS

Table 6.2. Read metrics of EV, cellular and subcellular RNA sequencing libraries

Table 6.3. Most abundant genes in small and long EV RNA libraries

Table 6.4. Most abundant mRNAs and miRNAs in EVs

References

1. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM: **Localization of ASH1 mRNA particles in living yeast.** *Mol Cell* 1998, **2:**437-445.
2. Schmid M, Jaedicke A, Du TG, Jansen RP: **Coordination of endoplasmic reticulum and mRNA localization to the yeast bud.** *Curr Biol* 2006, **16:**1538-1543.
3. Latham VM, Yu EH, Tullio AN, Adelstein RS, Singer RH: **A Rho-dependent signaling pathway operating through myosin localizes beta-actin mRNA in fibroblasts.** *Curr Biol* 2001, **11:**1010-1016.
4. Lecuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, Hughes TR, Tomancak P, Krause HM: **Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function.** *Cell* 2007, **131:**174-187.
5. Yao J, Sasaki Y, Wen Z, Bassell GJ, Zheng JQ: **An essential role for beta-actin mRNA localization and translation in Ca2+-dependent growth cone guidance.** *Nat Neurosci* 2006, **9:**1265-1273.
6. Neuman-Silberberg FS, Schupbach T: **Dorsoventral axis formation in Drosophila depends on the correct dosage of the gene gurken.** *Development* 1994, **120:**2457-2463.
7. Ephrussi A, Dickinson LK, Lehmann R: **Oskar organizes the germ plasm and directs localization of the posterior determinant nanos.** *Cell* 1991, **66:**37-50.
8. Pedder CM, Ford D, Hesketh JE: **Targeting of transcripts encoding membrane proteins in polarized epithelia: RNA-protein binding studies of the SGLT1 3'-UTR.** *Biochem Soc Trans* 2008, **36:**525-527.
9. Cui XA, Palazzo AF: **Localization of mRNAs to the endoplasmic reticulum.** *Wiley Interdiscip Rev RNA* 2014, **5:**481-492.
10. Moor AE, Golan M, Massasa EE, Lemze D, Weizman T, Shenhav R, Baydatch S, Mizrahi O, Winkler R, Golani O, et al: **Global mRNA polarization regulates translation efficiency in the intestinal epithelium.** *Science* 2017, **357:**1299-1303.
11. Leung AK: **The Whereabouts of microRNA Actions: Cytoplasm and Beyond.** *Trends Cell Biol* 2015, **25:**601-610.

12. Wilusz JE, Sunwoo H, Spector DL: **Long noncoding RNAs: functional surprises from the RNA world.** *Genes Dev* 2009, **23:**1494-1504.
13. Dunagin M, Cabili MN, Rinn J, Raj A: **Visualization of lncRNA by single-molecule fluorescence in situ hybridization.** *Methods Mol Biol* 2015, **1262:**3-19.
14. Rastan S: **X chromosome inactivation and the Xist gene.** *Curr Opin Genet Dev* 1994, **4:**292-297.
15. van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E, Hao W, MacInnes AW, Cuppen E, Simonis M: **Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes.** *Genome Biol* 2014, **15:**R6.
16. Cody NA, Iampietro C, Lécuyer E: **The many functions of mRNA localization during normal development and disease: From pillar to post.** *WIREs Dev Biol* 2013.
17. Maguire CA, Balaj L, Sivaraman S, Crommentuijn MH, Ericsson M, Mincheva-Nilsson L, Baranov V, Gianni D, Tannous BA, Sena-Esteves M, et al: **Microvesicle-associated AAV vector as a novel gene delivery system.** *Mol Ther* 2012, **20:**960-971.
18. Lefebvre FA, Lecuyer E: **Small Luggage for a Long Journey: Transfer of Vesicle-Enclosed Small RNA in Interspecies Communication.** *Front Microbiol* 2017, **8:**377.
19. Tokuhisa M, Ichikawa Y, Kosaka N, Ochiya T, Yashiro M, Hirakawa K, Kosaka T, Makino H, Akiyama H, Kunisaki C, Endo I: **Exosomal miRNAs from Peritoneum Lavage Fluid as Potential Prognostic Biomarkers of Peritoneal Metastasis in Gastric Cancer.** *Plos One* 2015, **10**.
20. Lefebvre FA, Benoit Bouvrette LP, Perras L, Blanchet-Cohen A, Garnier D, Rak J, Lecuyer E: **Comparative transcriptomic analysis of human and Drosophila extracellular vesicles.** *Sci Rep* 2016, **6:**27680.
21. Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, Lotvall JO: **Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells.** *Nat Cell Biol* 2007, **9:**654-659.
22. Kalra H, Simpson RJ, Ji H, Aikawa E, Altevogt P, Askenase P, Bond VC, Borras FE, Breakefield X, Budnik V, et al: **Vesiclepedia: a compendium for extracellular vesicles with continuous community annotation.** *PLoS Biol* 2012, **10:**e1001450.

23. Bolukbasi MF, Mizrak A, Ozdener GB, Madlener S, Strobel T, Erkan EP, Fan JB, Breakefield XO, Saydam O: **miR-1289 and "Zipcode"-like Sequence Enrich mRNAs in Microvesicles.** *Molecular Therapy-Nucleic Acids* 2012, **1**.
24. Batagov AO, Kuznetsov VA, Kurochkin IV: **Identification of nucleotide patterns enriched in secreted RNAs as putative cis-acting elements targeting them to exosome nano-vesicles.** *BMC Genomics* 2011, **12 Suppl 3:**S18.
25. Gall JG: **The origin of in situ hybridization - A personal history.** *Methods* 2016, **98:**4-9.
26. Lecuyer E, Yoshida H, Krause HM: **Global implications of mRNA localization pathways in cellular organization.** *Curr Opin Cell Biol* 2009, **21:**409-415.
27. Bhatt DM, Pandya-Jones A, Tong AJ, Barozzi I, Lissner MM, Natoli G, Black DL, Smale ST: **Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions.** *Cell* 2012, **150:**279-290.
28. Tilgner H, Knowles DG, Johnson R, Davis CA, Chakrabortty S, Djebali S, Curado J, Snyder M, Gingeras TR, Guigo R: **Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs.** *Genome Res* 2012, **22:**1616-1625.
29. Bahar Halpern K, Caspi I, Lemze D, Levy M, Landen S, Elinav E, Ulitsky I, Itzkovitz S: **Nuclear Retention of mRNA in Mammalian Tissues.** *Cell Rep* 2015, **13:**2653-2662.
30. Lefebvre FA, Benoit Bouvrette LP, Bergalet J, Lecuyer E: **Biochemical Fractionation of Time-Resolved Drosophila Embryos Reveals Similar Transcriptomic Alterations in Replication Checkpoint and Histone mRNA Processing Mutants.** *J Mol Biol* 2017.
31. Carlevaro-Fita J, Rahim A, Guigo R, Vardy LA, Johnson R: **Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells.** *RNA* 2016, **22:**867-882.
32. Chen Q, Jagannathan S, Reid DW, Zheng T, Nicchitta CV: **Hierarchical regulation of mRNA partitioning between the cytoplasm and the endoplasmic reticulum of mammalian cells.** *Mol Biol Cell* 2011, **22:**2646-2658.
33. Wang ET, Cody NA, Jog S, Biancolella M, Wang TT, Treacy DJ, Luo S, Schroth GP, Housman DE, Reddy S, et al: **Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins.** *Cell* 2012, **150:**710-724.

34. Sterne-Weiler T, Martinez-Nunez RT, Howard JM, Cvitovik I, Katzman S, Tariq MA, Pourmand N, Sanford JR: **Frac-seq reveals isoform-specific recruitment to polyribosomes.** *Genome Res* 2013, **23:**1615-1623.
35. Martinez-Nunez RT, Sanford JR: **Studying Isoform-Specific mRNA Recruitment to Polyribosomes with Frac-seq.** *Methods Mol Biol* 2016, **1358:**99-108.
36. Kopczynski CC, Noordermeer JN, Serano TL, Chen WY, Pendleton JD, Lewis S, Goodman CS, Rubin GM: **A high throughput screen to identify secreted and transmembrane proteins involved in Drosophila embryogenesis.** *Proc Natl Acad Sci U S A* 1998, **95:**9973-9978.
37. Diehn M, Eisen MB, Botstein D, Brown PO: **Large-scale identification of secreted and membrane-associated gene products using DNA microarrays.** *Nat Genet* 2000, **25:**58-62.
38. Diehn M, Bhattacharya R, Botstein D, Brown PO: **Genome-scale identification of membrane-associated human mRNAs.** *PLoS Genet* 2006, **2:**e11.
39. de Jong M, van Breukelen B, Wittink FR, Menke FL, Weisbeek PJ, Van den Ackerveken G: **Membrane-associated transcripts in Arabidopsis; their isolation and characterization by DNA microarray analysis and bioinformatics.** *Plant J* 2006, **46:**708-721.
40. Lerner RS, Seiser RM, Zheng T, Lager PJ, Reedy MC, Keene JD, Nicchitta CV: **Partitioning and translation of mRNAs encoding soluble proteins on membrane-bound ribosomes.** *Rna* 2003, **9:**1123-1137.
41. Marc P, Margeot A, Devaux F, Blugeon C, Corral-Debrinski M, Jacq C: **Genome-wide analysis of mRNAs targeted to yeast mitochondria.** *EMBO Rep* 2002, **3:**159-164.
42. Mercer TR, Neph S, Dinger ME, Crawford J, Smith MA, Shearwood AM, Haugen E, Bracken CP, Rackham O, Stamatoyannopoulos JA, et al: **The human mitochondrial transcriptome.** *Cell* 2011, **146:**645-658.
43. Mili S, Moissoglu K, Macara IG: **Genome-wide screen reveals APC-associated RNAs enriched in cell protrusions.** *Nature* 2008, **453:**115-119.
44. Blower MD, Feric E, Weis K, Heald R: **Genome-wide analysis demonstrates conserved localization of messenger RNAs to mitotic microtubules.** *J Cell Biol* 2007, **179:**1365-1373.

45. Sharp JA, Plant JJ, Ohsumi TK, Borowsky M, Blower MD: **Functional analysis of the microtubule-interacting transcriptome.** *Mol Biol Cell* 2011, **22:**4312-4323.
46. Eberwine J, Miyashiro K, Kacharmina JE, Job C: **Local translation of classes of mRNAs that are targeted to neuronal dendrites.** *Proc Natl Acad Sci U S A* 2001, **98:**7080-7085.
47. Job C, Eberwine J: **Localization and translation of mRNA in dendrites and axons.** *Nat Rev Neurosci* 2001, **2:**889-898.
48. Moccia R, Chen D, Lyles V, Kapuya E, E Y, Kalachikov S, Spahn CM, Frank J, Kandel ER, Barad M, Martin KC: **An unbiased cDNA library prepared from isolated Aplysia sensory neuron processes is enriched for cytoskeletal and translational mRNAs.** *J Neurosci* 2003, **23:**9409-9417.
49. Poon MM, Choi SH, Jamieson CA, Geschwind DH, Martin KC: **Identification of process-localized mRNAs from cultured rodent hippocampal neurons.** *J Neurosci* 2006, **26:**13390-13399.
50. Zivraj KH, Tung YC, Piper M, Gumy L, Fawcett JW, Yeo GS, Holt CE: **Subcellular profiling reveals distinct and developmentally regulated repertoire of growth cone mRNAs.** *J Neurosci* 2010, **30:**15464-15478.
51. Gumy LF, Yeo GS, Tung YC, Zivraj KH, Willis D, Coppola G, Lam BY, Twiss JL, Holt CE, Fawcett JW: **Transcriptome analysis of embryonic and adult sensory axons reveals changes in mRNA repertoire localization.** *RNA* 2011, **17:**85-98.
52. Khaladkar M, Buckley PT, Lee MT, Francis C, Eghbal MM, Chuong T, Suresh S, Kuhn B, Eberwine J, Kim J: **Subcellular RNA sequencing reveals broad presence of cytoplasmic intron-sequence retaining transcripts in mouse and rat neurons.** *PLoS One* 2013, **8:**e76194.
53. Sasaki Y, Gross C, Xing L, Goshima Y, Bassell GJ: **Identification of axon-enriched microRNAs localized to growth cones of cortical neurons.** *Dev Neurobiol* 2014, **74:**397-406.
54. Taliaferro JM, Vidaki M, Oliveira R, Olson S, Zhan L, Saxena T, Wang ET, Graveley BR, Gertler FB, Swanson MS, Burge CB: **Distal Alternative Last Exons Localize mRNAs to Neural Projections.** *Mol Cell* 2016, **61:**821-833.

55. Lefebvre FA, Cody N, Bouvrette LPB, Bergalet J, Wang X, Lecuyer E: **CeFra-seq: Systematic mapping of RNA subcellular distribution properties through cell fractionation coupled to deep-sequencing.** *Methods* 2017.
56. Dignam JD, Lebovitz RM, Roeder RG: **Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei.** *Nucleic Acids Res* 1983, **11:**1475-1489.
57. Thery C, Amigorena S, Raposo G, Clayton A: **Isolation and characterization of exosomes from cell culture supernatants and biological fluids.** *Curr Protoc Cell Biol* 2006, **Chapter 3:**Unit 3 22.
58. Vedeler A, Pryme IF, Hesketh JE: **The characterization of free, cytoskeletal and membrane-bound polysomes in Krebs II ascites and 3T3 cells.** *Mol Cell Biochem* 1991, **100:**183-193.
59. Jagannathan S, Nwosu C, Nicchitta CV: **Analyzing mRNA localization to the endoplasmic reticulum via cell fractionation.** *Methods Mol Biol* 2011, **714:**301-321.
60. Chomczynski P, Sacchi N: **Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction.** *Anal Biochem* 1987, **162:**156-159.
61. Simoes AE, Pereira DM, Amaral JD, Nunes AF, Gomes SE, Rodrigues PM, Lo AC, D'Hooge R, Steer CJ, Thibodeau SN, et al: **Efficient recovery of proteins from multiple source samples after TRIzol((R)) or TRIzol((R))LS RNA extraction and long-term storage.** *BMC Genomics* 2013, **14:**181.
62. Brocchieri L, Karlin S: **Protein length in eukaryotic and prokaryotic proteomes.** *Nucleic Acids Res* 2005, **33:**3390-3400.
63. Lotvall J, Hill AF, Hochberg F, Buzas EI, Di Vizio D, Gardiner C, Gho YS, Kurochkin IV, Mathivanan S, Quesenberry P, et al: **Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles.** *J Extracell Vesicles* 2014, **3:**26913.
64. Fujita A, Sato JR, Demasi MA, Sogayar MC, Ferreira CE, Miyano S: **Comparing Pearson, Spearman and Hoeffding's D measure for gene expression association analysis.** *J Bioinform Comput Biol* 2009, **7:**663-684.

65. Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, Lee BT, Learned K, Karolchik D, Hinrichs AS, et al: **The UCSC Genome Browser database: 2016 update.** *Nucleic Acids Res* 2016, **44:**D717-725.
66. Kickhoefer VA, Emre N, Stephen AG, Poderycki MJ, Rome LH: **Identification of conserved vault RNA expression elements and a non-expressed mouse vault RNA gene.** *Gene* 2003, **309:**65-70.
67. Persson H, Kvist A, Vallon-Christersson J, Medstrand P, Borg A, Rovira C: **The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs.** *Nat Cell Biol* 2009, **11:**1268-1271.
68. Luehr S, Hartmann H, Soding J: **The XXmotif web server for eXhaustive, weight matriX-based motif discovery in nucleotide sequences.** *Nucleic Acids Res* 2012, **40:**W104-109.
69. Brenner C: **HOMER: Software for motif discovery and next-gen sequencing analysis.**; 2012. .
70. Orhant L, Rondeau S, Vasson A, Anselem O, Goffinet F, Allach El Khattabi L, Leturcq F, Vidaud D, Bienvenu T, Tsatsaris V, Nectoux J: **Droplet digital PCR, a new approach to analyze fetal DNA from maternal blood: application to the determination of fetal RHD genotype.** *Ann Biol Clin (Paris)* 2016, **74:**269-277.
71. Zhang M, Pan Y, Qi X, Liu Y, Dong R, Zheng D, Chang Q, Zhang J, Fang W, Zhong Y: **Identification of New Biomarkers Associated With IDH Mutation and Prognosis in Astrocytic Tumors Using NanoString nCounter Analysis System.** *Appl Immunohistochem Mol Morphol* 2016.
72. Chevillet JR, Kang Q, Ruf IK, Briggs HA, Vojtech LN, Hughes SM, Cheng HH, Arroyo JD, Meredith EK, Gallichotte EN, et al: **Quantitative and stoichiometric analysis of the microRNA content of exosomes.** *Proc Natl Acad Sci U S A* 2014, **111:**14888-14893.
73. Cha DJ, Franklin JL, Dou Y, Liu Q, Higginbotham JN, Demory Beckler M, Weaver AM, Vickers K, Prasad N, Levy S, et al: **KRAS-dependent sorting of miRNA to exosomes.** *Elife* 2015, **4:**e07197.
74. Villarroya-Beltri C, Gutierrez-Vazquez C, Sanchez-Cabo F, Perez-Hernandez D, Vazquez J, Martin-Cofreces N, Martinez-Herrera DJ, Pascual-Montano A, Mittelbrunn

M, Sanchez-Madrid F: **Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs.** *Nature Communications* 2013, **4**.

75. Bobrie A, Colombo M, Krumeich S, Raposo G, Thery C: **Diverse subpopulations of vesicles secreted by different intracellular mechanisms are present in exosome preparations obtained by differential ultracentrifugation.** *J Extracell Vesicles* 2012, **1**.
76. Kowal J, Arras G, Colombo M, Jouve M, Morath JP, Primdal-Bengtson B, Dingli F, Loew D, Tkach M, Thery C: **Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes.** *Proc Natl Acad Sci U S A* 2016, **113:**E968-977.
77. Chakrabortty SK, Prakash A, Nechooshtan G, Hearn S, Gingeras TR: **Extracellular vesicle-mediated transfer of processed and functional RNY5 RNA.** *RNA* 2015, **21:**1966-1979.
78. Walter P, Blobel G: **Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum.** *Nature* 1982, **299:**691-698.
79. Reed JH, Sim S, Wolin SL, Clancy RM, Buyon JP: **Ro60 requires Y3 RNA for cell surface exposure and inflammation associated with cardiac manifestations of neonatal lupus.** *J Immunol* 2013, **191:**110-116.
80. Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, Belleannee C, Kucukural A, Serra RW, Sun F, et al: **Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals.** *Science* 2016, **351:**391-396.
81. Hung ME, Leonard JN: **A platform for actively loading cargo RNA to elucidate limiting steps in EV-mediated delivery.** *J Extracell Vesicles* 2016, **5:**31027.

Tables

Table 6. I. Most abundant proteins identified in EVs by LC-MS/MS

Gene	Protein Name	% of total
ACTB	Actin, cytoplasmic 1	5.75
TUBB	Tubulin beta chain	3.57
ACTC1	Actin, alpha cardiac muscle 1	2.98
TUBB4B	Tubulin beta-4B chain	2.94
TUBA4A	Tubulin alpha-4A chain	2.81
TUBA1B	Tubulin alpha-1B chain	2.78
AHCY	Adenosylhomocysteinase	2.53
KRT2	Keratin, type II cytoskeletal 2 epidermal	2.52
PPIA	Peptidyl-prolyl cis-trans isomerase A	2.37
KRT16	Keratin, type I cytoskeletal 16	2.23
RAP1B	Ras-related protein Rap-1b	1.91
PFN1	Profilin-1	1.69
KRT14	Keratin, type I cytoskeletal 14	1.61
RAN	GTP-binding nuclear protein Ran	1.49
KRT6B	Keratin, type II cytoskeletal 6B	1.47
KRT9	Keratin, type I cytoskeletal 9	1.47
MFGE8	Lactadherin	1.42
MT1E	Metallothionein-1E	1.26
HSPA8	Heat shock cognate 71 kDa protein	1.26
HSP90AB1	Heat shock protein HSP 90-beta	1.14

Table 6. II. Read metrics of EV, cellular and subcellular RNA sequencing libraries

	Small RNA population		Long RNA population	
Sample	Reads	Duplicates	Reads	Duplicates
EVs-1	15,528,785	95%	16,740,794	85%
EVs-2	18,394,535	97%	21,303,078	80%
Total-1	14,719,431	98%	19,777,448	21%
Total-2	15,382,715	98%	19,318,431	21%
Cytosol-1	15,124,350	88%	14,703,644	57%
Cytosol-2	14,689,253	87%	15,108,349	57%
Insoluble-1	14,735,241	91%	13,677,789	13%
Insoluble-2	15,234,032	89%	14,136,135	13%
Membrane-1	15,821,464	75%	14,747,973	38%
Membrane-2	15,414,685	78%	15,125,845	38%
Nuclear-1	14,514,307	92%	12,526,133	10%
Nuclear-2	16,777,561	90%	12,230,657	11%

Table 6. III. Most abundant genes in small and long EV RNA libraries

Small RNA population			Long RNA population		
Gene name	Gene biotype	% total reads	Gene name	Gene biotype	% total reads
RNY4	misc_RNA	14.56	RN7SL1	antisense	5.23
Y_RNA	misc_RNA	13.39	RN7SL2	misc_RNA	4.89
Y_RNA	misc_RNA	13.37	RPPH1	antisense	3.14
Y_RNA	misc_RNA	13.18	RN7SK	misc_RNA	1.59
RNY4P25	misc_RNA	4.53	RMRP	lincRNA	1.18
RNY4P18	misc_RNA	4.51	MALAT1	lincRNA	1.10
Y_RNA	misc_RNA	4.39	RNY1	misc_RNA	0.80
MIR17HG	processed_transcript	1.65	RNU4-2	snRNA	0.71
MIR92A2	miRNA	1.51	EEF1A1	protein_coding	0.68
RNY4P10	misc_RNA	1.10	Y_RNA	misc_RNA	0.46
CCNL1	protein_coding	1.10	EEF2	protein_coding	0.41
Y_RNA	misc_RNA	1.10	RNU5A-1	snRNA	0.39
RNY4P7	misc_RNA	1.07	SNORA73B	snoRNA	0.35
HOXB4	protein_coding	0.55	SNHG3	sense_intronic	0.32
MIR10A	sense_intronic	0.55	SNORD3A	lincRNA	0.31
MIR191	miRNA	0.41	RNU4-1	snRNA	0.30
MIR146B	miRNA	0.36	RPS18	protein_coding	0.28
RNY3	misc_RNA	0.29	FTL	protein_coding	0.27
Y_RNA	misc_RNA	0.23	RNY4	misc_RNA	0.23
Y_RNA	misc_RNA	0.23	RPLP0	protein_coding	0.22

Table 6. IV. Most abundant mRNAs and miRNAs in EVs

Long RNA library		Small RNA library	
mRNA name	% total reads	miRNA name	% total reads
EEF1A1	0.68	hsa-miR-486-5p	30.74
EEF2	0.41	hsa-miR-92a-3p	29.58
RPS18	0.28	hsa-miR-10a-5p	7.43
FTL	0.27	hsa-miR-191-5p	5.03
RPLP0	0.22	hsa-miR-146b-5p	4.40
PABPC1	0.21	hsa-miR-378a-3p	2.10
ACTB	0.21	hsa-miR-182-5p	1.92
RPL4	0.21	hsa-miR-21-5p	1.77
RPL13A	0.20	hsa-miR-148a-3p	1.38
RPL10	0.20	hsa-miR-25-3p	1.34
GAPDH	0.19	hsa-miR-92b-3p	0.76
HSP90AB1	0.19	hsa-miR-30d-5p	0.73
FTH1	0.19	hsa-miR-22-3p	0.70
RPS8	0.18	hsa-miR-26a-5p	0.65
TPT1	0.18	hsa-miR-151a-3p	0.50
RPL3	0.18	hsa-miR-186-5p	0.50
GNB2L1	0.18	hsa-miR-320a	0.43
CIT	0.17	hsa-miR-16-5p	0.41
MT-ND4	0.17	hsa-miR-99b-5p	0.37
RPS2	0.17	hsa-miR-142-5p	0.32

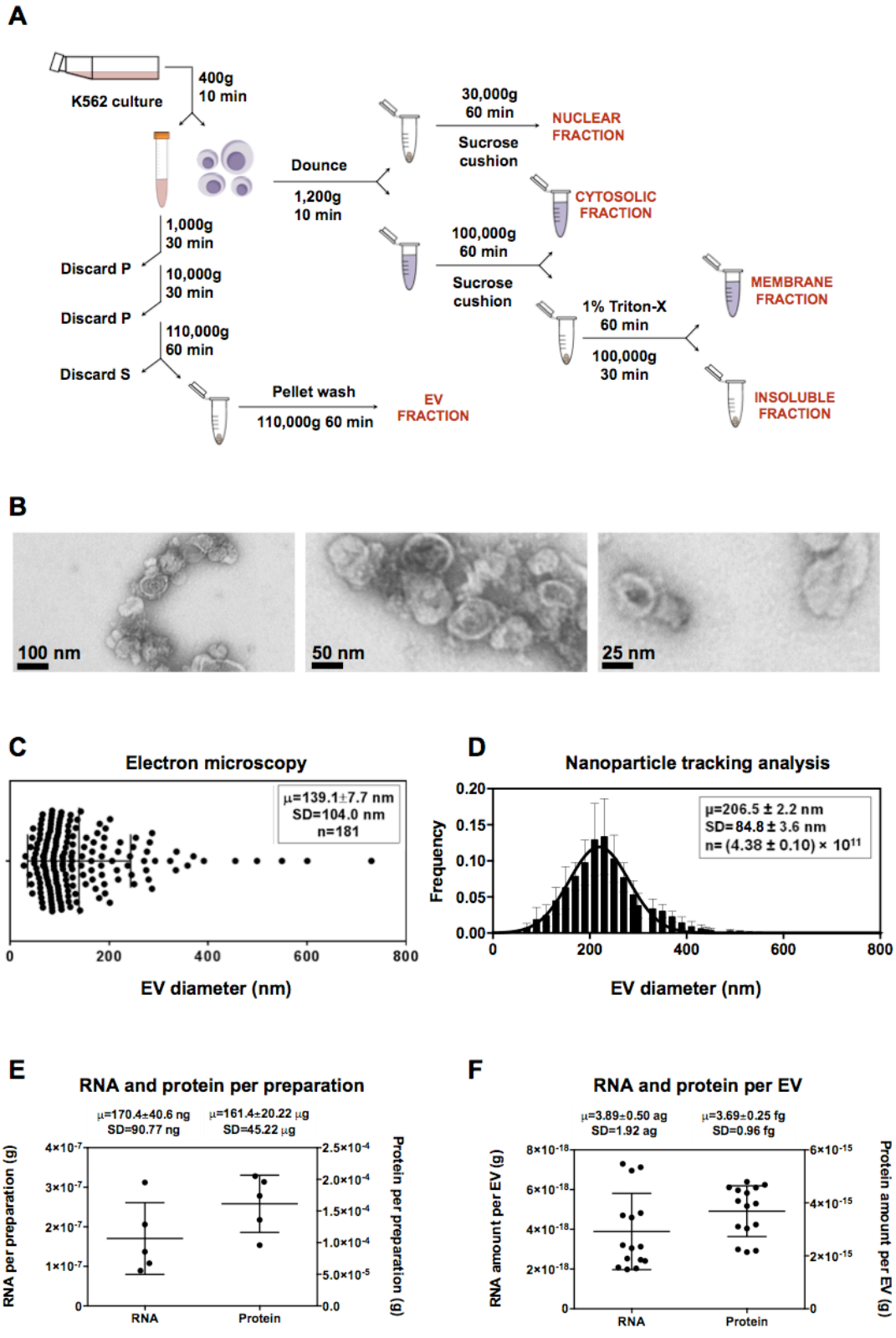


Figure 6.1. Experimental workflow and structural characterization of K562 EVs

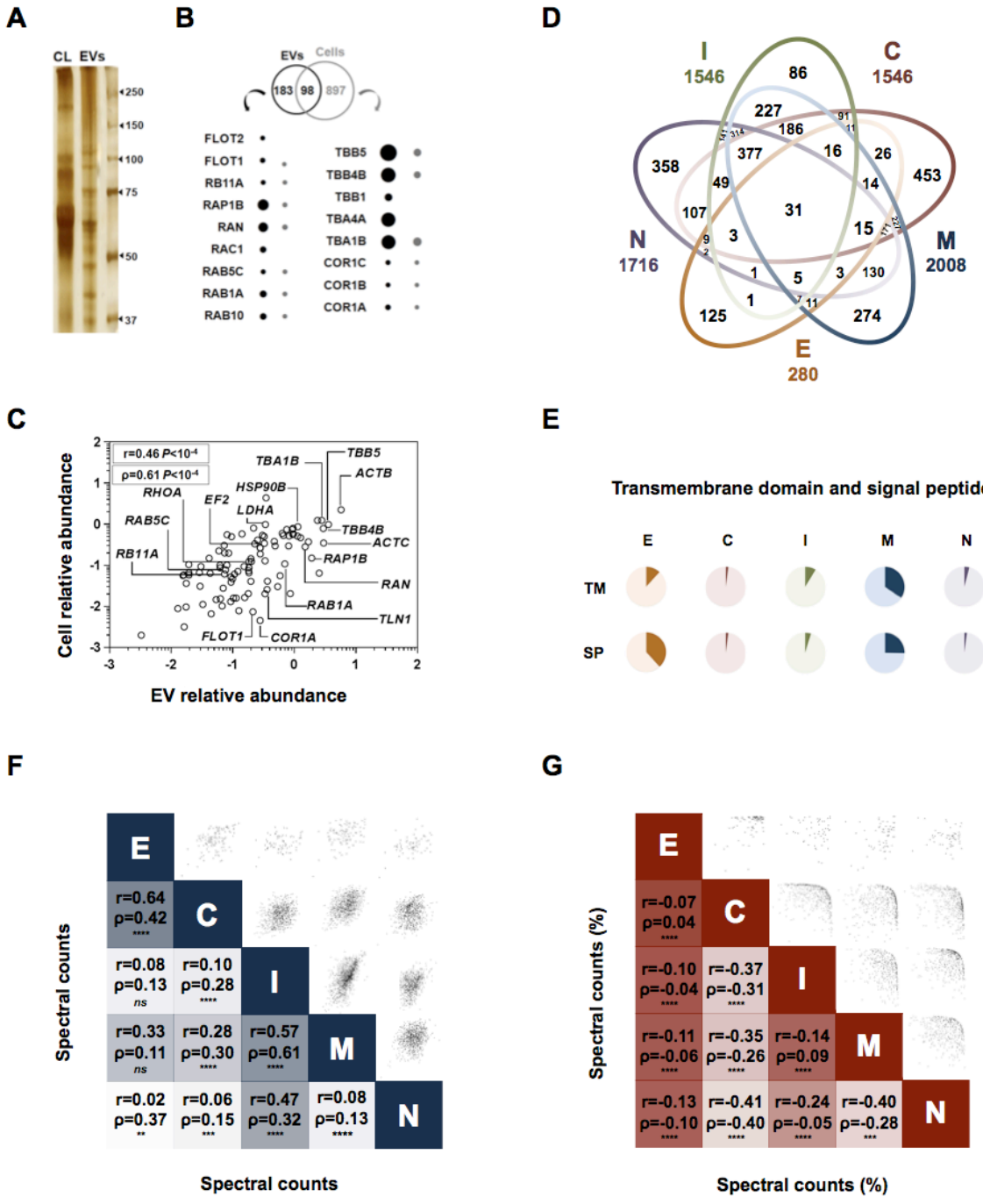


Figure 6.2. The EV protein profile is related to the cytosolic fraction

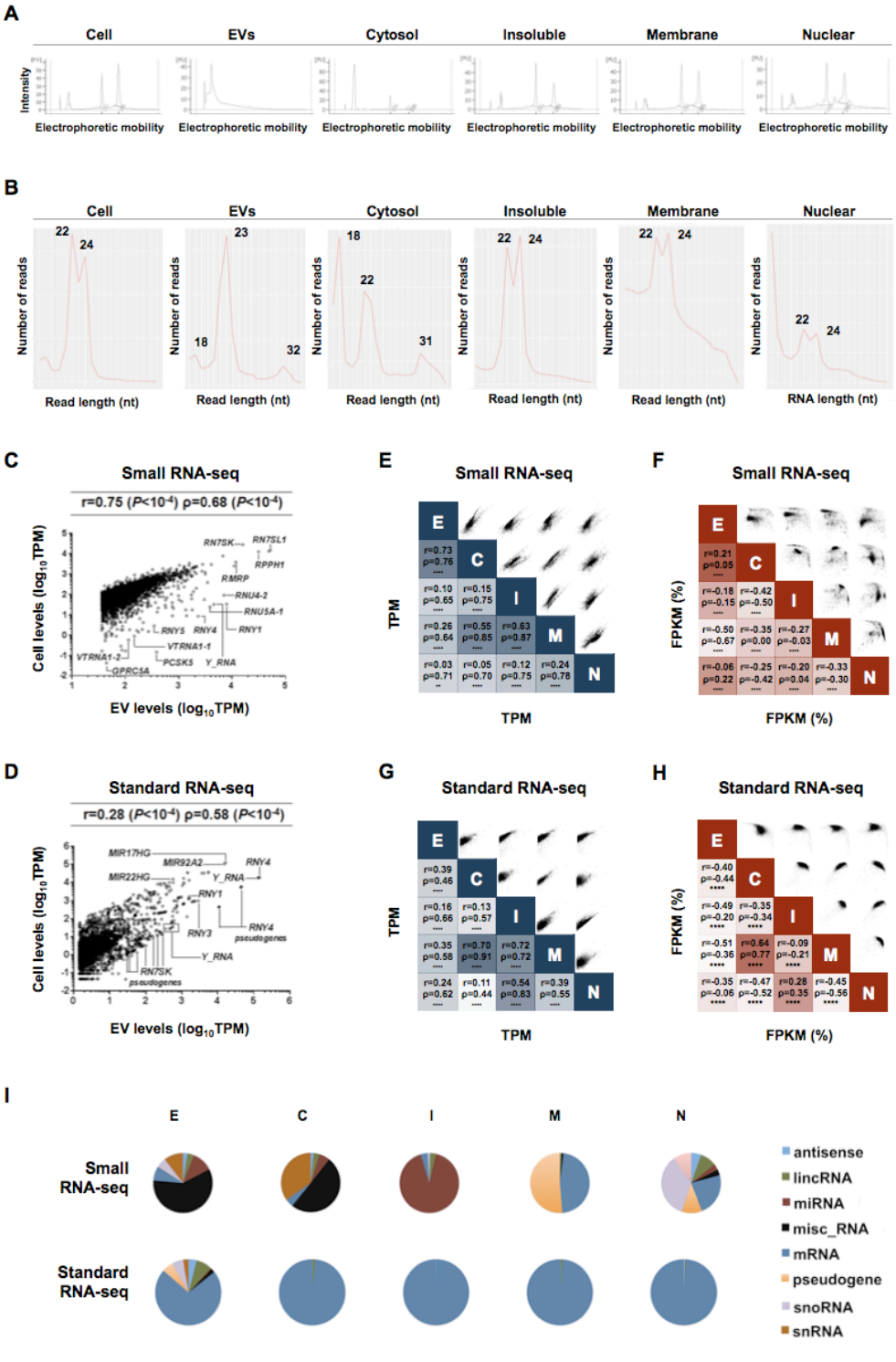


Figure 6.3. Transcriptomic analyses link the cytosolic and EV repertoires of small RNA

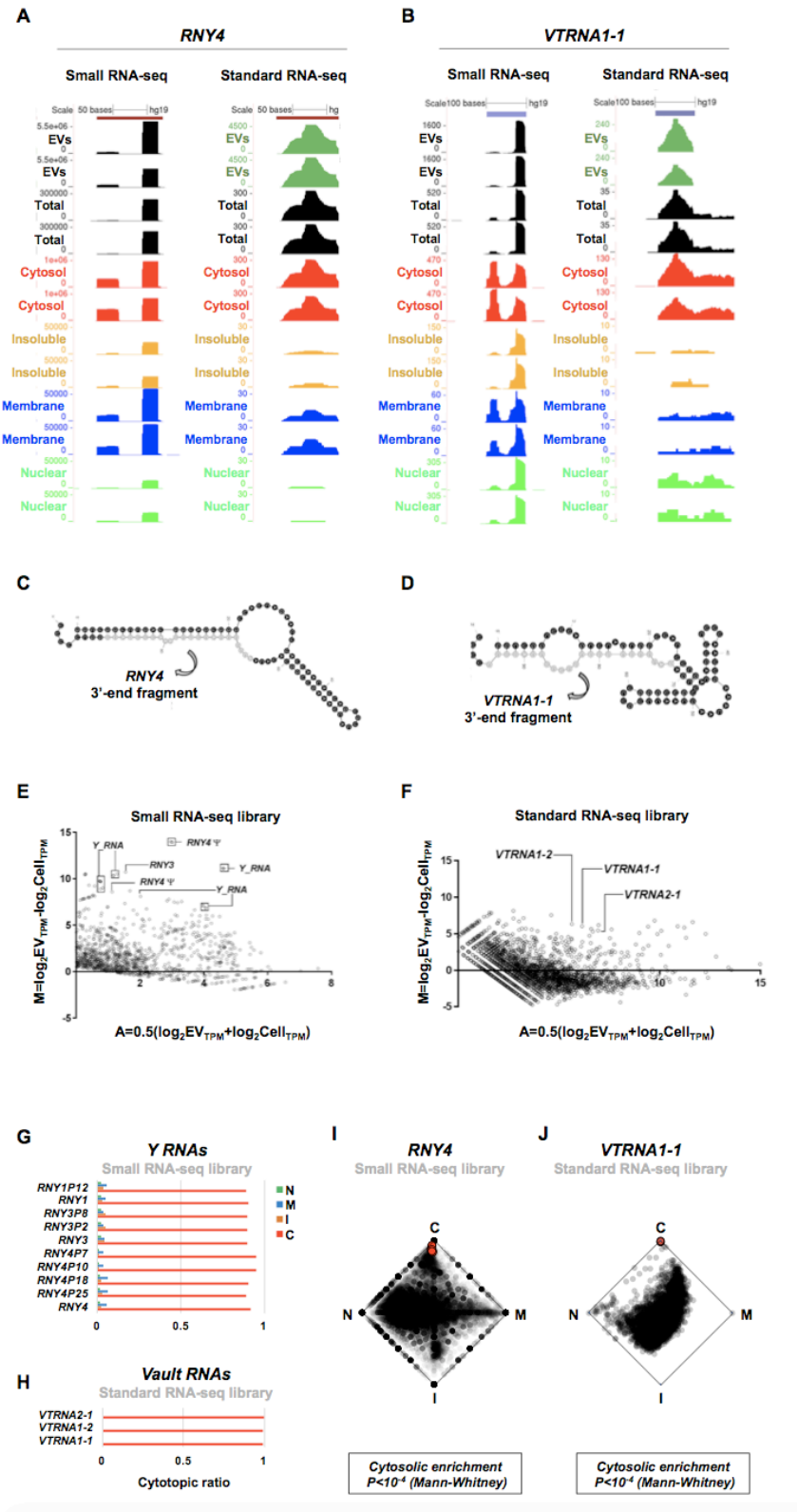


Figure 6.4. 3'-end fragments and full-length transcripts of Y RNAs and Vault RNAs are enriched in the EV and cytosolic fractions

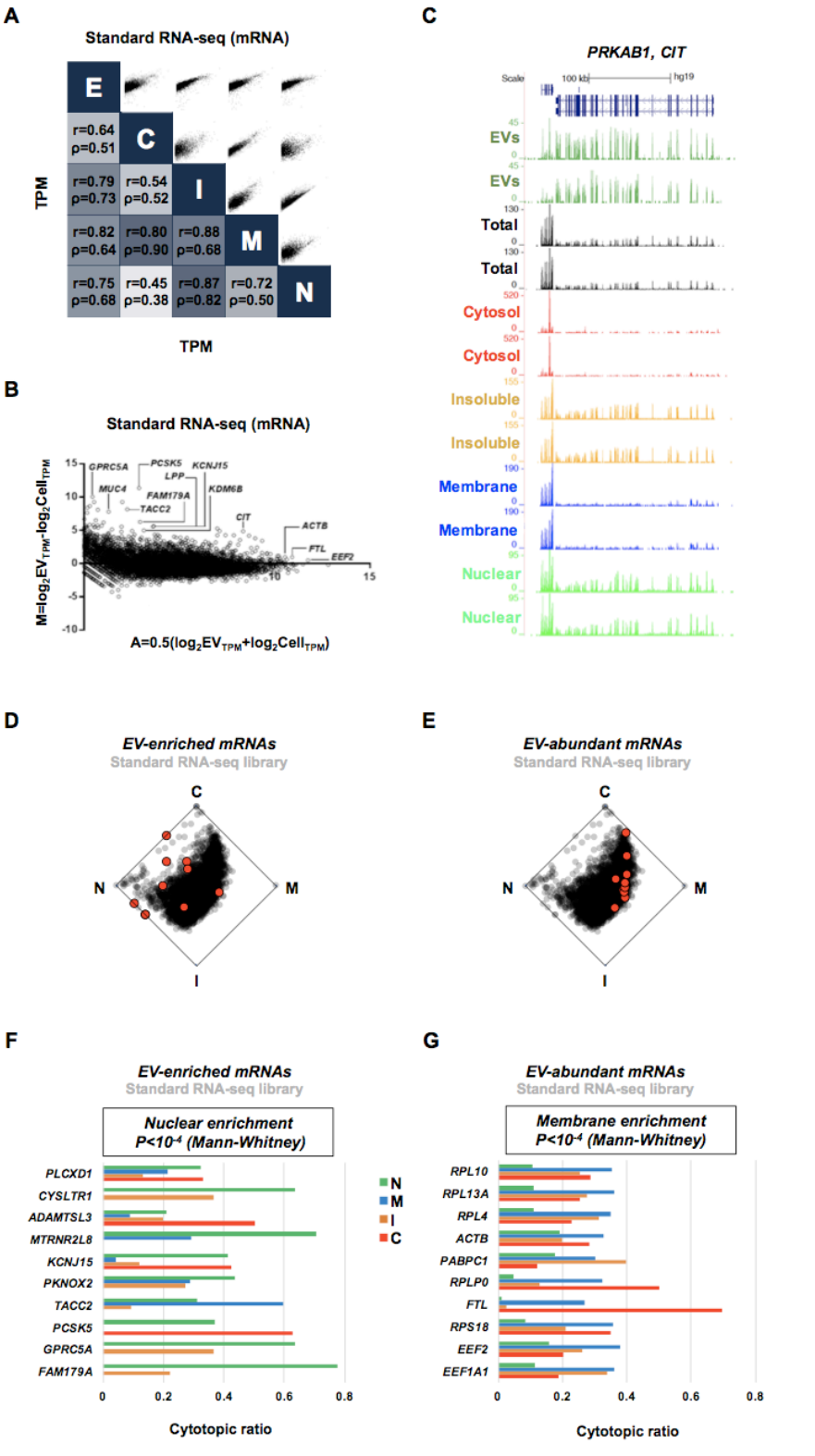


Figure 6.5. EV-enriched mRNAs are predominantly localized in the nucleus

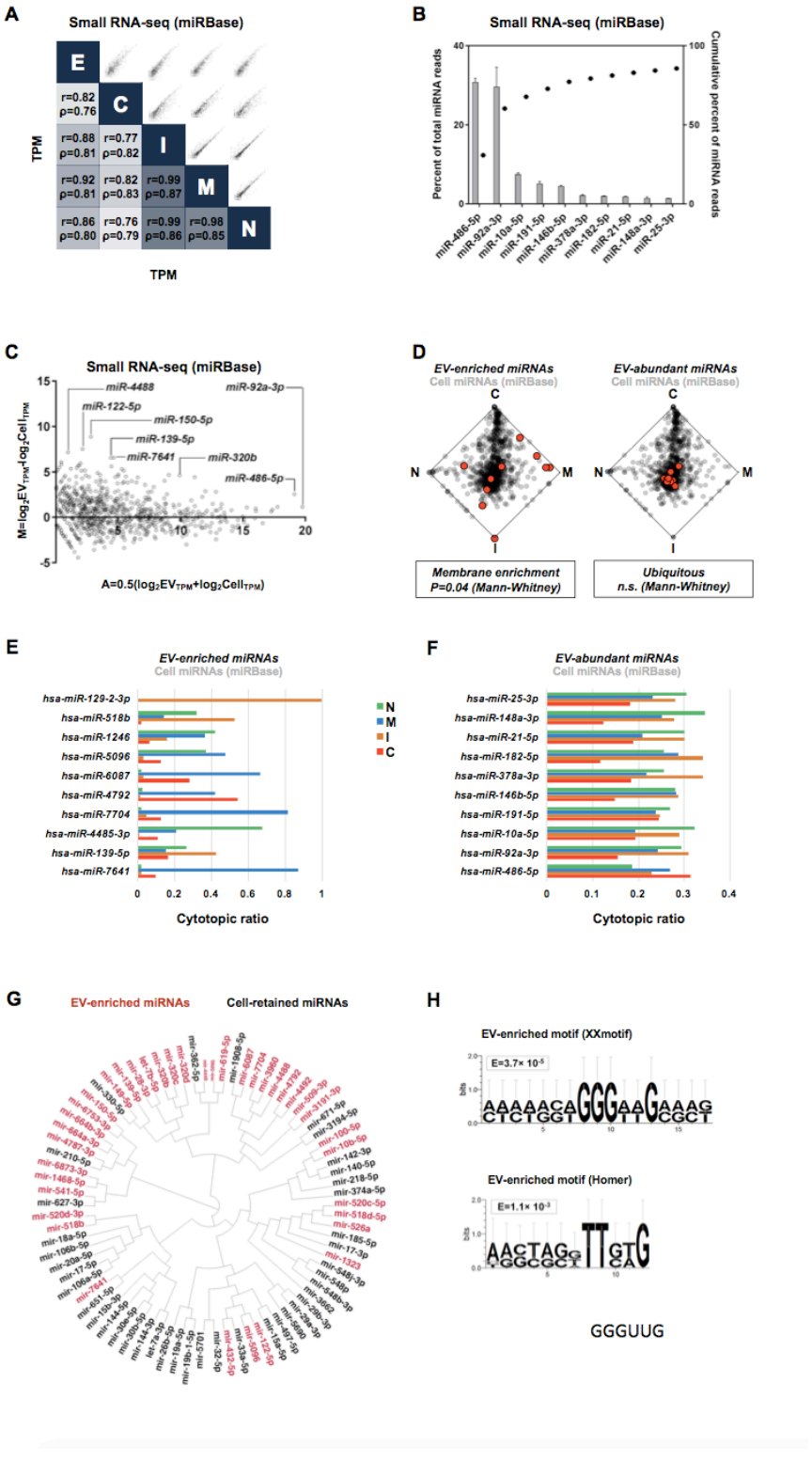


Figure 6.6. Sequence determinants of miRNA enrichment in EVs

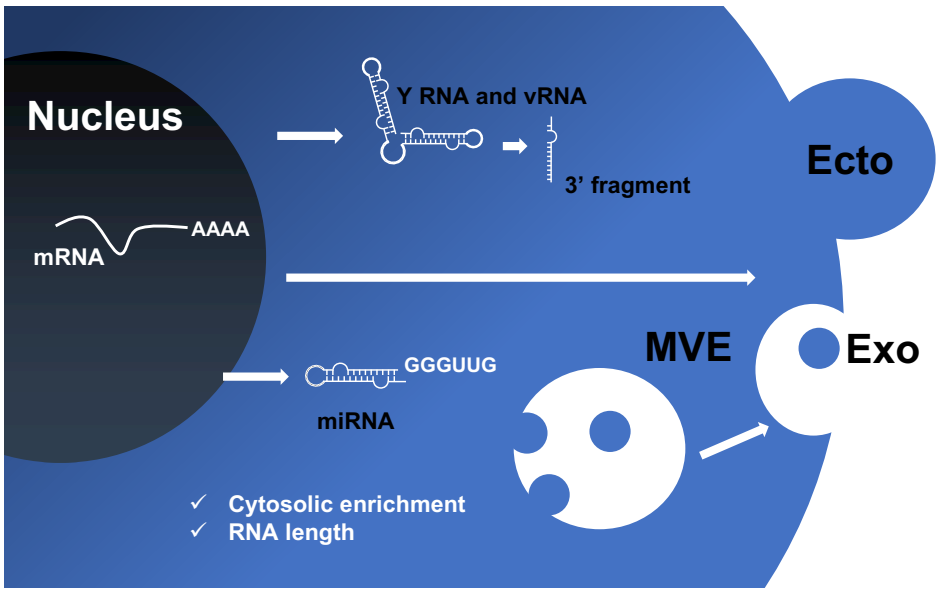


Figure 6.7. Model of RNA localization to EVs

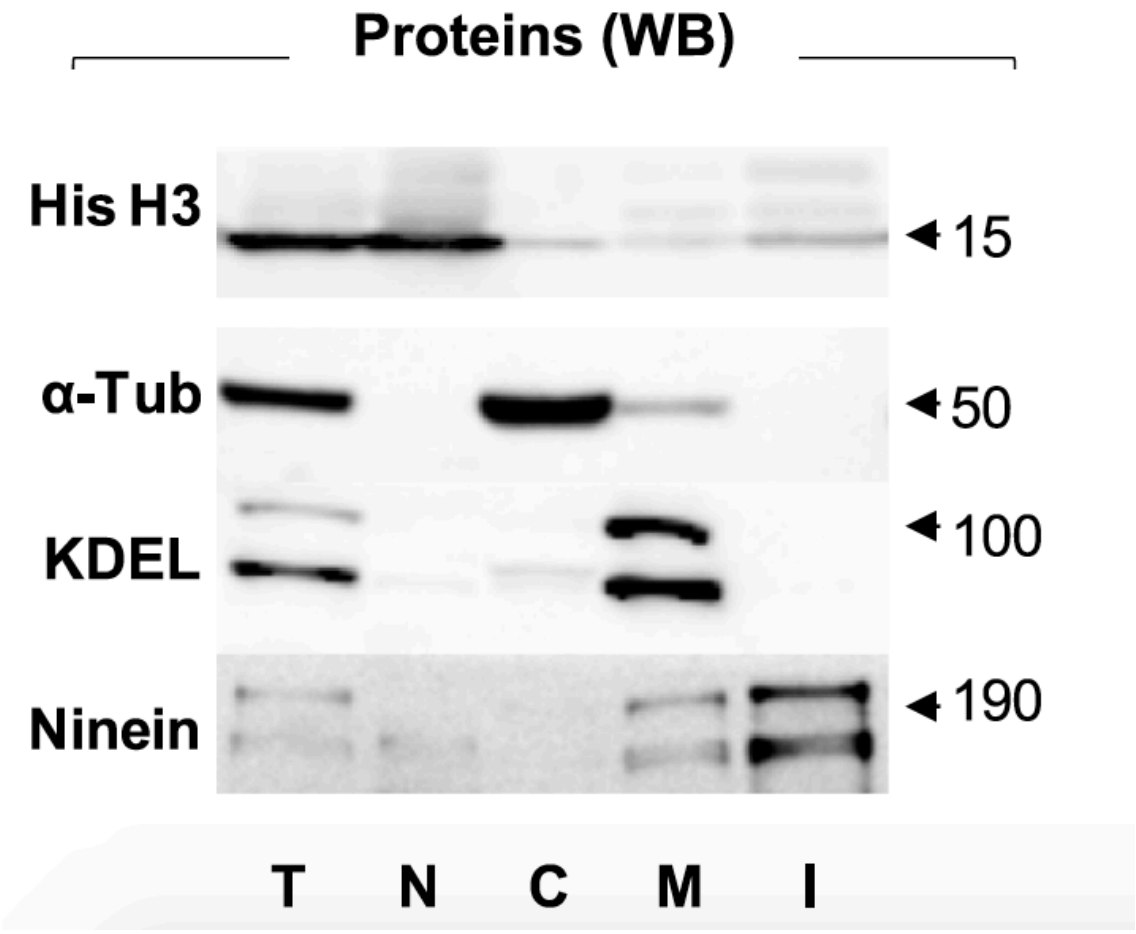


Figure 6S.1. **Western blot validation of K562 subcellular fractionation**

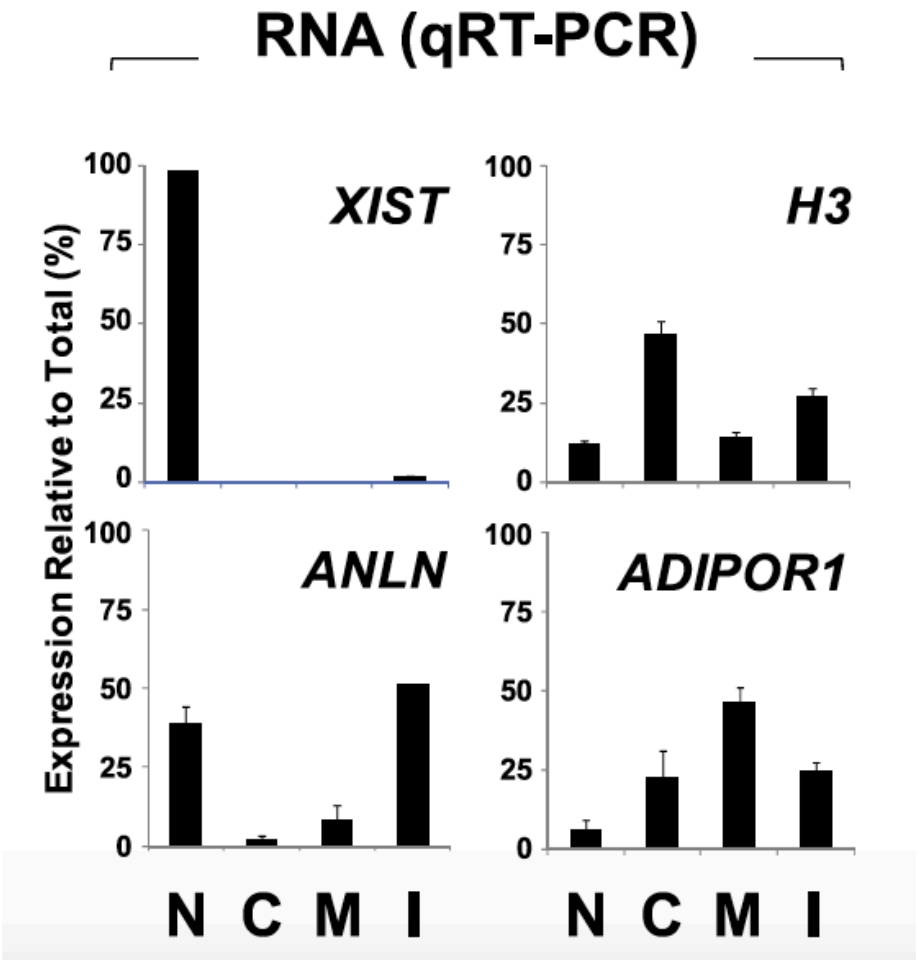


Figure 6S.2. **RT-qPCR validation of K562 subcellular fractionation**

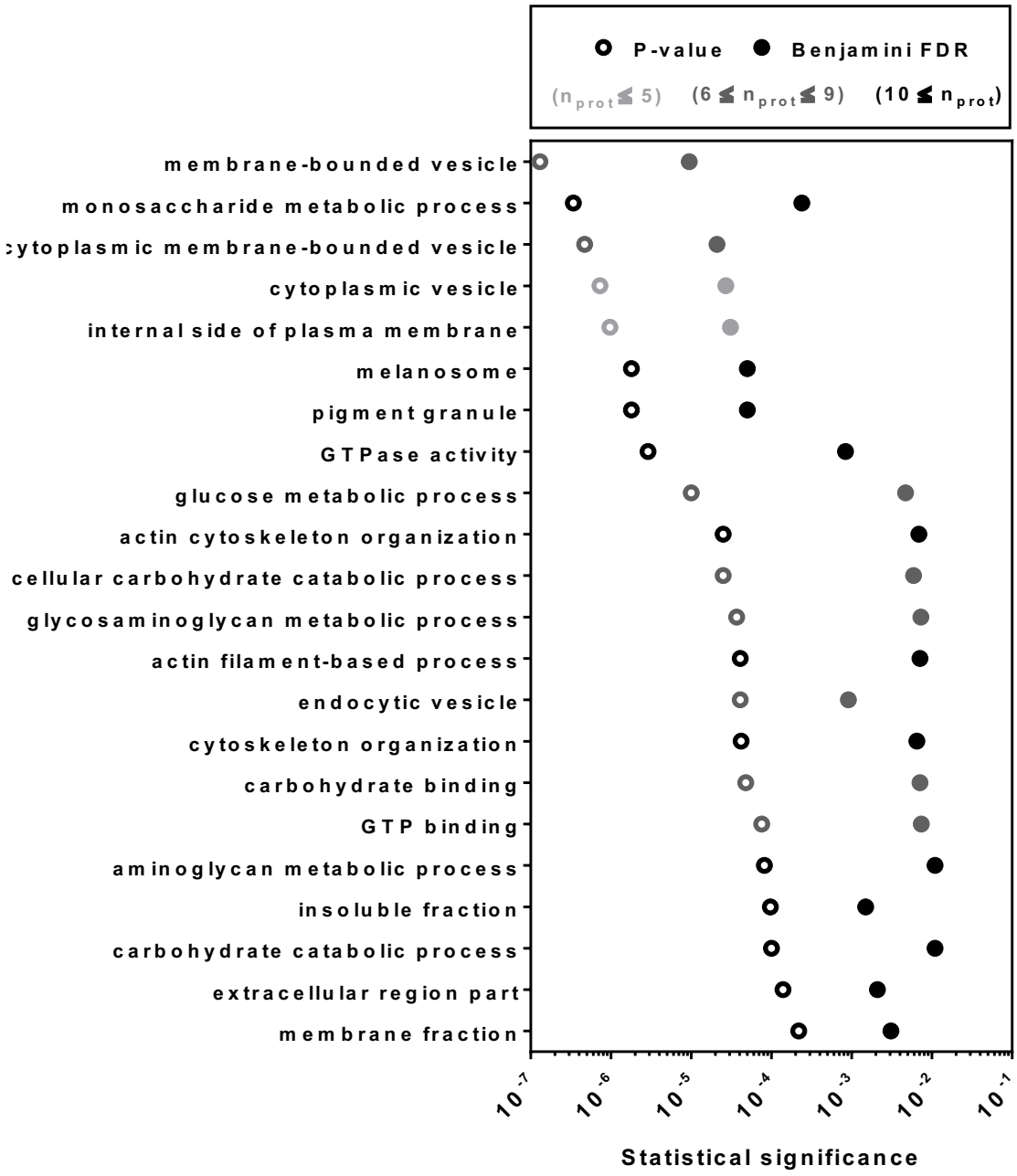


Figure 6S.3. **Gene ontology analysis of EV-enriched proteins**

Préface au chapitre 7

Ce chapitre est présenté sous la forme d'un article de recherche, publié dans le journal *Scientific Reports* : (Ref : Lefebvre, F. A., Bouvrette, L. P. B., Perras, L., Blanchet-Cohen, A., Garnier, D., Rak, J., & Lécuyer, É. (2016). Comparative transcriptomic analysis of human and Drosophila extracellular vesicles. Scientific Reports, 6, 27680.).

L'article propose une comparaison des propriétés morphologiques et du contenu transcriptomique des vésicules extracellulaires issues de deux lignées cancéreuses humaines et de deux lignées embryonnaires de Drosophile. Il met en évidence un degré de conservation élevé, tant au niveau morphologique que transcriptomique.

Dans le contexte de la présente thèse, les résultats exposées dans cet article font suite au développement de l'approche CeFra-seq (Chapitre 4) et regroupent les premiers travaux réalisés pendant mes études doctorales.

J'ai écrit l'intégralité du manuscrit. J'ai effectué les expériences et les analyses liées aux vésicules extracellulaires ainsi que la préparation des figures présentées dans cette publication. Louis Philip Benoit Bouvrette, étudiant au PhD au laboratoire, a contribué aux analyses de séquençage à haut débit. Lilyanne Perras, à l'époque une étudiante au baccalauréat en biochimie, m'a aidé à collecter les vésicules et à les caractériser par *nanoparticle tracking analysis*. Alexis Blanchet-Cohen, alors responsable de la plateforme de bio-informatique de l'IRCM, nous a aidé à compléter les analyses informatiques et a contribué aux discussions ayant mené à l'écriture du manuscrit. Delphine Garnier, alors stagiaire post-doctorale au laboratoire du Dr Janusz Rak, a contribué à m'initier aux techniques de purification de vésicules. Dr Éric Lécuyer, mon superviseur, a contribué à l'étude en amassant les financements requis et en relisant le texte pour y suggérer diverses améliorations.

Chapitre 7 : Comparative transcriptomic analysis of human and *Drosophila* extracellular vesicles (Article #5)

Chapitre 7 : Article #5

Comparative transcriptomic analysis of human and *Drosophila* extracellular vesicles

Fabio Alexis Lefebvre[1,2], Louis Philip Benoit Bouvrette[1,2], Lilyanne Perras[1,2], Alexis Blanchet-Cohen[1], Delphine Garnier[3,4], Janusz Rak[3], Éric Lécuyer[1,2,5,*]

1- Institut de Recherches Cliniques de Montréal (IRCM), Montréal, QC H2W 1R7, Canada
2- Département de Biochimie, Université de Montréal, Montréal, QC H3T 1J4, Canada
3- Montreal Children's Hospital, Research Institute of the McGill University Health Centre, Montreal, QC H4A 3J1, Canada
4- INSERM UMR 1064-ITUN; CHU de Nantes, Nantes 44093, France
5- Division of Experimental Medicine, McGill University, Montréal, QC H3A 1A3, Canada

***-** Corresponding author:

Dr. Eric Lécuyer
IRCM, RNA Biology Laboratory
110 Avenue des Pins, Ouest
Montréal, Québec, Canada, H2W 1R7
Email: eric.lecuyer@ircm.qc.ca

Abstract

Extracellular vesicles (EVs) are membrane-enclosed nanoparticles containing specific repertoires of genetic material. In mammals, EVs can mediate the horizontal transfer of various cargos and signaling molecules, notably miRNA and mRNA species. Whether this form of intercellular communication prevails in other metazoans remains unclear. Here, we report the first parallel comparative morphologic and transcriptomic characterization of EVs from *Drosophila* and human cellular models. Electronic microscopy revealed that human and *Drosophila* cells release similar EVs with diameters ranging from 30 to 200 nm, which contain complex populations of transcripts. RNA-seq identified abundant ribosomal RNAs, related pseudogenes and retrotransposons in human and *Drosophila* EVs. Vault RNAs and Y RNAs abounded in human samples, whereas small nucleolar RNAs involved in pseudouridylation were most prevalent in *Drosophila* EVs. Numerous mRNAs were identified, largely consisting of exonic sequences displaying full-length read coverage and enriched for translation and electronic transport chain functions. By analogy with human systems, these sizeable similarities suggest that EVs could potentially enable RNA-mediated intercellular communication in *Drosophila*.

Introduction

Although unprotected RNA molecules display short half-lives in biological fluids such as human serum[1], circulating RNAs can be stabilized in some circumstances, notably via their incorporation within extracellular vesicles (EVs)[2-4]. EV is an umbrella term referring to membrane-delimited nanostructures released by many eukaryotic and prokaryotic cells[5]. Exosomes, an intensively studied class of small EVs (40-100 nm), arise as intraluminal vesicles in endosomal compartments, while microvesicles, or plasma membrane-shed EVs, tend to be larger (100-400 nm) and are released via an actin-dependent abscission process[5,6]. Longly dismissed as cellular debris, EVs currently stand as established intercellular shuttles of genetic material, with functional implications particularly salient in immunology[3,7] and cancer biology[2,8]. Transcriptomic studies have revealed complex, cell type-specific extracellular RNA (exRNA) repertoires in EVs from diverse biological fluids and cell lines[9-11]. Tumor-shed EVs can transfer functional transcripts to stromal cells and thus remodel the tumor microenvironment[12-15]. Intercellular shuttling of miRNA activity has been described in diverse systems[13,14]. At least one EV-associated long intergenic non-coding RNA (lincRNA) contributes to tumor progression in a hepatocellular cancer model[12], while the transfer of EVs containing *GFP* mRNA leads to fluorescent protein expression in recipient endothelial cells[16]. Together, these results emphasize the functionality of diverse exRNA biotypes in various mammalian models.

Recently, increasing characterization efforts have extended our appreciation of exRNA repertoires in multiple biological species. For instance, lipid vesicles released by the Gram-negative bacterium *Vibrio cholerae* contain diverse RNAs enriched for intergenic sequences[17]. Exosome-like structures released by the protozoan *Trypanosoma cruzi* encapsulate large transcript populations dominated by short sequences derived from ribosomal RNAs (rRNA) and transfer RNAs (tRNA)[18]. Meanwhile, a comparative analysis of fungal exRNA identified a predominance of small nucleolar RNAs (snoRNA) and tRNAs in *Saccharomyces cerevisiae*, *Candida albicans*, *Paracoccidiodes brasiliensis* and *Cryptococcus neoformans*[19]. The arthropod *Drosophila melanogaster* stands as a key model organism that has enabled discoveries of paramount importance over the last century. In blastoderm *Drosophila* embryos, a high proportion of mRNAs adopts spatially resolved patterns, accumulating near subcellular

structures such as plasma membrane domains[20,21]. EVs have been implicated in *Drosophila* larval development, where a pool of the Wnt ligand is released from imaginal discs in association with exosomal membranes[22,23], possibly contributing to dissemination of the morphogenic signal and resulting cell fate and body patterning commitments. Although proteomic analyses have identified several novel factors in EVs purified from *Drosophila* cell cultures[24], the exRNA repertoire remains, to our knowledge, hitherto unexplored. In this comparative study, we used a uniform experimental pipeline to characterize EVs and define exRNA repertoires in two *Drosophila* and two human cell lines. Our morphologic and transcriptomic observations reveal considerable similarities across EVs from these distant metazoan systems: they contain comparable amounts of RNA largely consisting of short ribosomal sequences, retrotransposons, other non-coding RNAs and mRNA signatures enriched for translational functions.

Results and Discussion

Size characterization of human and *Drosophila* EVs

We investigated EVs in two *Drosophila* cell lines, Dm-D17-c3 (D17), derived from third instar larvae haltere discs[25,26], and S2R+, a macrophage-like S2 isolate from a late stage embryo primary culture[27]. Both are semi-adherent cell lines expressing hemocyte markers that are characterized by the activation of diverse survival pathways[28]. In contrast to S2R+, D17 cells are highly motile and can form cell-cell junctions[25]. Since human tumor-shed EVs have received considerable attention, we reasoned that inclusion of such models in our analysis along with *Drosophila* samples would provide instructive comparisons. We opted for the EGFR-driven, epidermoid carcinoma line A431[29] and the highly differentiated hepatocellular carcinoma line HepG2[30].

As a first approach to EV profiling, we performed nanoparticle tracking analyses (NTA) using a Nanosight device on cell culture supernatants cleared of floating cells (30 min at 2000×g). We calculated the total number of particles found in each preparation based on NTA [μ_{HepG2}= (9.01 ± 1.92) $\times 10^{10}$; μ_{A431}= (8.80 ± 0.76) $\times 10^{10}$; μ_{D17}=(6.72 ± 2.2) $\times 10^{10}$; μ_{S2R+}=(1.24 ± 0.35) $\times 10^{11}$]. The differences found when comparing HepG2 and A431 EV counts (P=0.84, Mann-Whitney), D17 and S2R+ counts (P=0.057, Mann-Whitney) and all four cell lines (P=0.044, Kruskal-Wallis) were not clearly significant (**Figure 7.S1**). All samples contained nanostructures of heterogeneous size, displaying characteristic Gaussian distributions ($0.87 \leq R^2_{Gaussian} \leq 0.91$; **Figure 7.1A**). Particle diameters of the two *Drosophila* lines were not significantly distinct from one another [μ_{D17}=151.0 ± 2.9 nm ; μ_{S2R+}=150.9 ± 3.0 nm ; P=0.97, t-test], while human A431 particles were slightly larger than their HepG2 counterparts [μ_{A431}=238.6 ± 3.4 nm ; μ_{HepG2}=219.3 ± 3.2 nm ; P=0.002, t-test]. Most notably, the mean diameters of human particles were significantly larger than those of *Drosophila* samples (P=0.008, one-tailed t-test). Analysis of cell media supplemented with depleted FBS revealed the absence of contaminant particles. Having established that all four cell models release nanostructures, we performed a standard EV differential ultracentrifugation-based purification. Purified EVs from *Drosophila* D17 and human HepG2 cells were analyzed by transmission electron microscopy (TEM), confirming the prevalence of cup-shaped, exosome-like structures (**Figure 7.1B**). Visual inspection of TEM images suggested that disrupted membrane fragments

or protein aggregates were generally absent from the preparations. We took advantage of electron micrographs to carry out comparative manual quantifications of EV diameters (**Figure 7.1C**). *Drosophila* D17 EVs displayed a significantly smaller diameter than human HepG2 EVs (μ_{D17}=47.9±1.8 nm ; μ_{HepG2}=62.8±2.1 nm, P <10^{-4}, t-test). Although the size distributions determined by NTA and TEM partially overlapped, we observed a slight discrepancy when comparing the two techniques, with NTA pointing at larger average size than TEM. Indeed, small EVs (<50 nm) were abundant in micrographs of HepG2 and especially D17 EVs, but NTA profiles were largely exempt of bodies smaller than 75 nm. It should be noted that uranyl acetate staining can alter EV morphology, likely causing EVs to collapse and present smaller diameter than what would be observed in solution[31]. Furthermore, aggregates formed by several EVs may be interpreted as single particles by the Nanosight instrument, leading to an overestimation of EV diameter distributions in NTA results. We sometimes observed aggregated EVs during TEM analyses in spite of extensive agitation, a feature that likely arises as a consequence of the ultracentrifugation procedure. Aggregation appeared more prevalent in the case of HepG2 than D17 EVs, which may underlie inherent differences in surface properties of EVs, possibly mediated by the expression of specific membrane proteins. HepG2 cells are considerably more adhesive than D17 cells, which are easily detached from plastic surfaces without trypsin. Whether EVs mirror the adhesion properties of their parental cell type is an interesting and seemingly unexplored question.

Discrepancies between NTA and TEM could thus result from morphological alterations induced by sample preparation for TEM, errors in NTA underlying aggregation, inherent limitations of each technique or a combination of these factors. Previous studies have relied on cryo-electron tomography (ET)[32] to circumvent the artifacts associated with heavy metal staining while assessing aggregation and derive reliable estimates of tridimensional EV diameter. In a future study, it would be interesting to systematically contrast NTA results with TEM and ET estimates of *Drosophila* EV size to yield a more robust comparison. Nonetheless, our TEM and NTA results both indicate that *Drosophila* D17 EVs are smaller than human HepG2 EVs.

Human and *Drosophila* EVs enclose complex populations of protected small RNAs

We extracted and quantified protein and DNAse-treated RNA from biological triplicates of *Drosophila* S2R+ and D17 EVs, in conjunction with human HepG2 and A431 EVs. We found

that all EV samples, collected over a 48h window, contained ~100 - 250 µg of protein and ~200 - 650 ng of RNA (**Figure 7.S2**). No significant difference was found between the variance of these distributions (*P*=0.11 and *P*=0.91, ANOVA). We took advantage of total EV counts determined by NTA and attempted to infer estimates of total protein and RNA mass per EV. This effort was motivated by recent reports showing that miRNA counts per EV are very low, less than one copy per EV on average, raising doubts about the potential functional impact of EV-mediated miRNA transfer[33]. Although the precision of our quantifications are limited by the biases of NTA count estimations, the sensitivity of photometric biomolecule quantification and the inherent heterogeneity of EV size and composition, we found that individual EVs isolated from all cell lines contain ~1-3 femtogram/fg (10^{-15} g) of protein [μ_{HepG2}=2.61 ± 0.74 fg ; μ_{A431}=2.37 ± 34 fg ; μ_{D17}=2.17 ± 0.77 fg ; μ_{S2R+}=1.29 ± 0.42 fg]. Our calculations indicate that S2R+ EVs contain significantly more protein (*P*=0.0045, Mann-Whitney) than D17 EVs. Individual RNA quantification revealed lower values in the range of attograms/ag (10^{-18} g) that displayed strong variability [μ_{HepG2}=5.71 ± 2.60 ag ; μ_{A431}=5.83 ± 3.07 ag ; μ_{D17}=4.26 ± 2.07 ag ; μ_{S2R+}=2.57 ± 1.20 ag] (**Figure 7.S3**). Arithmetic estimates indicate that 1.0 ag of a single-stranded RNA accounts for approximately 37 copies of a molecule containing 50 nt, or 3.7 copies for a 500 nt-long molecule[34]. Although our experimental setup doesn't enable estimations of specific RNA molecule counts, these NTA results are consistent with the presence of a few dozens of RNA molecules in an average EV for all cell types considered. Whether such amounts are sufficient to drive functional changes in recipient cells upon *in vivo* EV transfers remains unclear.

We next performed an RNAse protection assay, where fresh EV pellets were split in two equal parts, either submitted to an RNAse A treatment followed by RNA extraction or a direct RNA extraction. Only a minor proportion of exRNA was degraded upon RNAse A treatment in all 4 EV types, a difference that was not statistically significant, while parallel treatments of total RNA extracted from HepG2 and D17 cells led to complete degradation (**Figure 7.S4**). These results are consistent with a topological exclusion of exRNA from the surrounding solution by intact EV lipid membranes. We then conducted bioanalyzer capillary electropheresis on all exRNA and total cellular RNA types to compare their size distributions. Small species ranging in size from 50 - 250 nt were most prevalent in all exRNA electrophoretic profiles, although signatures of longer transcripts were also present (**Figure 7.S5**). Human and *Drosophila* exRNA

profiles were highly similar and mature 18S and 28S peaks were largely absent of all preparations, in accordance with previous reports[2,3,19,30].

Ribosomal RNA and related pseudogenes are predominant in human and *Drosophila* EVs

Morphological and structural similarities shared between human and *Drosophila* EVs led us to investigate exRNA repertoires in both species. We performed RNA-seq on duplicates of *Drosophila* D17 and S2R+ exRNA, along with human HepG2 and A431 exRNA. To provide unaltered portrayals of exRNA, we chose not to perform any selection or depletion procedure prior to sequencing. To enable comparisons with matching cell transcriptomes, we analyzed duplicates of rRNA-depleted libraries from HepG2 and D17 cellular RNA in parallel. While most previous transcriptomic studies examining EVs have opted to sequence the small RNA fraction, we chose to generate standard sized libraries, which should capture the bulk of EV RNAs as determined by our bioanalyzer profiles (**Figure 7.S5**). Chiefly, it should emphasize protein-coding and long non-coding RNA repertoires, which have received less attention than EV-associated miRNAs and would typically not be traced in the small RNA fraction. However, our approach cannot emphasize the presence of transcripts shorter than 50 nt, such as mature miRNAs, which would require the preparation of a small RNA-seq library.

Various rRNA sequences were predominant in all exRNA libraries (**Table 7.1**). Pseudogenes derived from 5S rRNA, such as human *RNA5SP145* and *RNA5SP149*, were especially abundant (**Figure 7.S6**), along with the mitochondrial *RNR1* (12S) and *RNR2* (16S) rRNAs (**Figure 7.S7**). Full-length bidirectional read distributions mapped to these loci in human EVs and similar species prevailed in *Drosophila* exRNA (**Table 7.S1**). The 28S and 18S ribosomal RNAs were the single most abundant sequences in all samples. This observation is in sharp contrast with the scarcity of the corresponding ~2 and ~3 kb peaks on EV bioanalyzer imprints. Such apparent inconsistency could underlie the cleavage of EV-targeted 18S and 28S to yield shorter transcripts. Indeed, although read coverage for most transcripts in EVs mimicked corresponding cellular signatures, our analyses can't ensure that they correspond to full-length RNAs. Since we sequenced 50 nt-long paired-end reads, arrays of cleaved RNA fragments could have potentially resulted in similar coverage signatures. Ribosomal RNA turn-over remains poorly characterized and while the 18S and 28S rRNAs are associated with longer cytoplasmic half-lives than polyadenylated RNAs, rRNA stability in fibroblasts is considerably affected by

growth conditions[35]. It is tempting to speculate that EV targeting could serve as a selective clearance mechanism for rRNAs, a view that would be compatible with fragmentation of full-length 18S and 28S. Abundance of rRNAs in EVs has been reported in some systems, notably human serum and urine samples[36], although most studies have rather outlined the absence of long rRNAs in EVs[3,10] . These differences likely reflect both inherent specificity of the model considered and contrasting experimental strategies (eg. small vs long RNA libraries). In some models, apoptotic bodies and microvesicles have been reported to contain higher proportions of rRNAs than exosomes[10]. While several large structures (>100 nm) seen on our TEM micrographs may constitute microvesicles, apoptotic bodies should be rare in our preparations, since we routinely monitored cell death before our experiments and consistently observed rates lower than 5% in all cultures.

Besides hypothetical fragmentation, ambiguous mapping issues may also account for the discrepancy between bioanalyzer profiles and abundant 18S and 28S sequencing reads. Indeed, we found two poorly characterized rRNA-like short sequences (<200 nt) that were highly abundant in human EVs and relatively rare in cells, *AC079949.1* and *AL161626.1,* annotated as "novel miRNAs" in Ensembl and as "rRNA pseudogenes" in GATExplorer. These miRNA loci overlap with large ribosomal subunit repeats[37] (**Figure 7.S8**), suggesting that reads arising from these species or similar rRNA-related sequences could potentially have been mapped to 28S genes by read alignment tools, inflating the proportion of rRNA reads in EVs. *AC079949.1* and *AL161626.1* were first identified in human EV extracts and are mentioned in a patent request regarding the use of mesenchymal stem cell-derived EVs for wound therapy[38]. Asymmetric and bidirectional read distributions consistent with an abundant sense-antisense RNA pair were observed for *AC079949.1* in exRNA (**Figure 7.S8**). Such sense-antisense pairs can anneal and form double-stranded RNA, initiating interference pathways through Dicer activation[39,40]. While the functional role of *AC079949.1* remains to be determined, the contribution of short transcripts to rRNA regulation is demonstrated in mammalian cells[41]. Moreover, it is becoming clear that diverse mammalian tissues express "specialized" ribosomes that bear diverse rRNA, ribosomal proteins and isoform specificities[42]. EV transfer could potentially modulate recipient cell ribosomal repertoire, perhaps through phenocopying of donor cell rRNAs. It is thus tempting to speculate that the abundance of rRNA-related "novel miRNA" species in EVs reflects an additional layer of intercellular regulation in ribosomal biogenesis or translational fine-tuning.

exRNA distributions correlate across cell types in human and *Drosophila*

To further characterize EV RNA repertoires, we next subtracted rRNA reads and submitted adjusted libraries to genomic alignments and expression analyses (**Table 6.2**). We retrieved GENCODE annotations using Ensembl Biomart for the 1,000 most abundant transcripts in each library and compared exRNA and cellular RNA biotype abundance distributions (**Figure 7.2**). Interestingly, a strong positive correlation was observed between human HepG2 and A431 exRNA biotypes (**Figure 7.2A**; Pearson's r=0.92, $P=3\times10^{-3}$). The miscellaneous RNA category was strongly overrepresented in human exRNA, accounting for over one third of A431 and over half of HepG2 exRNA distributions, as opposed to 1% for HepG2 cell RNA. In *Drosophila* D17 EVs, snoRNAs represented nearly half of the distribution, whereas S2R+ EVs contained a majority of mRNAs (**Figure 7.2B**).

We then compared EV and cellular RNA libraries by correlating relative abundance values of individual transcripts (**Figure 7.3-7.4**). Although many abundant cellular RNAs were present within EVs, HepG2 exRNA levels were poorly correlated to cellular levels (**Figure 7.3A**; Pearson's r=0.24, $P<10^{-4}$) and more closely resembled A431 exRNA levels (**Figure 7.3B**; Pearson's r=0.77, $P<10^{-4}$). In accordance with previous reports, three groups of "miscellaneous" Polymerase III products[43] were strongly overrepresented in human exRNA samples: vault RNAs, Y RNAs and signal recognition particle components, 7SL RNAs. Interestingly, vault RNAs were nearly absent of HepG2 cellular libraries, while the relative levels of vault paralogues was nearly identical in A431 and HepG2 EVs (**Figure 7.3C**). Although the vault ribonucleoprotein (RNP) complex has been linked to antineoplastic drug resistance[44-46], the bulk of vault RNA transcripts do not associate with this RNP complex[47]. A fraction is rather processed by Dicer into miRNA-like regulatory transcripts that can downregulate the catabolic cytochrome CYP3A4[48]. While *VTRNA1-1* shows full-length coverage in exRNA (**Figure 7.S9**), its eventual intercellular transfer could potentially impact xenobiotic metabolism in recipient cells. In addition, a strong dissymmetry was noted among 7SL paralogue distribution : *RN7SL3*, the sole detectable cellular paralogue, was barely present in exRNA, while *RN7SL1* and *RN7SL2* accounted for over 90% of total 7SL abundance in exRNA and were undetected in HepG2 cells (**Figure 7.3D**). Multiple sequence alignments[49] revealed that *RN7SL1* and *RN7SL3* have nearly identical sequences

(97.6% identity score). Presumably, the few nucleotides that distinguish these paralogues lead to the establishment of differential interactions resulting in extensive EV targeting of *RN7SL1* and cellular retention of *RN7SL3*.

In accordance with human data, *Drosophila* D17 exRNA levels presented lower correlations with corresponding cellular levels (**Figure 7.4A**; Pearson's r=0.45, $P<10^{-4}$) than with S2R+ exRNA (**Figure 7.4B**; Pearson's r=0.58, $P<10^{-4}$). Multiple snoRNAs were overrepresented in D17 exRNA samples, especially H/ACA box species involved in site-specific pseudouridylation of 18S and 28S rRNA (**Table 7.S2**). Components of atypical snoRNPs were also abundant in human and *Drosophila* exRNA, notably human *RPPH1* and *Drosophila RNAseP:RNA*, which function in tRNA maturation[50]. The human and *Drosophila* transcripts of RNAse MRP, a multifunctional ribozyme notably involved in 5.8S rRNA processing[51], were abundant in both exRNA and cellular RNA, along with diverse paralogues of spliceosomal U5 snRNA (**Figure 7.S10**). The *Drosophila CG13900* gene, which encodes a putative spliceosomal factor[52], contains 9 snoRNAs within its introns, two of which were highly abundant in exRNA (**Figure 7.S11**). Read coverage at this locus strongly suggests that mRNAs and associated intronic snoRNAs constitute independent transcriptional units, with divergent fates regarding their incorporation into EVs. We also observed low levels of diverse tRNAs in human and *Drosophila* EVs (**Figure 7.S12**). Several, but not all *Drosophila* snoRNAs and tRNAs, displayed asymmetric read coverage relative to corresponding cells, similar to the pre-miRNA *AC079949.1* discussed above. Such patterns are intriguing in light of accumulating evidence of miRNA-like transcripts derived from tRNAs and snoRNAs involved in RNA interference pathways[53-55]. Indeed, a recent study found that most small RNAs in mature mammalian sperm correspond to 3'-end fragments of tRNAs (tsRNAs), which may modulate cholesterol metabolism in the offspring[56]. Whether snoRNAs and tRNAs that encode miRNA-like substrates are preferentially targeted to EVs remains unclear, but has been suggested[57]. Investigating snoRNAs and tRNAs after selecting for the small RNA population of *Drosophila* EVs may provide further interesting observations regarding tsRNAs.

Transposable elements are conserved EV components in human and *Drosophila*

Transposable sequences such as long interspersed elements (LINEs) and *Alu* elements have been described as major components of exRNA, notably in glioblastoma models[9,58]. We

took advantage of the RepeatMasker inventory of interspersed repeats and low complexity sequence genomic coordinates to systematically survey repeats in exRNA[59]. Alignments revealed that repeats are collectively overrepresented in exRNA relative to cellular RNA samples, especially in *Drosophila* (**Table 7.2**). Interestingly, the single most abundant repeat sequence in *Drosophila* D17 exRNA was a short (~150 nt) antisense fragment of the retrotransposon *TART*, a telomere-specific LINE-like element involved in chromosome length maintenance[60] (**Figure 7.S13**). Upon DNA replication, yeast and mammalian cells depend on the reverse-transcriptase activity of the telomerase complex to regenerate G-rich repeats and assemble the end of chromosomes[61-63]. Drosophilids rely on retrotransposition of telomeric LINE-like elements as an alternative solution to the end-replication problem[64-66]. Other *Drosophila* retrotransposons, notably *Copia* and related sequences were also highly abundant. This 5 kb-long element displayed full-length coverage in S2R+ and D17 exRNA (**Figure 7.S14**). Interestingly, as observed in the case of *TART*, a resolved antisense peak of approximately 50 nt mapped to the central region of the *Copia* sequence in D17 and S2R+ exRNA. Repeat-associated small interfering RNA (rasiRNA) form a class of Piwi-interacting transcripts involved in transposable element silencing during *Drosophila* gonadogenesis[67-70]. Whether rasiRNAs are present in *Drosophila* exRNA is unclear, but it is tempting to suggest that the short antisense sequences mapped to *Copia* and *TART* in exRNA may function as small interfering RNA precursor.

Human endogenous retrovirus long terminal repeats, notably members of the ERV1, ERVL and ERVK subfamilies were the most abundant repeats in HepG2 and A431 exRNA. A few full-length L1 retrotransposons, several Alu, 7SL and tRNA repeats were also prevalent in human exRNA (**Table 7.S3**). By contrast, α-satellite sequences and low-complexity repeats such as poly-purine or poly-pyrimidine tracts were rare in human exRNA samples. While evidence remains scarce, it is tempting to presume that extensive EV targeting of retroviral sequences constitutes an adaptation that promotes invasion via horizontal transfers.

Human and *Drosophila* exRNA contain full-length mRNA signatures enriched for translation-related functions

Over 1,000 mRNA signatures were traced in all exRNA samples. We validated RNA-seq results by RT-qPCR analyses targeting 7 mRNAs in exRNA and cellular RNA of *Drosophila* D17 and human HepG2 samples. As expected, FPKM and cycle threshold values (CT) were significantly anti-correlated ($-0.92 \leq$ Pearson's $r \leq -0.75$) (**Figure 7.5A-B**). Importantly, D17 exRNA displayed low amounts of mRNAs, which were more abundant in S2R+ exRNA (**Figure 7.2B**) and mRNA level comparisons between D17 EVs and cells revealed a weaker correlation than global comparisons (**Figure 7.5C**; Pearson's r=0.24, $P<10^{-4}$). By contrast, human HepG2 mRNA levels in EVs were closer to corresponding cell levels than described above for global comparisons in human HepG2 (**Figure 7.5D**; Pearson's r=0.32, $P<10^{-4}$). Accordingly, several highly expressed cellular mRNAs were traced in EVs, notably mitochondrial mRNAs involved in the respiratory chain, such as cytochrome oxidase, NADH:ubiquinone reductase and ATPase subunit mRNAs (**Table 7.S4**). Numerous ribosomal protein and translation elongation factor mRNAs were also abundant in all samples. Protein-coding transcripts typically displayed full-length read coverage in cells and EVs, consistent with the EV export of mature and potentially functional mRNAs, as observed for the ribosomal gene *RPLP1* in human and *Drosophila* (**Figure 7.5E-F**). Gene ontology enrichment analyses identified 18 terms common to all four exRNA types, 12 of which contain the words "translation" or "ribosome" (**Figure 7.5G-H** and **Table 7.S5**). In addition, the ferritin light chain transcript was among the most abundant mRNAs in human HepG2 and A431 exRNA. Ferritin proteins assemble into large shell-like complexes that enclose and store iron ions[71]. These proteins have been identified in EVs of immune origin from mice and human models[72-74]. Interestingly, another important iron regulator, the transferrin receptor, figures among the first factors described in human EVs and its release from reticulocytes is linked to their maturation into erythrocytes[75-77]. EV targeting of these mRNAs and encoded proteins may reflect an intricate layer of iron metabolism regulation. The most abundant mRNA in *Drosophila* D17 exRNA was *Arc1* (*Activity-regulated cytoskeletal protein 1*), orthologous to a mammalian plasticity protein involved in synaptogenesis[78]. Indeed, mammalian *Arc* mRNA is targeted to rat neuronal dendrites via a *cis*-regulatory motif found in its 3'UTR[79]. If this mechanism is conserved in *Drosophila*, it could promote *Arc1* accumulation to plasma membrane domains, resulting in its preferential incorporation within EVs.

Whether a subset of cellular mRNAs undergoes selective targeting to EVs through a sequence-specific mechanism remains unclear. If specific RNA motifs are involved in sorting mRNAs to EVs and if such motifs are conserved, a fraction of orthologous human and *Drosophila* sequences should exhibit a common propensity to accumulate in EVs. To test this hypothesis, we took advantage of our vast EV repertoire of human and *Drosophila* mRNAs to investigate global abundance correlations across orthologs of the two species. Using an integrative ortholog prediction tool[80], we retrieved 1,140 pairs of orthologous mRNAs represented in our D17 and HepG2 cellular and EV datasets. We first compared cellular relative abundance values, which revealed a strong positive correlation (**Figure 7.S15**; Pearson's r=0.61, $P<10^{-4}$). These distributions were largely dominated by abundant mRNAs encoding ribosomal proteins. When comparing the relative abundance distributions of orthologs in EV, we documented a slightly weaker correlation (**Figure 7.S15**; Pearson's r=0.54, $P<10^{-4}$). Therefore, while the expression levels of orthologous mRNAs are strongly related, our comparative analysis does not suggest that gene-specific mRNA enrichment to EVs is a globally conserved feature.

Previous comparative studies have identified so-called "exRNA-exclusive" transcripts, likely undetected in cognate cells due to highly efficient subcellular transport processes, high cellular turnovers or a combination thereof[81]. To systematically and stringently survey exRNA exclusive transcripts, we compared exRNAs reaching a 5 FPKM threshold to all transcripts detected in cognate cellular libraries. Interestingly, several of the 24 "exRNA-exclusive" *Drosophila* mRNAs encode neuronal membrane-associated proteins, notably *Snap25* (Synaptosomal-associated protein)[82], along with extracellular matrix factors, such as members of the mucin family (*Muc26B*, *Mu4B*)[83] (**Table 7.S6**). The functional relevance of expressing neuronal factors in haltere disc cells is unclear and extensive EV targeting of corresponding mRNAs could constitute a strategy to clear transcriptional noise products and prevent aberrant protein expression. This interpretation is in line with the membrane protein clearance function of EV targeting, well established during reticulocyte differentiation to erythrocytes[76]. Considering the prevalence of mRNA localization and spatially restricted translation during *Drosophila* embryogenesis[21,84], it also appears conceivable that functional cis-regulatory RNA elements present in these sequences promote their targeting to membrane domains, as suggested above for *Arc1* mRNA. Among the 5 exRNA-exclusive human HepG2 mRNAs, *ALB* (*Albumin*) and *APOB* (*Apolipoprotein B*) encode secreted proteins involved in transmembrane transport[85,86] while

TSPN16 (*Tetraspanin 16*) belongs to a class of membrane-spanning factors described as EV protein markers[85,86]. In line with our findings in *Drosophila*, the exclusivity of these mRNAs within EVs suggests the prevalence of instructive targeting signals in their sequence that promote local translation of protein factors associated with membrane or extracellular localization.

Conclusion and perspective

To the best of our knowledge, we provide here the first morphological and transcriptomic comparative analysis of human and *Drosophila* EVs. Our work revealed that several features of EVs are considerably conserved in these distant metazoan species, notably the abundance of ribosomal sequences and retrotransposons, including sense-antisense RNA pairs. *Drosophila* EVs released by S2R+ and D17 cells are enriched for diverse snoRNAs, while human EVs produced by HepG2 and A431 cells contain strong signatures of miscellaneous RNAs, such as vault and Y RNAs. While EM and NTA analyses have shown considerable heterogeneity in our preparations regarding EV size and morphology, transcriptomic analyses were performed on EV populations, and therefore can't determine whether individual EVs display disparate RNA repertoires. Diverse methods such as immunoaffinity captures or density-based separation have been developed to isolate exosomes from preparations containing larger vesicles[87], and it would be interesting to optimize these tools for *Drosophila* EVs. Previous studies have suggested that mammalian exosomes and microvesicles may contain contrasting molecular signatures, with exosomes showing higher amounts of RNA than microvesicles[10]. Approaches amenable to single EV capture and RNA sequencing have yet to be developed due to the prohibitively small size of these structures, but would be highly appealing since they may reveal the heterogeneity of EV RNA repertoires, notably in *Drosophila*. Furthermore, the use of long-read RNA sequencing platforms[88], such as single-molecule real-time sequencing (15,000 bp per read) or pyrosequencing (700 bp per read), represent a interesting avenue to further characterize exRNA populations. Such approaches would help illuminate the precise repertoire or RNA isoforms and fragmentation intermediates present within EV specimens. Refined means of intercellular communication have emerged and expanded throughout the evolutionary history of metazoans. While exRNA shuttling likely contributes to cell-cell signaling in mammals, prevalence and functional relevance of the phenomenon remains unexplored in *Drosophila.* In light of our findings, it seems likely that diverse *Drosophila* lineages, in line with their human counterparts,

rely on exRNA to convey intercellular communication. This hypothesis could be explored via loss-of-function studies.

Material and Method

Cell culture and EVs purification

Human HepG2 and A431 cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 1% of a 1:100 solution of penicillin and streptomycin (pen/strep) and 10% fetal bovine serum (depleted FBS) (Wisent). Before use, FBS was depleted of bovine EVs by ultracentrifugation (110,000 g, 18h, 4°C). Human cells were cultured at 37 °C, in an atmosphere of 5% CO_2 in T-flasks of 175 cm^2 and routinely detached using 0.25% trypsin upon reaching 80% confluence. *Drosophila* D17 and S2R+ cells were respectively maintained in Shield and Sang Insect Medium (M3) and in Schneider medium. Both media were supplemented with 1% of a 1:100 pen/strep solution and 10% depleted FBS. M3 medium additionally contained insulin (20 µg/mL). *Drosophila* cells were cultured at 25°C. Cell death rate was routinely monitored by trypan blue staining (Sigma-Aldrich) and consistently remained below 5% for all cell types.

Isolation of EVs

For EV isolation, cells were kept at low passage (P<10) and cultured for approximately 48h starting from ~10^7 cells. EVs were isolated according to an established differential ultracentrifugation protocol[89]. Briefly, fresh culture supernatants (~80 mL) were cleared of floating (10 min at 400×g) and dead cells (10 min at 2000×g) using a 5810 R centrifuge (Eppendorf). Cell debris were removed (30 min at 10,000×g) and EVs were pelleted (70 min at 110,000×g) using the Sw28 and the Sw32 Ti rotor (Beckman Coulter) in an L8-70M machine (Beckmann). Preparations were extensively washed with PBS and pelleted (70 min at 110,000×g) using a RP100-AT4 rotor (Sorvall) and a RC-M100 micro ultracentrifuge (Sorvall). All steps were performed at 4 °C and EVs were immediately processed for analysis.

Electron microscopy

Steps were performed as previously described [89]. Briefly, 10 µL of fresh EVs preparations diluted in 100 µL PBS were loaded on previously discharged formvar-coated copper grids and

allowed to adhere for 20 min. The grids were then applied to freshly prepared drops of 2% uranyl acetate for 30 to 40 s and washed six times for two minutes with water. Excessive water was removed by absorption and grids were left to dry for 30 min. Samples were imaged on a Tecnai 12 120 kV transmission electron microscope. Contrast was enhanced with the software Photoshop (Adobe).

Nanoparticle Tracking Analysis

Cell depleted supernatants (30 min at 2000×g) were analyzed by nanoparticle tracking analysis using an LM-10 machine (Nanosight) according to the manufacturer's instructions. Samples were submitted to 3 successive analyses of 30 s using the default settings of the instrument. 3 washes were performed with water between samples. DMEM, M3 and Schneider media containing depleted FBS were used as negative controls. Biological triplicates were analyzed.

Isolation and characterization of exRNA

EVs and corresponding cellular pellets were resuspended in 1 mL of TRIzol™ reagent (Ambion) and processed according to the manufacturer's instructions RNA extracts were purified with the RNA Clean & Concentrator™-5 system (Zymo Research). *In-column* DNase I (New England BioLabs) treatment, RNA washes and elution steps were performed according to the manufacturer's instructions. One additional centrifugation step was included (5 min at 16,000×g) to ensure complete removal of the washing buffer. RNA samples were eluded in 12 µL of RNAse-free water (Wisent). Absorbance distributions were immediately quantified using a NanoDrop 2000c spectrophotometer. RNA samples were pure ($A_{260}/A_{280} \geq 2.0$)($2.00 \leq A_{260}/A_{230} \leq 2.25$). Aliquots of ~5 ng were submitted to capillary electrophoresis on a Bioanalyzer 2100 machine (Agilent). RNAse protection assay was performed as previously described [19] to confirm the intraluminal topology of exRNA within EVs. D17 and HepG2 cellular RNA extracts were used as controls. RNAse A (Qiagen) was inactivated by heat (10 min at 65 ºC) and RNA was extracted as described above. For calculations related to the estimation of RNA copy number per EV, we averaged the mass of each nucleoside at 325 Da. For example (1 ag of a 50 nt RNA): 10^{-18} g × 6.02 × 10^{23} molecules × mol^{-1} × (50 nt × $molecule^{-1}$ × 325 g × mol^{-1} × $molecule)^{-1}$ = 37.04 molecules.

Library generation for RNA RNA-seq

Biological duplicates of sequencing libraries were prepared from high quality RNA extracts (50 ng exRNA and 500 ng cellular RNA) using the Illumina TruSeq Stranded RNA Kit according to the manufacturer's instructions. The TruSeq PE Clusterkit v3-cBot-HS was used on an Illumina HiSEq 2000 machine.

***In silico* analysis of RNA sequencing data**

Read quality was confirmed using FastQC v0.10.1 and trimming was performed with Trimmomatic when deemed necessary. Read alignment was performed using Tophat v2.0.10 on the human GRCh37/hg19 and the *Drosophila* BDGP5.78/dm3 genomes, respectively. Alignment BAM files were used to generate bigWig files, which were submitted to the UCSC genome browser for read coverage visualization. Reads mapping to RepeatMasker v4.0.6 sequence coordinates were counted with BEDTools. Ribosomal RNA sequences were filtered out by first mapping the reads to FASTA files of genes annotated as rRNA in Ensembl. Remaining reads were then re-aligned to reference genomes with Tophat. BAM files were used for expression analyses with Cuffdiff v2.2.1 without effective length correction. Transcripts per million ($TPM_i=[FPKM_i/\Sigma_j FPKM_j]\times 10^6$) were used as relative abundance units. A 5 FPKM threshold was applied. For biotype and correlation analyses, the 1,000 most abundant transcripts were considered, which accounted for over 90% of identified transcripts.

Reverse transcription quantitative PCR (RT-qPCR)

The M-MLV reverse transcriptase (Invitrogen) was used according to the manufacturer's instructions to synthesize cDNA in triplicate starting from 100 ng of exRNA and cell RNA. Priming was performed with random hexamers (Promega). Real-time PCR was performed on the ABI ViiA7 instrument (Life) using SYBR Green reagent (Applied Biosystems). Cycling conditions were as follows : 50 °C for two minutes and 95 °C for 10 min followed by 40 cycles of 95 °C for 15 s, 59 °C for 60 s. Melting curve analysis was performed. Validated primer

sequences targeting exon junctions were retrieved from the GETprime database[86]. Sequences are shown below.

Table 6. I. Primer sequences

Homo sapiens sequences		
	Forward primer	Reverse primer
TPT1	5'-CACGATGAGATGTTCTCCG-3'	5'-TCCTACTGACCATCTTCCC-3'
PABPC1	5'-CACTGGCATGTTGTTGGAG-3'	5'-CTTCATCAACCTTAGAACGGAG-3'
ATF4	5'-ATGATTACCTGGAGGTGGC-3'	5'-CTCCTTGCTGTTGTTGGAG-3'
PTBP1	5'-AGTTCTTCCAGAAGGACCG-3'	5'-GTTGTGCAGGTCAATGAGG-3'
HDGF	5'-GTGACGGTGATAAGAAGGG-3'	5'-TTTAGGAGAGTCCTCCAGC-3'
G3BP1	5'-GTAGAGGAACCTGAAGAAAGAC-3'	5'- ATGTCATTACTGACAACTGCC-3'
BRAF	5'- CTATTGGACAAATTTGGTGGG-3'	5'-GTATTCTTCATAGGCCTCCAG-3'
Drosophila melanogaster sequences		
	Forward primer	Reverse primer
arc1	5'-TAGAAGGTATCAGCGACGAG-3'	5'-GCCATACCGTAGAACAGCA-3'
pabp2	5'-GCATACATTGAGTTTGGTTCC-3'	5'-CGACATTACCTTTATTTGACGC-3'
sqd	5'-CACGGCAAGATCTTTGTCG-3'	5'-TCCACCTCGACGATATTGC-3'
cam	5'-GAAACTCACAGACGAGGAG-3'	5'-CATCATAGTCACGAATTCTTCG-3'
rin	5'-CAAGGGTGACTTTGAGCAG-3'	5'-GACATTTCCGAAGCGTGAG-3'
hmu	5'-TCTACTGGACGGACTCCTC-3'	5'-TATTTGAAGAGACGGCCGG-3'
imp	5'-CTCTACGAATAAGGGTGAACTC-3'	5'-CGTCCAATCAAATTGTTGTGG-3'

Functional annotation analysis

The 1,000 most abundant mRNAs in each EV type were used for functional analysis. Process, function and component GO terms were retrieved with DAVID. Terms associated with FDR<10^{-3} were considered.

General Statistics

Statistic tests were performed with GraphPad Prism v6.

Data access

Datasets are available on GEO under the accession number GSE76173.

Acknowledgments

This project was supported by the Cancer Research Society of Canada (grant #20227) and start-up funding from the IRCM to EL, as well as a foundation grant from the Canadian Institutes of Health Research (CIHR) to JR. EL is a Junior 2 scholar of the Fonds de Recherche du Québec-Santé (FRQS) and a Louisa Gale Scholar of the Canadian Cancer Society. FAL was supported by scholarships from the Canadian Institutes of Health Research (CIHR), the National Sciences and Engineering Council of Canada, and from FRQS. DG was supported by a Thelma Adams Fellowship from the Montreal Children's Hospital Foundation.. The authors express their gratitude to core facilities staff, notably Jeannie Mui (McGill Facility for Electronic Microscopy) and Odile Neyret (IRCM Molecular Biology Platform).

Author contributions

The study was designed by FAL, DG, JR and EL. FAL generated and analyzed all data regarding *Drosophila* and human HepG2 EV morphology, while A431 EVs were purified by DG, who also contributed to nanoparticle tracking analyses and training of FAL. All RNA extractions and RT-qPCR were performed by FAL. LPBB, LP, AB-C and FAL all contributed to transcriptomic analyses. FAL mounted figures and wrote the manuscript. All authors contributed to manuscript revision.

Competing financial interests

We declare no potential conflict of interest.

Figure and table legends

Figure 7.1. Size distributions of EVs released by human and *Drosophila* cells

(a) Histograms depicting the diameter distribution of particles in cell-depleted supernatant, as determined by nanoparticle tracking analyses (Nanosight) for human (*H.s.*) and *Drosophila* (*D.m.*) cells. Mean (μ) and standard deviation (σ) with associated standard error measurements (s.e.m) are overlaid on histograms. **(b)** Representative transmission electron micrographs of human HepG2 (left panels) and *Drosophila* D17 (right panels) EVs purified from culture supernatants and stained with uranyl acetate. Scale bars = 50 nm. **(c)** Whisker plot of diameter distributions of HepG2 (left) and D17 (right) EVs, as determined by direct quantification of electronic micrographs. The number (n) of EVs quantified is indicated.

Figure 7.2. Human and *Drosophila* EVs enclose various types of transcripts

(a,b) Pie chart depictions of biotype abundances for the 1,000 most abundant transcripts identified by RNA-seq in human (A431 and K562) (a) and *Drosophila* (S2R+ and D17) (b) cells lines and their derived EVs. Biotype relative abundance was determined on the basis of TPM values. Reads mapping to rRNA were excluded from this analysis. Biotypes associated with values inferior to 0.5% were grouped into the "others" category.

Figure 7.3. Correlative comparisons of exRNA secreted by human cell lines

(a,b) HepG2 exRNA levels were compared to HepG2 cellular RNA (a) and A431 exRNA (b) levels. Pearson's correlations (r) and associated p-values are indicated at the top of each graph. Select groups of transcripts are identified. **(c,d)** Bar charts representing the distributions of sequencing reads mapped to paralogues of vault RNAs (c) and 7SL RNAs (d) in human EVs and cells. Values at the top of each column refer to the total number of reads mapped to these transcripts.

Figure 7.4. Correlative comparisons of exRNA secreted by *Drosophila* cell lines

(a,b) D17 exRNA levels were compared to D17 cellular RNA (a) and S2R+ exRNA (b) levels. Pearson's correlations (r) and associated p-values are displayed at the top of each graph. Select

groups of transcripts are identified. **(c,d)** Bar charts representing the distribution of sequencing reads mapped to lncRNAs (c) and snoRNAs (d) in *Drosophila* EVs and cells. Values at the top of each column refer to the total number of reads mapped to these transcripts.

Figure 7.5. Characterization of mRNAs secreted within Human and *Drosophila* EVs
(a,b) Comparative analysis of expression levels of select mRNAs via RNA-seq and qRT-PCR. Cycle threshold (CT) values determined by qRT-PCR are negatively correlated to FPKM values determined by RNA-seq for various *Drosophila* D17 (a) and human HepG2 (b) mRNAs in EVs and cells. **(c,d)** Relative expression levels of mRNAs extracted from D17 (c) or HepG2 cells and EVs (d). Select groups of transcripts are identified. Pearson's correlations (r) and associated p-values are indicated at the top of each graph. **(e,f)** UCSC genome browser views of *Rplp1* mRNA shows strictly exonic read coverage in human HepG2 (e) and *Drosophila* D17 (f) EVs and cells. **(g)** Venn diagram depicting the overlap of enriched gene ontology (GO) terms displayed for human and *Drosophila* EVs mRNAs. The number of enriched GO terms retrieved per sample is shown. **(h)** Examples of translation-related GO terms identified in all EV samples. Associated false discovery rates (FDR) are provided.

Figure 7.S1. EV count per preparation
EV counts were determined on the basis of NTA analysis, considering total particles in cultures established from 10^7 cells over the course of 48h. The significance of the differences observed between lines of the same species was evaluated through a Mann-Whitney test, while the prevalence of a global difference was evaluated using a Kruskal-Wallis test. Differences are deemen non-significant.

Figure 7.S2. Human and *Drosophila* EVs contain similar amounts of proteins and RNA
Quantification of total protein (left panel) and RNA (right panel) in EV preparations derived from human and *Drosophila* cells. No significant difference was found across samples by an ANOVA test.

Figure 7.S3. Average protein and RNA content per individual EV Estimations of total protein (top panel) and RNA (bottom panel) mass found in individual EV were obtained from

results of photometric macromolecule quantification and NTA vesicle counts. The significance of the differences observed between lines of the same species was evaluated through a Mann-Whitney test while the prevalence of a global difference was evaluated using a Kruskal-Wallis test. One femtogram (fg) corresponds to 10^{-15} g, while one attogram (ag) corresponds to 10^{-18} g.

Figure 7.S4. The bulk of RNA found in human and *Drosophila* EVs is protected from degradation by RNAse A

EV preparations derived from human HepG2 and A431 and *Drosophila* D17 and S2R+ cells were split in two equal pools. The first pool (grey bars) was incubated with RNAse A followed by inactivation and total RNA extraction, while the second pool (black bars) was directly submitted to RNA extraction. Previously quantified samples of cellular HepG2 and D17 total RNA were used as controls.

Figure 7.S5. RNA in human and *Drosophila* EVs largely consists of short sequences

Length distributions of human and *Drosophila* EV (left panels) and cellular (right panels) total RNA were evaluated by bioanalyzer capillary electrophoresis. 'R' refers to the 18S and 28S rRNA peaks, while 'M' refers to the exogenous marker peak.

Figure 7.S6. Bidirectional read distributions map to 5S pseudogenes in human HepG2 EVs

Strand-specific read coverage of the genomic region surrounding the RNA5SP145 (chr3:150,905,886-150,906,010, top panel) and RNA5SP149 (chr3:179,879,674-179,879,765; bottom panel) in human HepG2 EVs and cells. The y-axis refers to read coverage (no normalization). Screen shots were taken from the UCSC genome browser.

Figure 7.S7. Mitochondrial rRNAs display abundant bidirectional read coverage in HepG2 EVs and cells

Strand-specific read coverage of the genomic region surrounding the 12S (MT-RNR1; chrM:650-1,603) and 16S (MT-RNR2; chrM:1,673-3,230) loci in human HepG2 EVs and cells. Note that rRNA depletion was performed prior to RNA-seq in HepG2 cellular extracts.

Figure 7.S8. Abundant bidirectional signatures of an rRNA-like novel miRNA in human HepG2 EVs

Strand-specific read coverage of the genomic region surrounding the *AC079949*.1 "novel miRNA" locus in human HepG2 EVs and cells. This miRNA overlaps with a large subunit rRNA repeat (chr12:127,650,453-127,650,987), H3K27 acetylated and DNAse I sensitive region.

Figure 7.S9. Abundant full-length read distributions map to vault RNA 1-1 in HepG2 EVs

Strand-specific read coverage surrounding the *VTRNA1-1* locus (chr5:140,090,861-140,090,958) in HepG2 EVs and cells.

Figure 7.S10. Spliceosomal U5 snRNAs are abundant in human and *Drosophila* EVs

Strand-specific read coverage of a *Drosophila* U5 snRNA gene (U5:63BC; chr3L:3,090,801-3,090,923) and a human ortholog (RNU5B-1; chr15:65,597,015-65,597,130) in D17 (A) and HepG2 (B) EVs and cells.

Figure 7.S11. Intronic snoRNAs display differential targeting to *Drosophila* EVs

Strand-specific read coverage of three snoRNAs encoded from intronic regions of CG13900 (chr3L:826,720-833,385) in *Drosophila* EVs and cells.

Figure 7.S12. Evidence of tRNA targeting to *Drosophila* and human EVs

Strand-specific read coverage of two proximal *Drosophila* tRNAs (chr2L:8,683,381-8,683,935; top panel) and two human tRNAs (chr12:125,423,890-125,425,085; bottom panel). The latter example shows evidence of bidirectional tRNA sequences in EVs and asymmetric read coverage in EVs and cells.

Figure 7.S13. A narrow read distribution maps to a central region of the retrotransposon TART in *Drosophila* EVs and cells

Strand-specific read coverage of the TART locus (chrX:15,069-25,164) in Drosophila EVs and cells.

Figure 7.S14. A narrow antisense read distribution maps to a central region of the retrotransposon Copia in *Drosophila* EVs

Strand-specific read coverage surrounding a Copia element repeat (chr2L:2,294,373-2,298,967) in *Drosophila* S2R+ EVs, D17 EVs and D17 cells.

Figure 7.S15. Correlative analysis of orthologous protein-coding genes

Relative TPM abundance values of Drosophila D17 mRNAs were correlated to their corresponding human HepG2 orthologs. Cellular (top panel) and EV (bottom panel) populations are displayed as scatter plots. Pearson's correlations (r) and associated p-values are indicated at the top of each graph.

Table 7.S1. Read distributions of the fifty most abundant rRNA genes in human and *Drosophila* EV samples

Table7.S2. List of the ten most abundant transcripts in human and *Drosophila* EV samples following rRNA subtraction

Table 7.S3. List of the ten most abundant repeat element-derived RNAs in human and *Drosophila* EVs

Table 7.S4. List of the ten most abundant mRNAs in human and *Drosophila* EVs

Table 7.S5. List of the ten most significantly enriched gene ontology (GO) terms associated with human and *Drosophila* EV mRNAs.

Table 7.S6. Exhaustive list of "EV-exclusive" mRNAs in human HepG2 and *Drosophila* D17 cells.

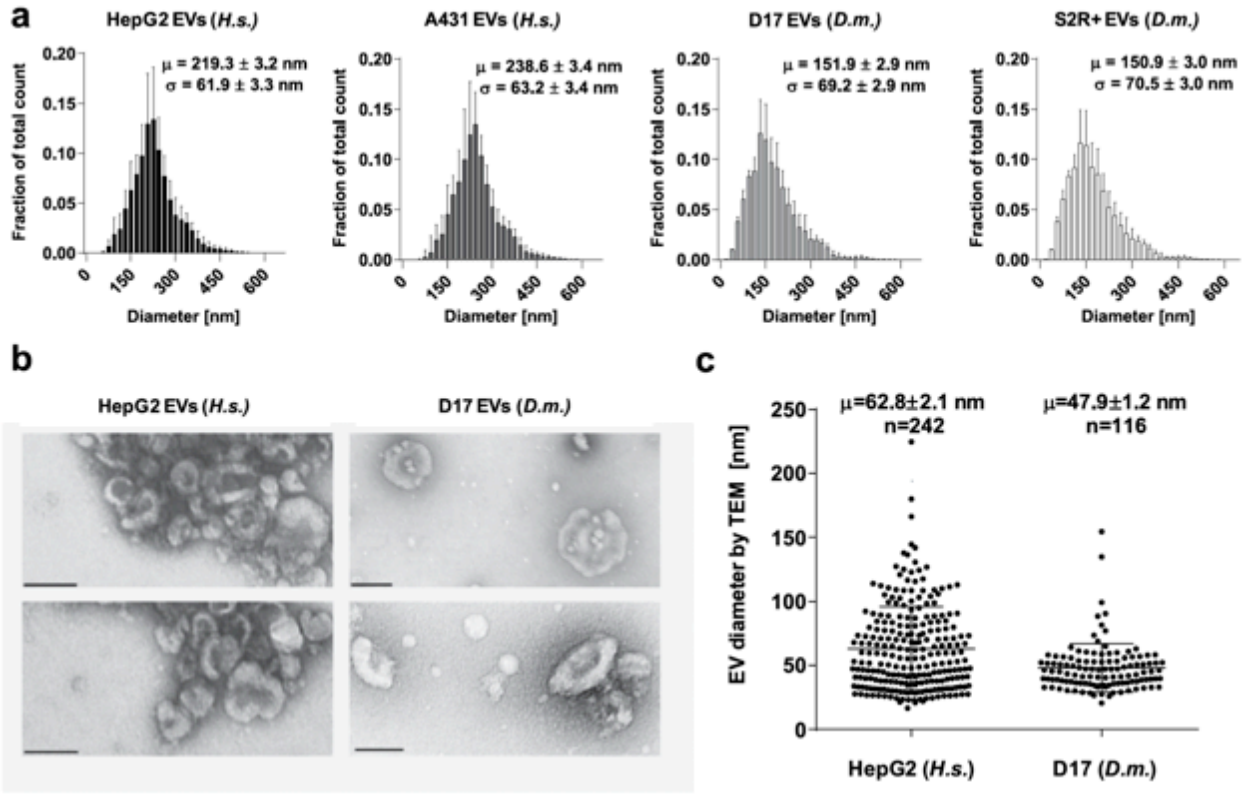


Figure 7. 1. **Size distributions of EVs released by human and *Drosophila* cells**

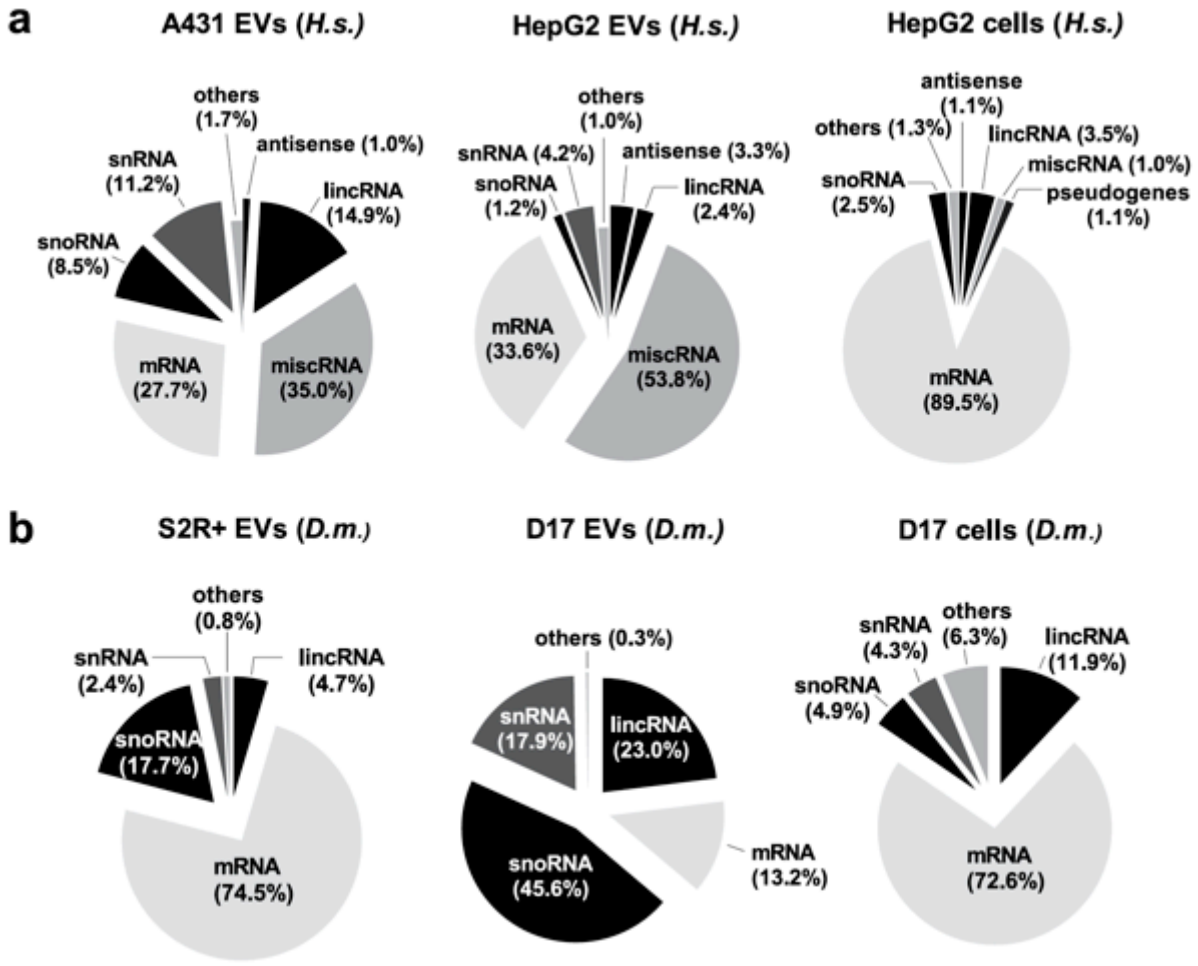


Figure 7. 2. Human and *Drosophila* EVs enclose various types of transcripts

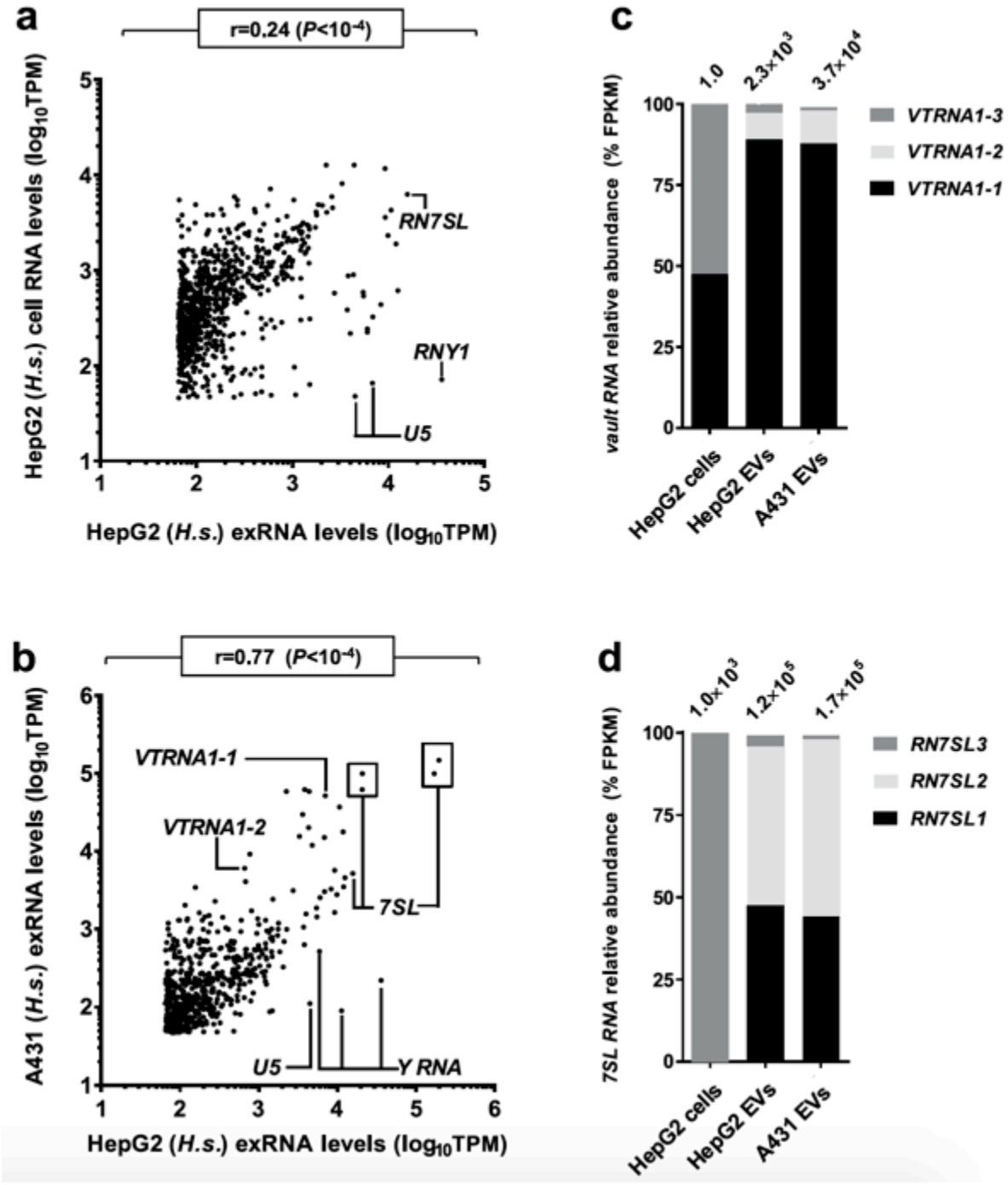


Figure 7. 3. Correlative comparisons of exRNA secreted by human cell lines

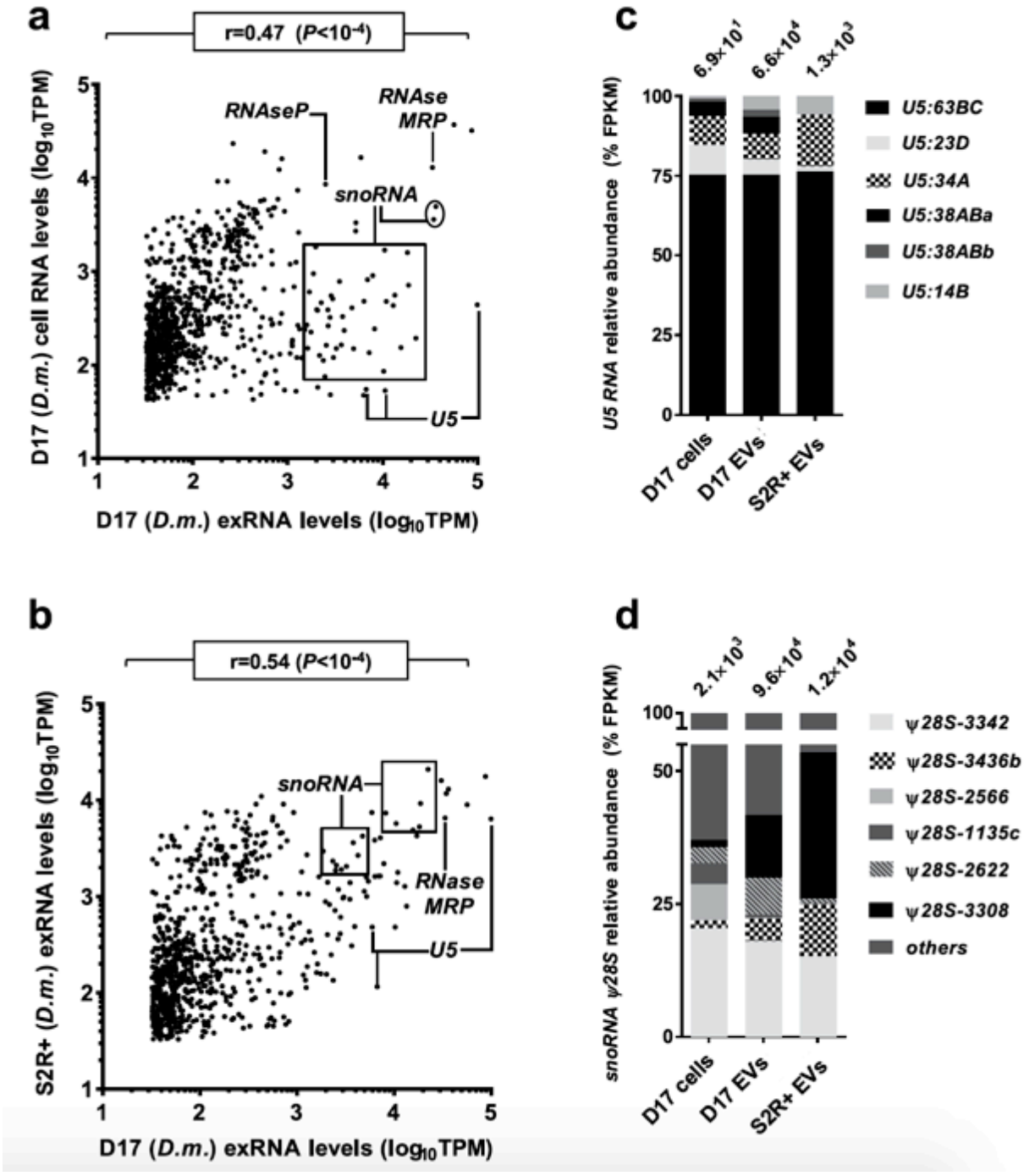


Figure 7. 4. Correlative comparisons of exRNA secreted by *Drosophila* cell lines

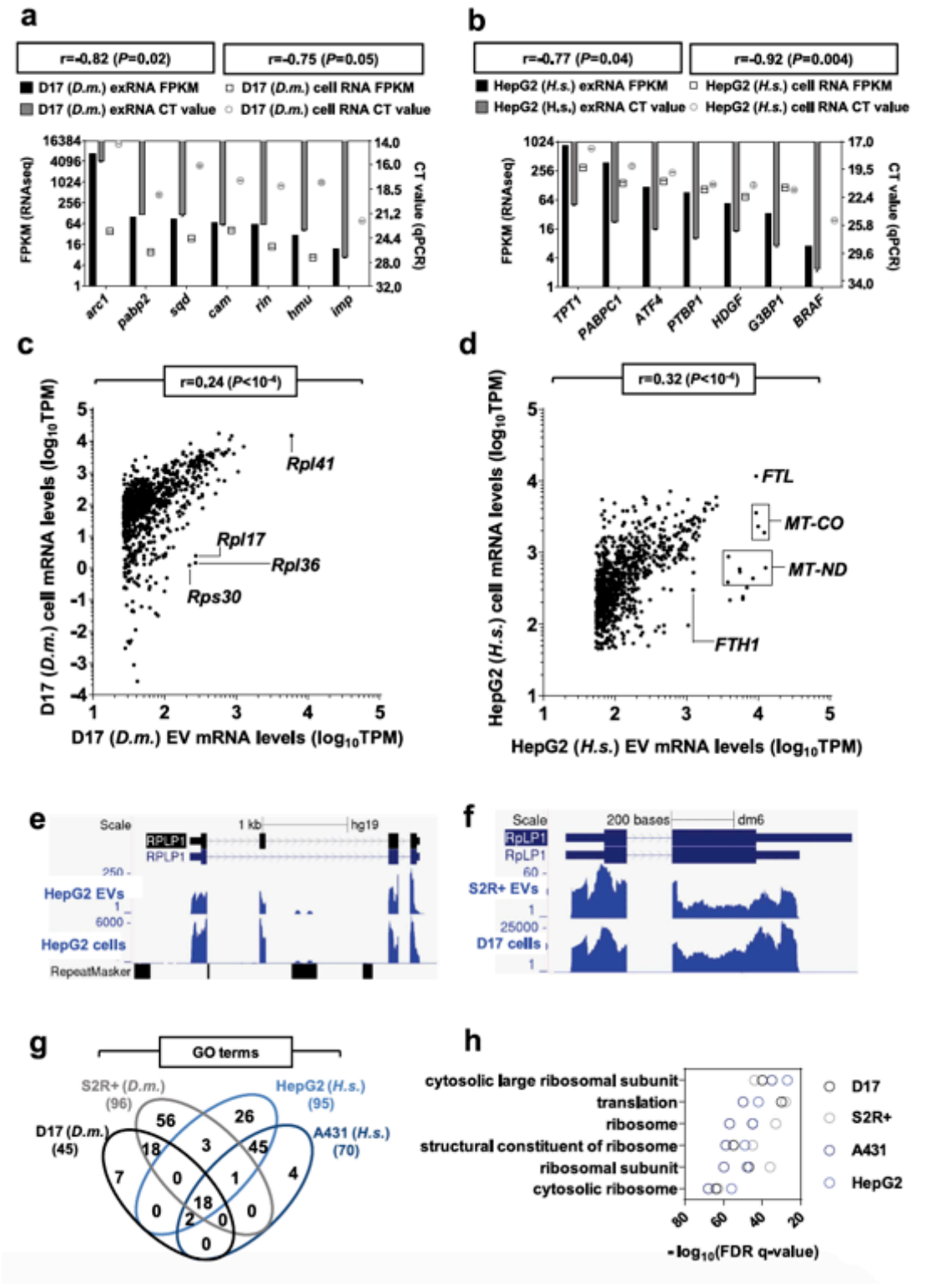


Figure 7. 5. Characterization of mRNAs secreted within Human and *Drosophila* EVs

Table 1. Read metrics of human and *Drosophila* rRNA sequences

Library	Mapped reads ($\times 10^6$)	rRNA reads ($\times 10^6$)	rRNA reads (%)
S2R+ exRNA (i)	15.13	12.84	84.91
S2R+ exRNA (ii)	19.29	17.74	91.98
D17 exRNA (i)	16.7	15.03	90.01
D17 exRNA (ii)	15.09	14.24	94.41
D17 cell RNA (i)	40.89	0.83	2.05
D17 cell RNA (ii)	38.01	0.86	2.27
A431 exRNA (i)	8.10	7.96	98.39
A431 exRNA (ii)	5.42	5.16	95.29
HepG2 exRNA (i)	7.06	6.45	91.38
HepG2 exRNA (ii)	6.97	5.55	79.73
HepG2 cell RNA (i)	71.51	0.49	0.69
HepG2 cell RNA (ii)	72.66	0.63	0.88

Table 2. Read metrics of human and *Drosophila* repeated elements and other transcripts

Library	Reads left upon rRNA filtration ($\times 10^6$)	Repeats reads ($\times 10^6$)	Repeats reads (%)	Other RNA reads ($\times 10^3$)	# Other RNA ≥ 5 FPKM
S2R+ exRNA (i)	2.29	1.49	65.07	250	3,905
S2R+ exRNA (ii)	1.55	1.09	70.32	180	
D17 exRNA (i)	1.67	0.83	49.70	81	4,472
D17 exRNA (ii)	0.85	0.74	87.06	51	
D17 cell RNA (i)	40.06	12.99	32.43	27,200	3,944
D17 cell RNA (ii)	37.15	11.83	31.84	25,300	
A431 exRNA (i)	0.14	0.02	16.43	110	9,969
A431 exRNA (ii)	0.26	0.05	18.85	270	
HepG2 exRNA (i)	0.61	0.10	16.39	850	4,754
HepG2 exRNA (ii)	1.42	0.24	16.90	2220	
HepG2 cell RNA (i)	71.02	4.92	6.93	68,850	6,537
HepG2 cell RNA (ii)	72.03	5.09	7.07	69,850	

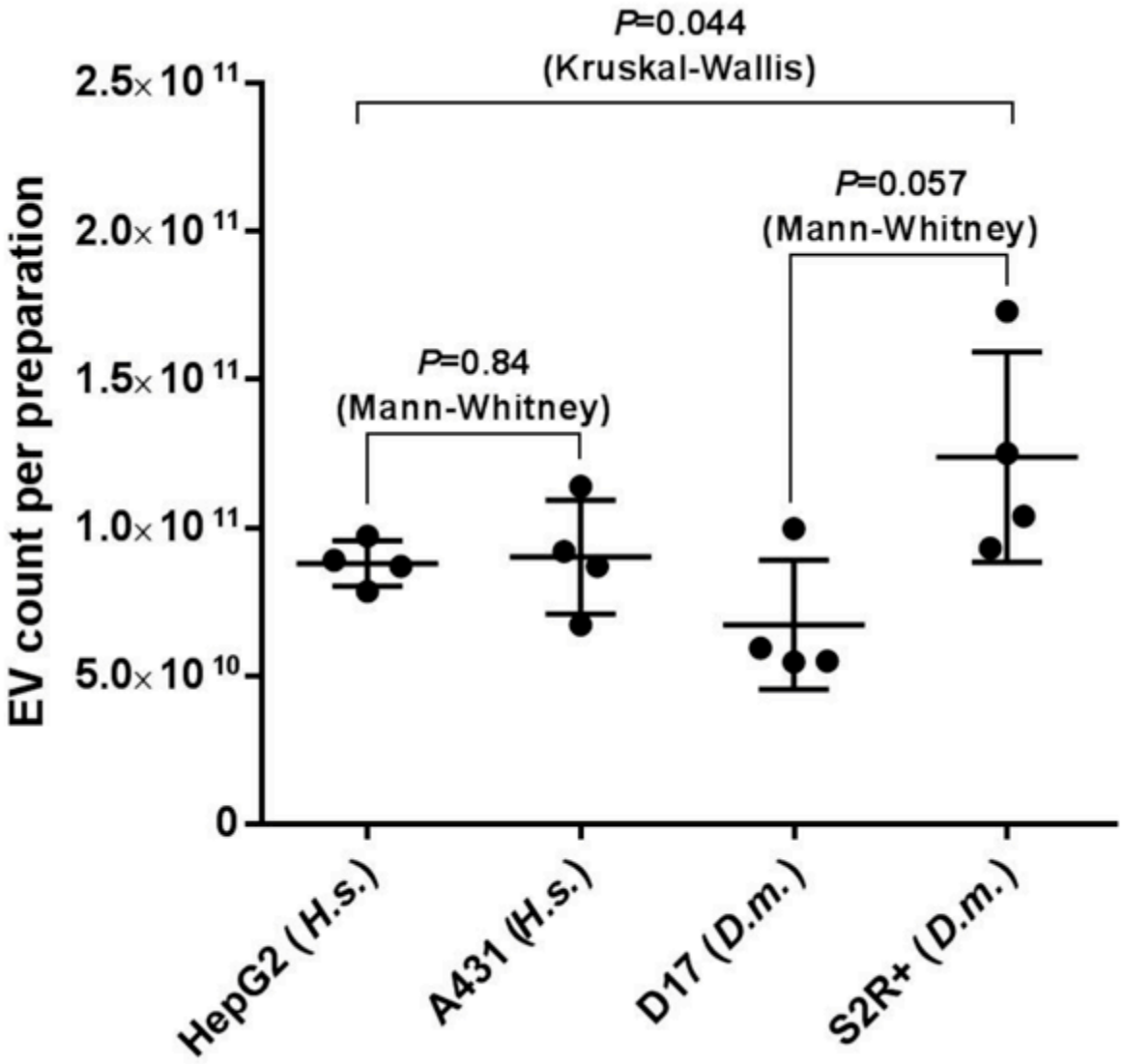


Figure 7S. 1. **EV count per preparation**

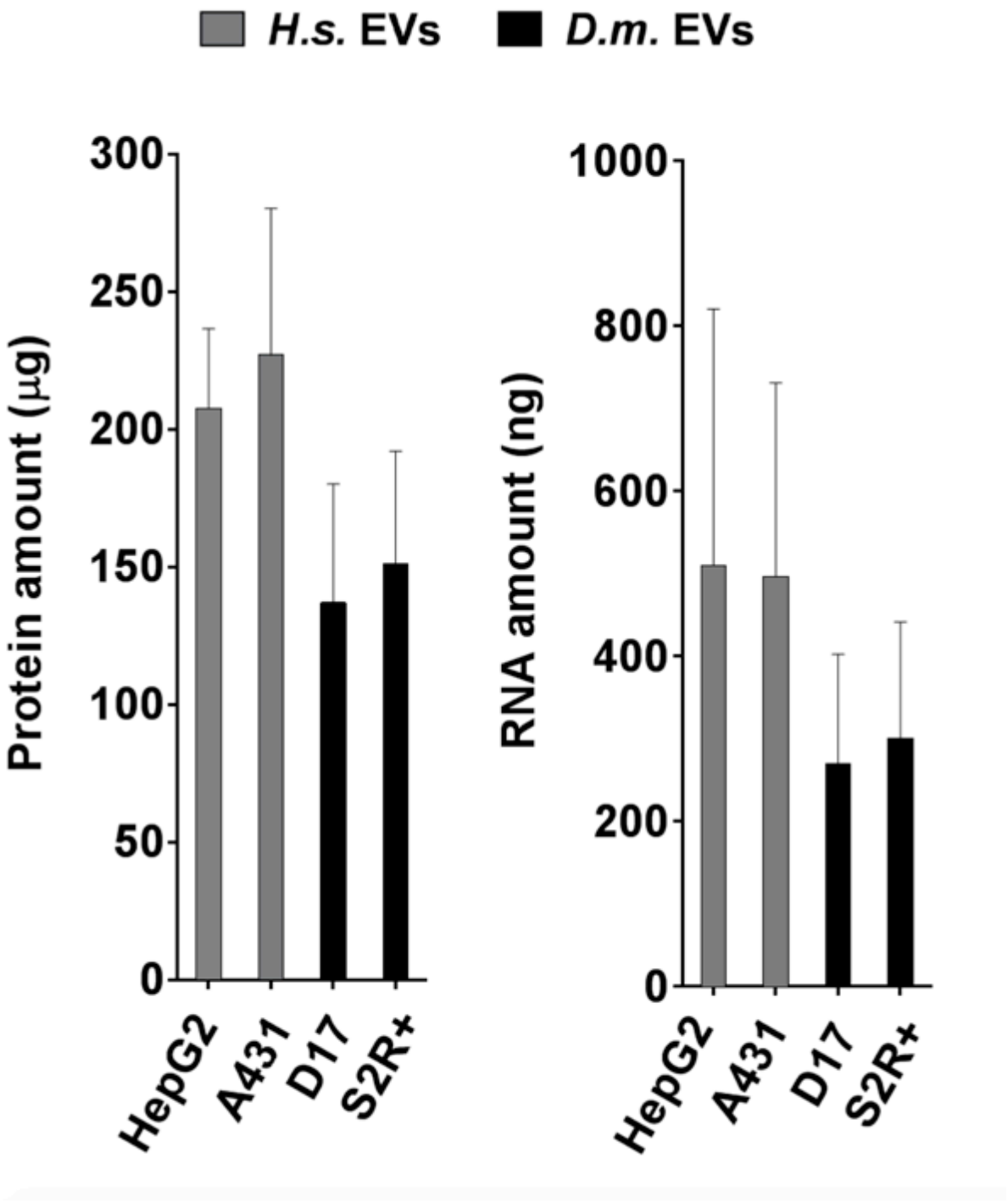


Figure 7S. 2. **Human and *Drosophila* EVs contain similar amounts of proteins and RNA**

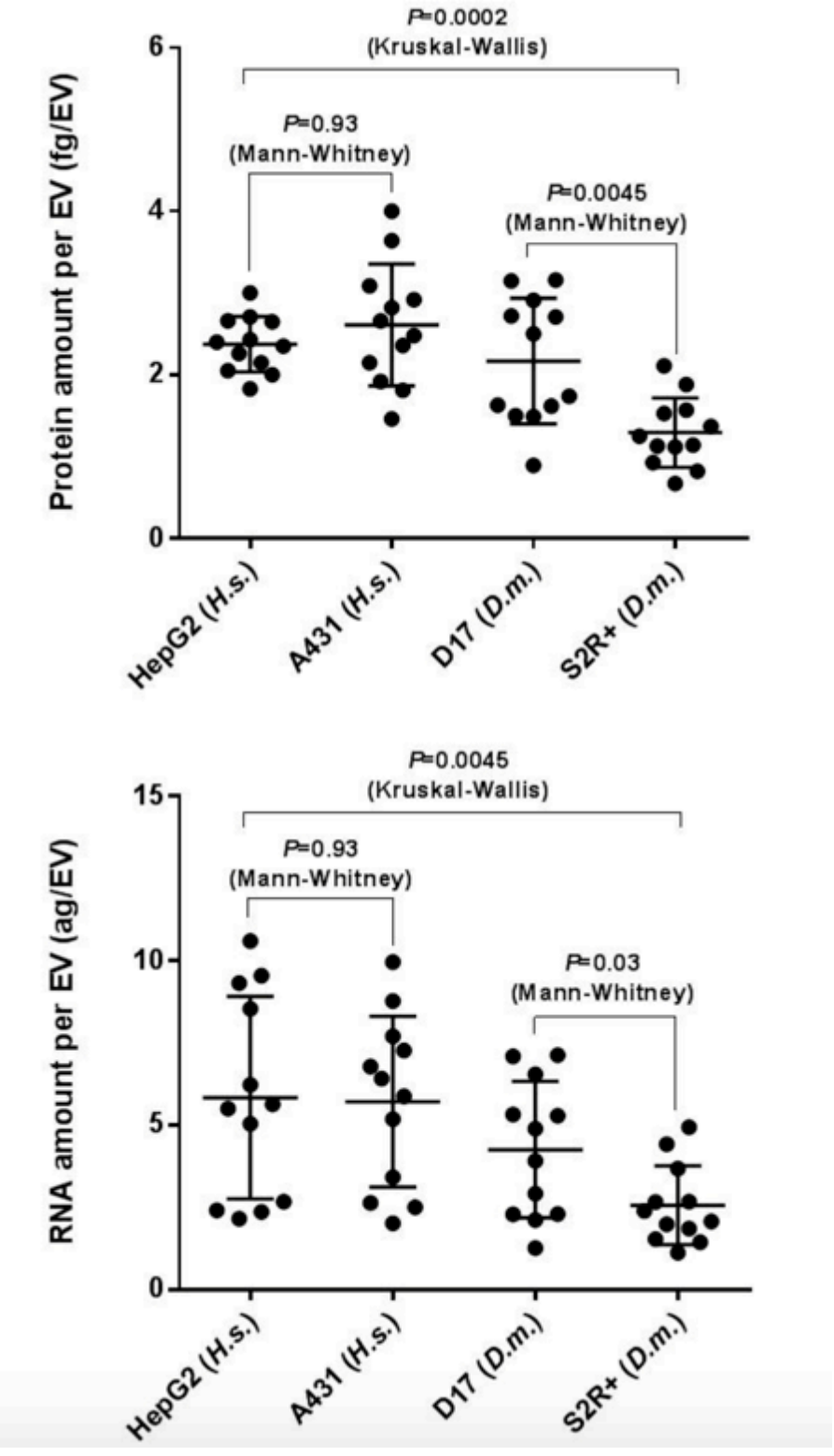


Figure 7S. 3. **Average protein and RNA content per individual EV**

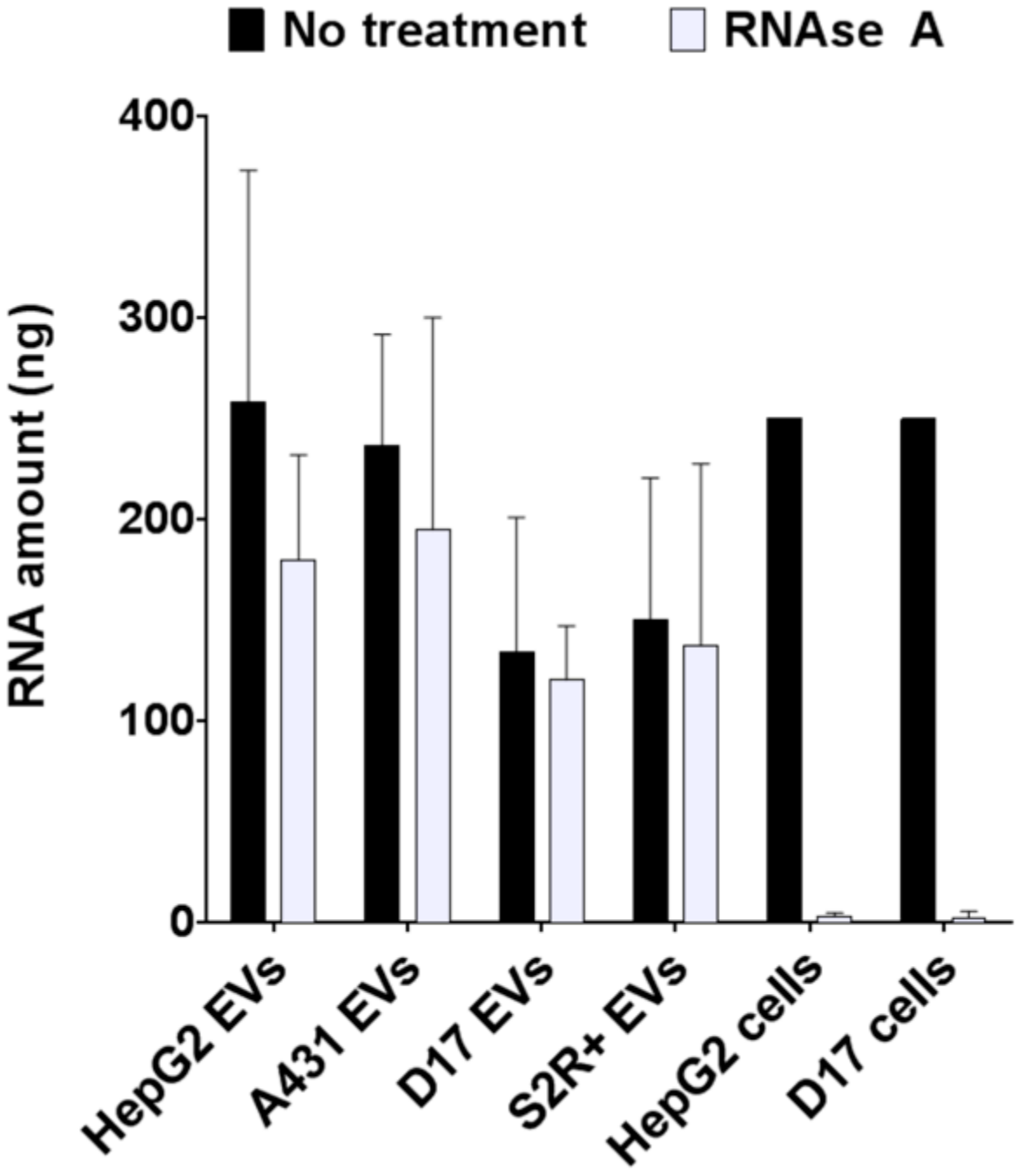


Figure 7S. 4. **The bulk of RNA found in human and *Drosophila* EVs is protected from degradation by RNAse A**

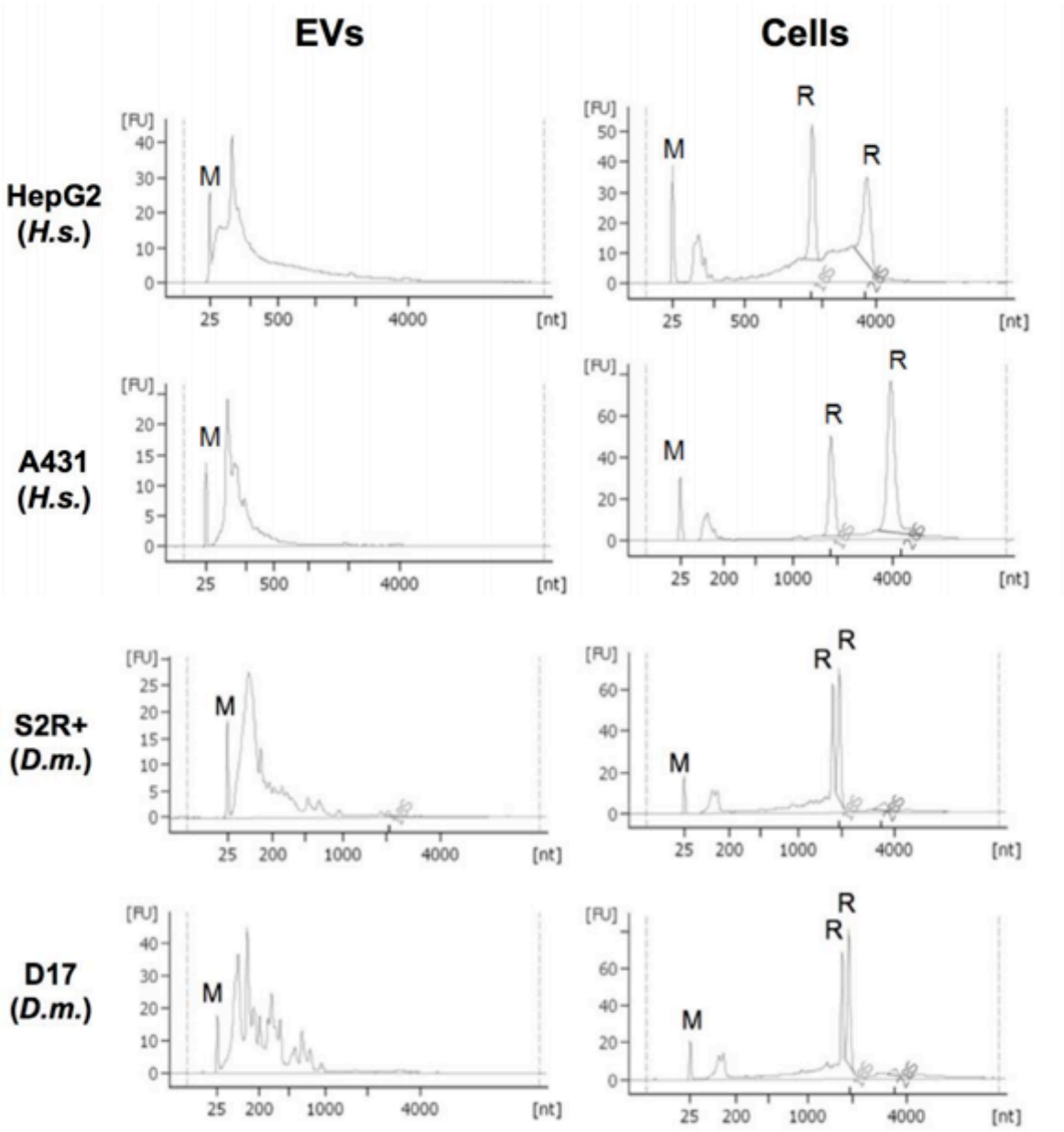


Figure 7S. 5. **RNA in human and *Drosophila* EVs largely consists of short sequences**

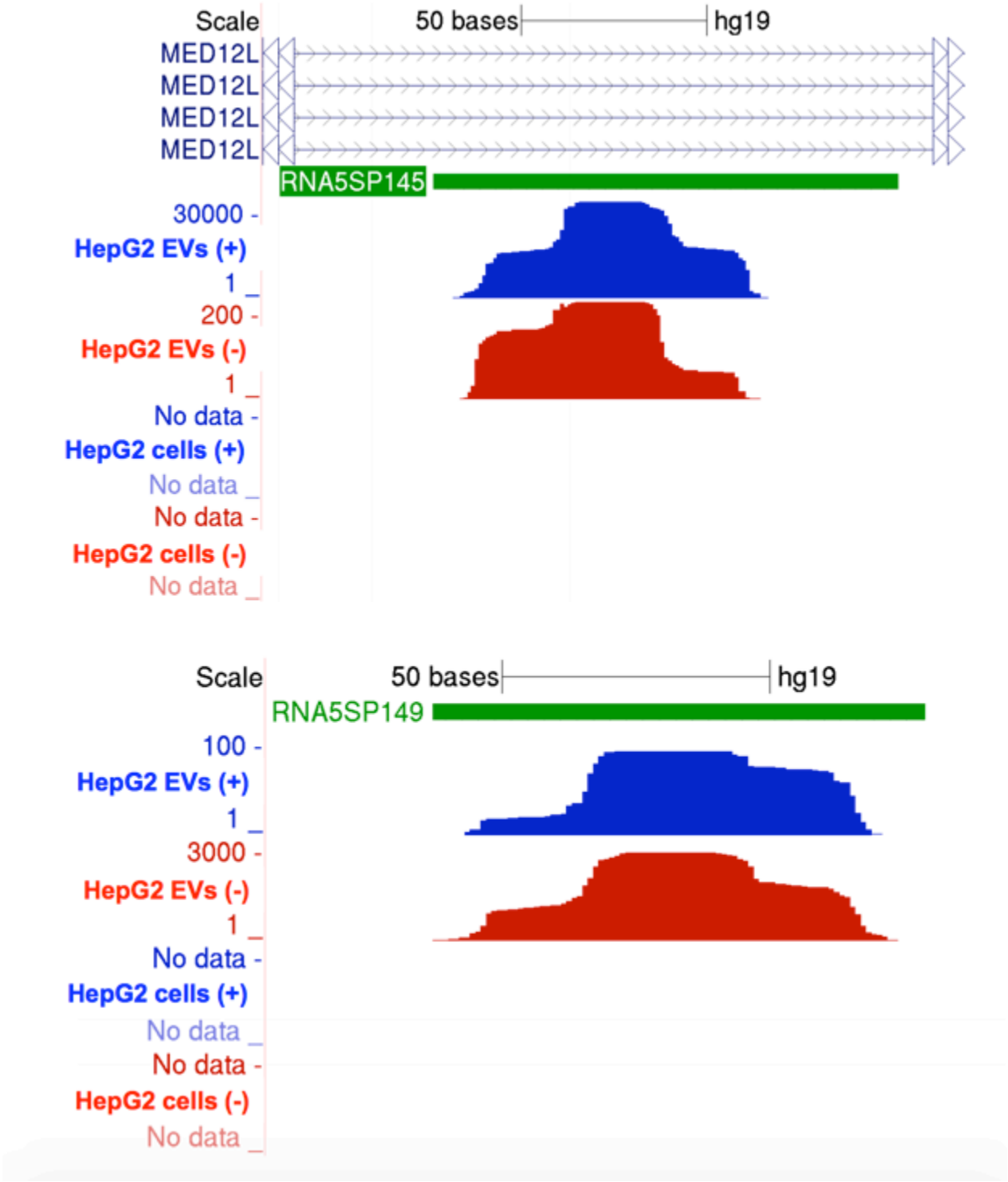


Figure 7S. 6. Bidirectional read distributions map to 5S pseudogenes in human HepG2 EVs

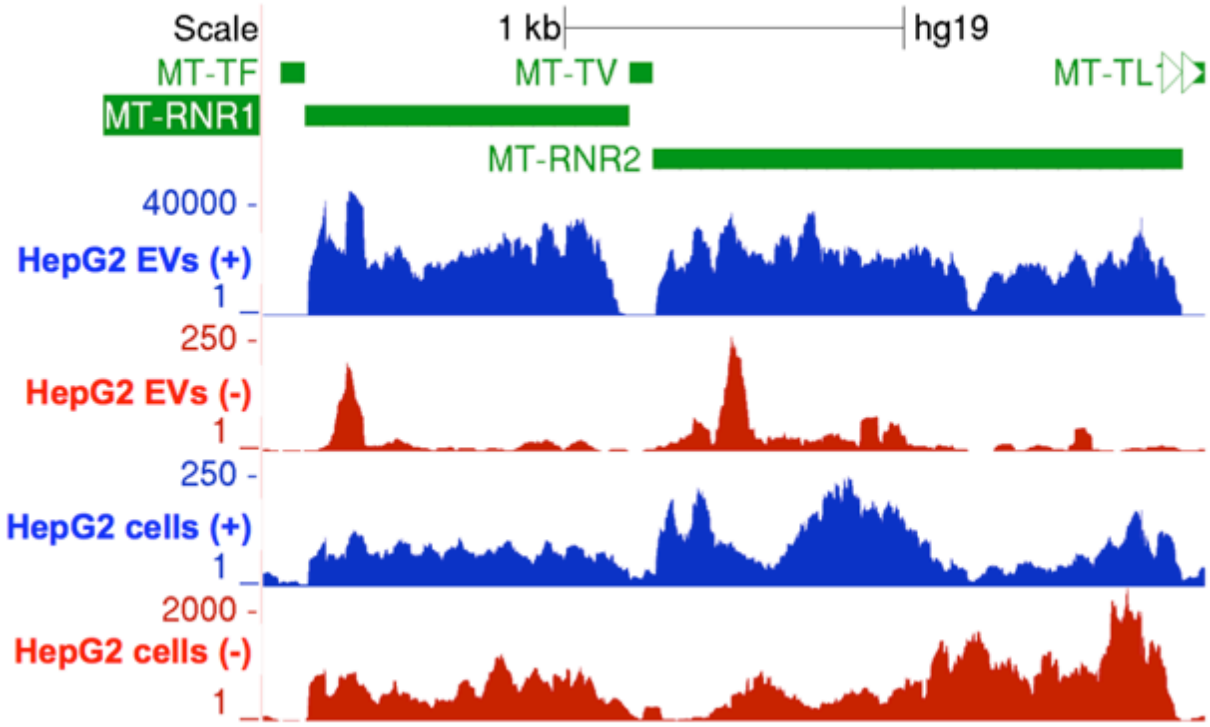


Figure 7S. 7. Mitochondrial rRNAs display abundant bidirectional read coverage in HepG2 EVs and cells

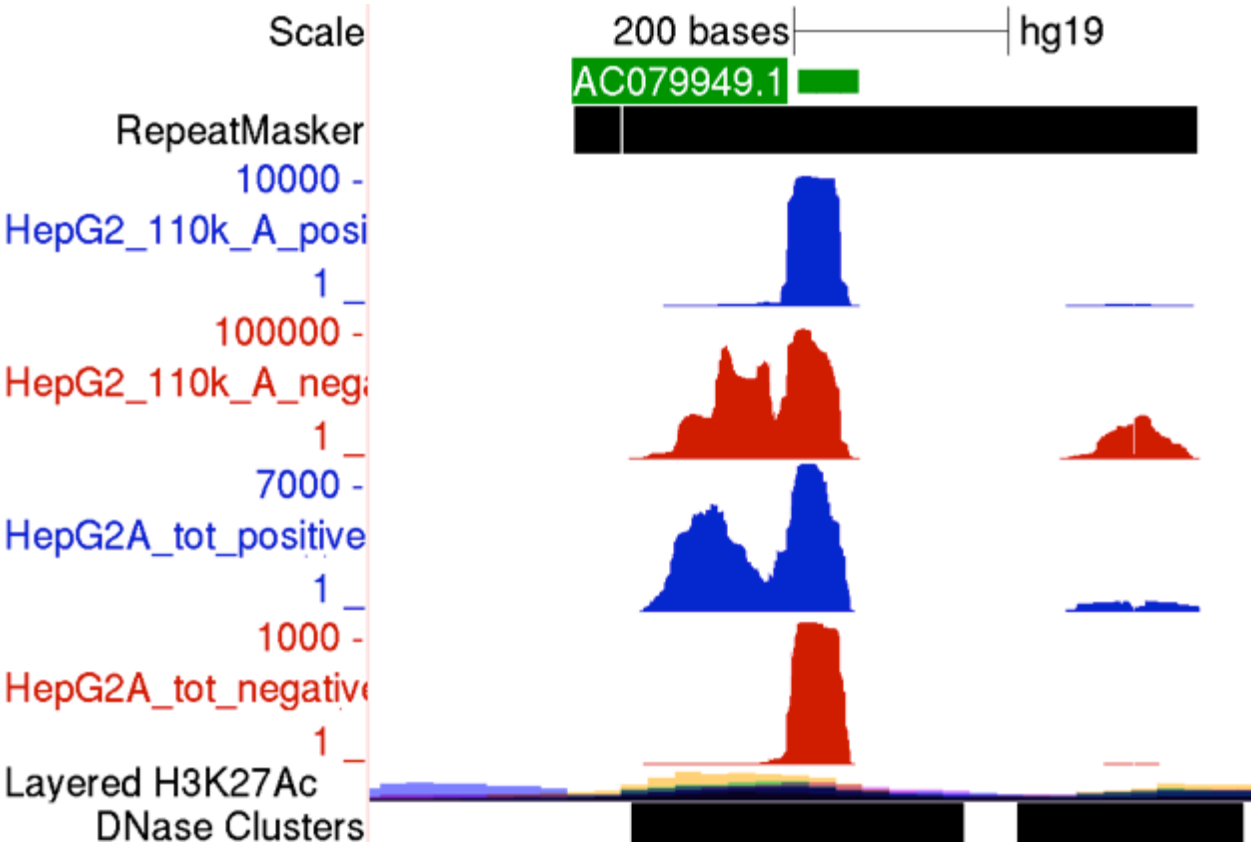


Figure 7S. 8. Abundant bidirectional signatures of an rRNA-like novel miRNA in human HepG2 EVs

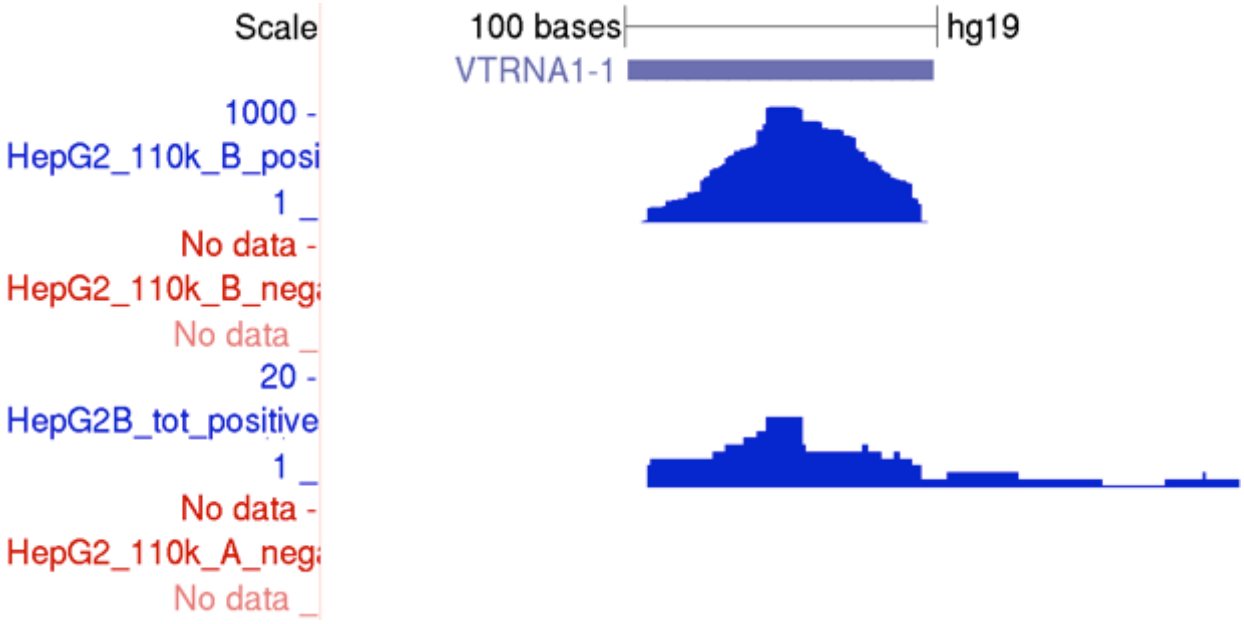


Figure 7S. 9. Abundant full-length read distributions map to vault RNA 1-1 in HepG2 EVs

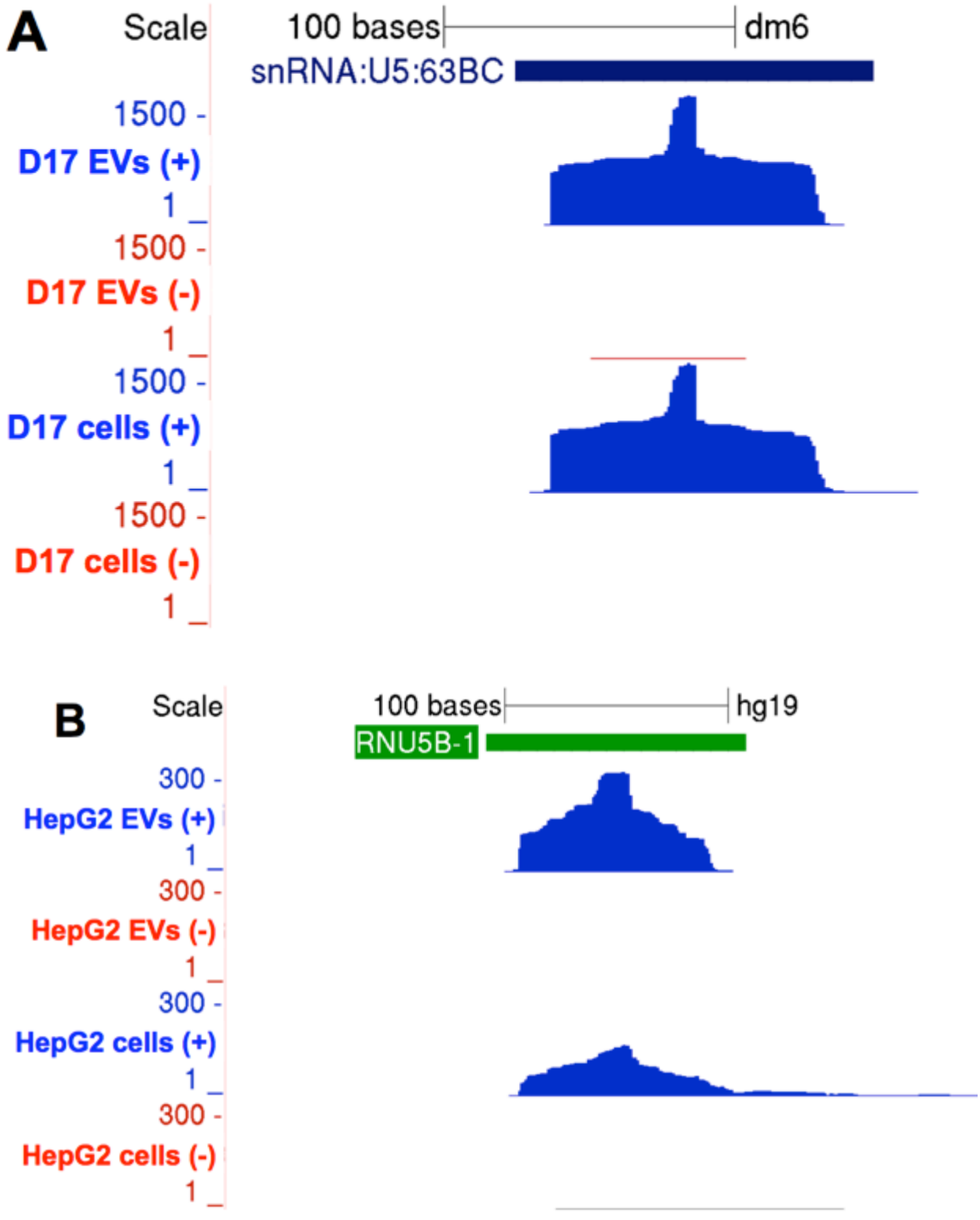


Figure 7S. 10. **Spliceosomal U5 snRNAs are abundant in human and *Drosophila* EVs**

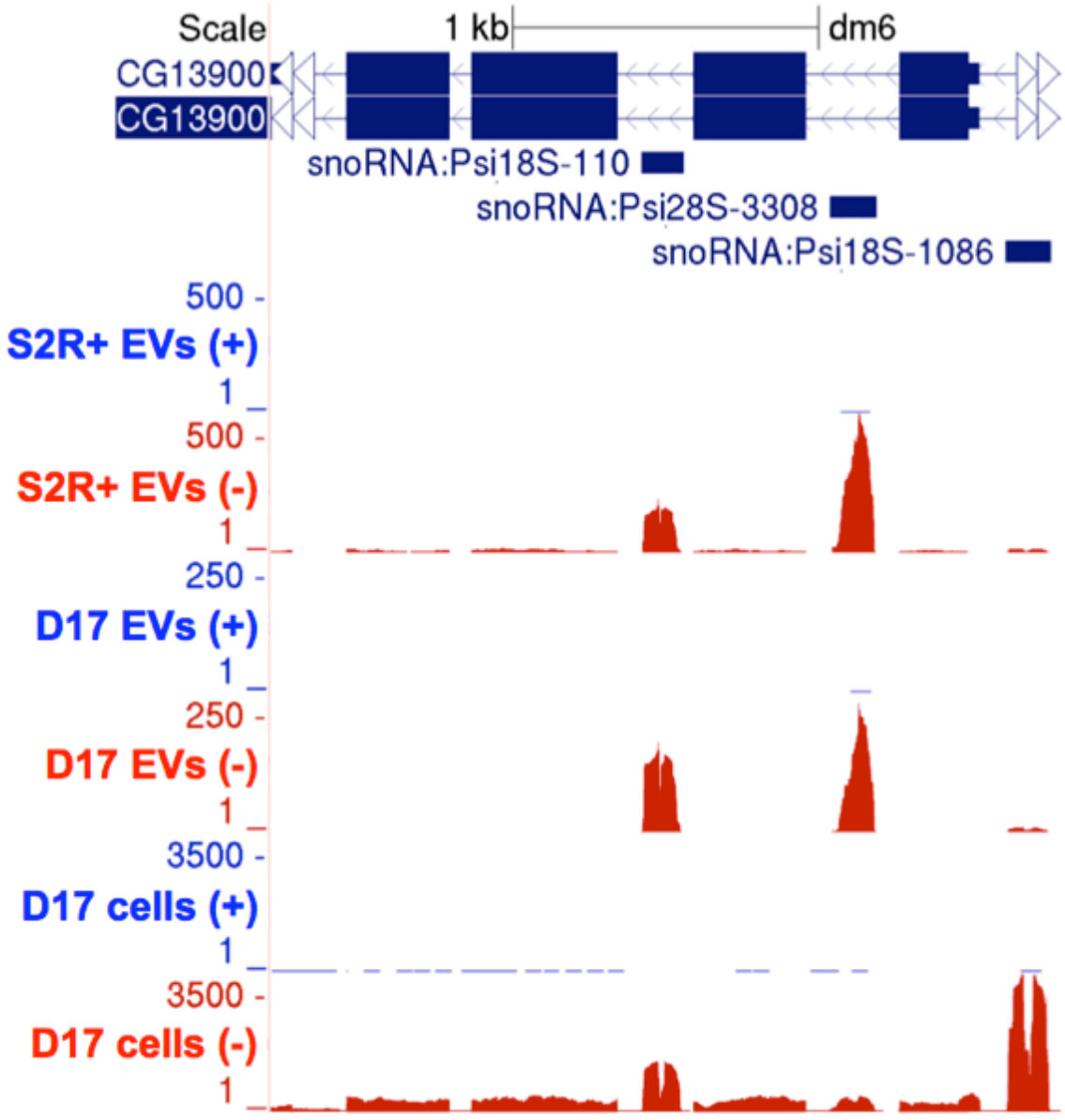


Figure 7S. 11. Intronic snoRNAs display differential targeting to *Drosophila* EVs

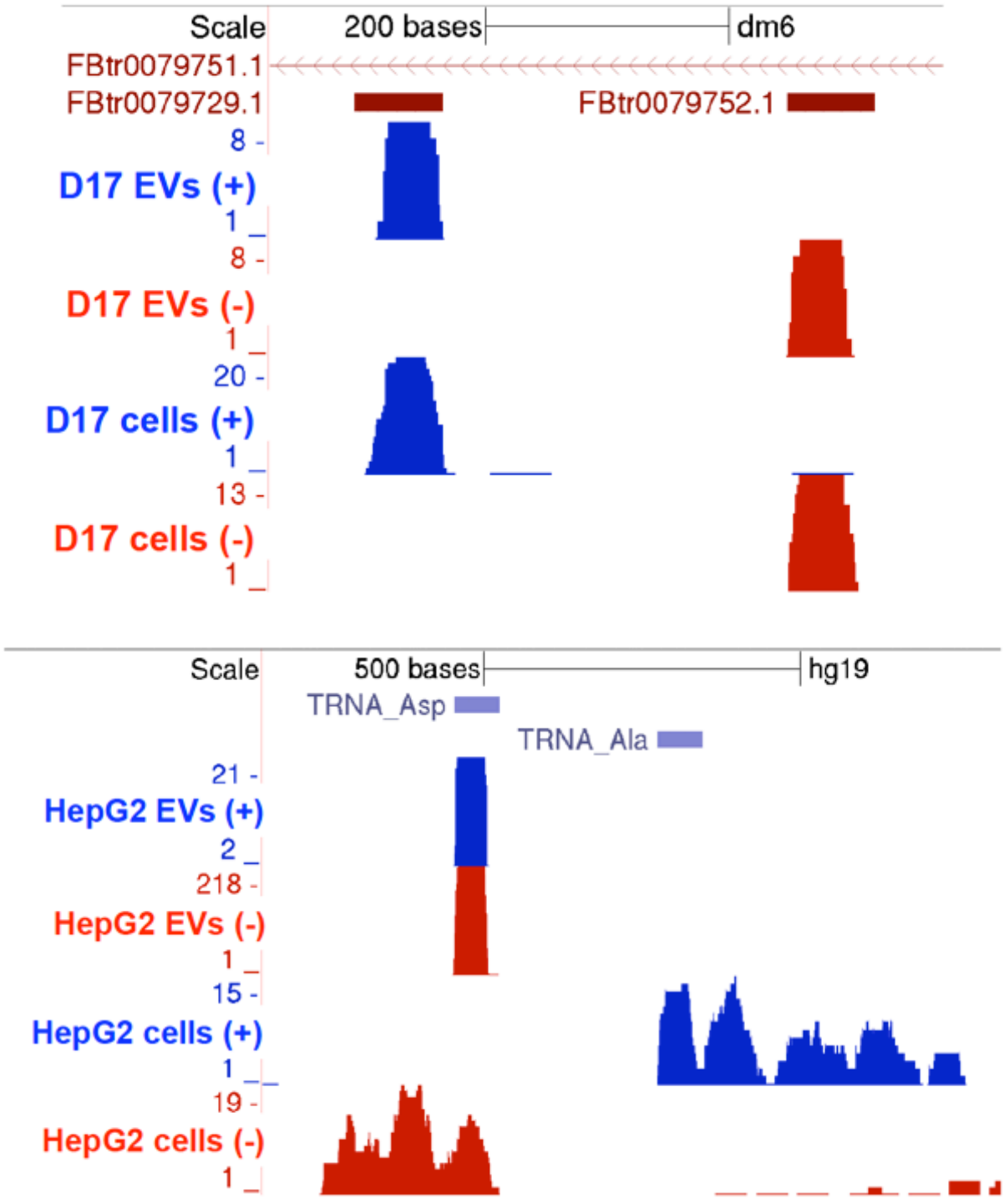


Figure 7S. 12. **Evidence of tRNA targeting to *Drosophila* and human EVs**

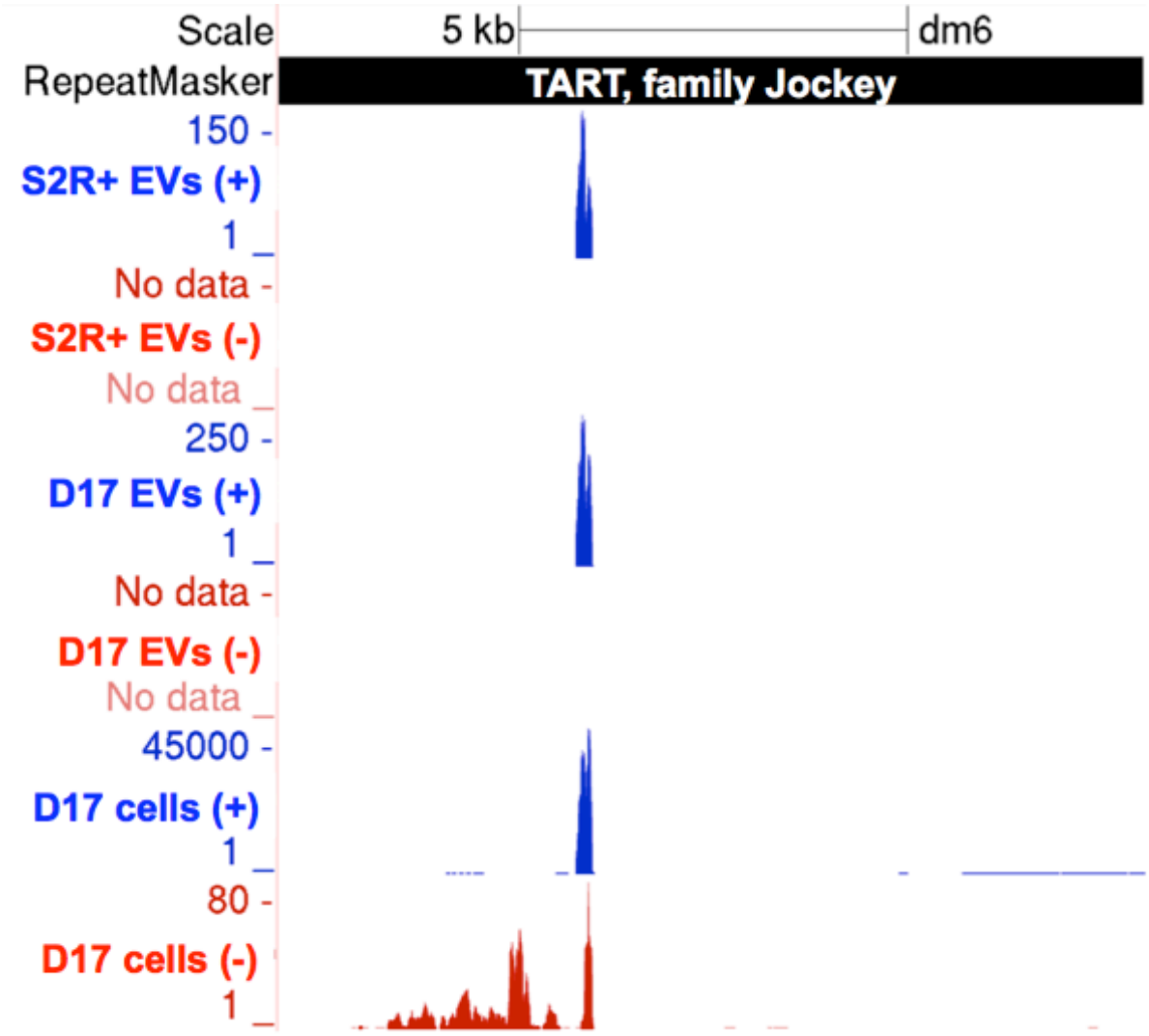


Figure 7S. 13. A narrow read distribution maps to a central region of the retrotransposon TART in *Drosophila* EVs and cells

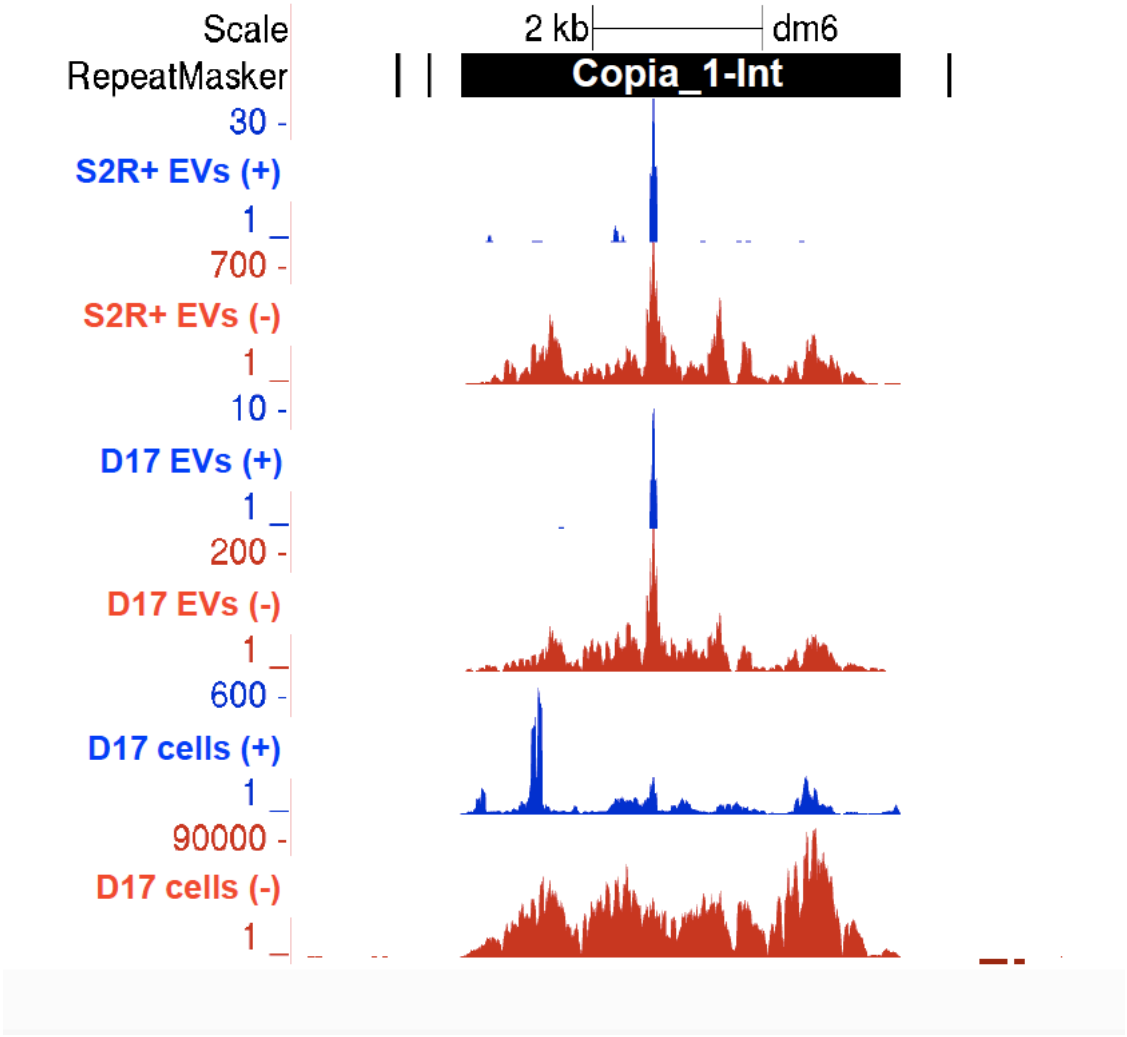
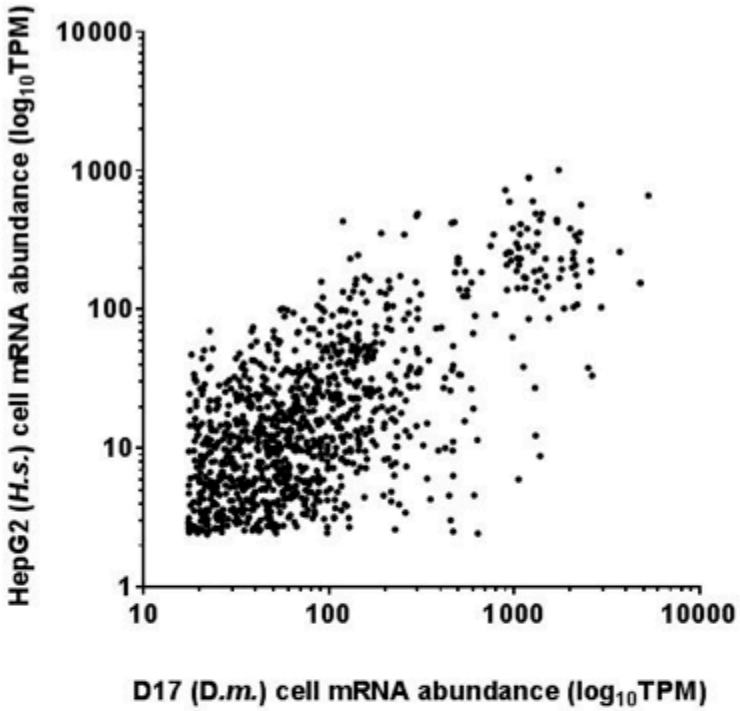


Figure 7S. 14. A narrow antisense read distribution maps to a central region of the retrotransposon Copia in *Drosophila* EVs



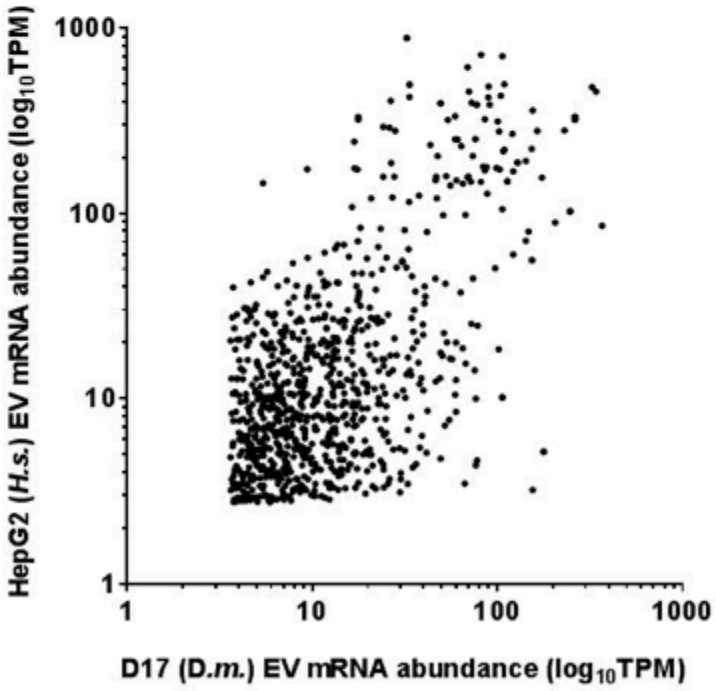


Figure 7S. 15. **Correlative analysis of orthologous protein-coding genes**

Table 6S. I. Read distributions of the fifty most abundant rRNA genes in human and *Drosophila* EV samples

ensembl_gene_id	gene_id	A431 Evs	HepG2 Evs	HepG2 cells
ENSG00000266658	RNA28S5	18320021	16414483	853851
ENSG00000272060	RNA18S5	15308624	10290624	285306
ENSG00000210082	MT-RNR2	81302	396901	26866
ENSG00000211459	MT-RNR1	29807	276904	12367
ENSG00000241335	RNA5-8S5	77325	66110	138
ENSG00000242716	RNA5-8S5	77305	66095	191
ENSG00000201925	RNA5S15	10704	23295	4
ENSG00000200624	RNA5S6	10313	23071	1
ENSG00000199910	RNA5S10	10265	22870	1
ENSG00000199352	RNA5S1	10554	23056	1
ENSG00000199337	RNA5S3	10362	22704	0
ENSG00000202257	RNA5S16	10289	22646	1
ENSG00000199334	RNA5S11	10052	22313	0
ENSG00000199270	RNA5S12	10487	22525	2
ENSG00000201588	RNA5S2	10323	22291	5
ENSG00000202526	RNA5S13	10111	22266	4
ENSG00000200343	RNA5S8	10221	22204	2
ENSG00000202521	RNA5S7	10236	22213	3
ENSG00000201355	RNA5S14	10259	22273	0
ENSG00000199396	RNA5S5	9988	21973	2
ENSG00000200381	RNA5S4	10216	21964	4
ENSG00000200370	RNA5S17	10105	21766	1
ENSG00000201321	RNA5S9	3951	5204	0
ENSG00000199415	RNA5SP370	2059	4456	0
ENSG00000199523	RNA5SP226	2045	4254	0
ENSG00000199480	RNA5SP357	1910	4096	0
ENSG00000199654	RNA5SP389	1917	4070	0
ENSG00000201185	RNA5SP202	2002	3199	0
ENSG00000199994	RNA5SP145	630	1380	0
ENSG00000201822	RNA5SP149	262	475	0
ENSG00000200741	RNA5SP161	332	344	0
ENSG00000252623	RNA5SP481	165	278	0
ENSG00000201861	RNA5SP298	132	255	0
ENSG00000222806	RNA5SP225	46	91	0
ENSG00000266658	RNA28S5	624	125	3
ENSG00000201413	RNA5SP141	18	45	0
ENSG00000200558	RNA5SP429	25	45	0
ENSG00000202411	RNA5SP259	21	34	0
ENSG00000199564	RNA5SP502	10	26	0
ENSG00000200278	RNA5SP352	16	25	0
ENSG00000251705	RNA5-8SP6	62	35	0
ENSG00000272351	5S_rRNA	10	9	0
ENSG00000199839	RNA5SP150	6	8	0
ENSG00000200275	RNA5SP199	7	8	0
ENSG00000201041	RNA5SP242	4	5	0
ENSG00000201096	RNA5SP387	2	4	0
ENSG00000200408	RNA5SP74	1	3	0
ENSG00000200687	RNA5SP335	7	4	0
ENSG00000202187	RNA5SP355	5	5	0
ENSG00000200115	RNA5SP440	0	2	0

Flybase ID	Gene name	S2R+ Evs	D17 Evs	D17 cell
FBgn0085797	CR41540	5306258	3591189	295662
FBgn0085813	CR41602	3517800	3380805	118729
FBgn0085742	CR40546	2429095	1567232	186204
FBgn0085774	CR40779	2260038	2136917	16865
FBgn0085805	CR41583	2106021	2505405	137649
FBgn0085823	CR41617	2092710	3075441	2922
FBgn0085819	CR41609	1596444	930954	134872
FBgn0085773	CR40766	1407175	726834	26546
FBgn0085825	CR41619	1344789	763627	4271
FBgn0085807	CR41590	868201	1204822	12725
FBgn0085765	CR40677	780867	816985	77306
FBgn0085815	CR41605	746258	1407179	6373
FBgn0085757	CR40621	588246	818844	10962
FBgn0085743	CR40560	560630	423825	27134
FBgn0085769	CR40728	472500	405368	1632
FBgn0085817	CR41607	423919	638905	3677
FBgn0085764	CR40668	405530	545564	148161
FBgn0085796	CR41539	379780	396090	4901
FBgn0085822	CR41613	376405	400427	3161
FBgn0085795	CR41535	375645	318151	62542
FBgn0085761	CR40642	372968	395060	3086
FBgn0085759	CR40640	287741	269986	2839
FBgn0085770	CR40734	263840	335867	56804
FBgn0085760	CR40641	255087	304084	2777
FBgn0085828	CR42195	231387	328168	3164
FBgn0085777	CR40959	230728	256935	937
FBgn0085799	CR41544	194809	250138	2754
FBgn0085771	CR40741	190510	545404	115922
FBgn0085766	CR40679	164617	202766	11915
FBgn0085753	CR40596	160787	144314	56757
FBgn0085758	CR40639	118711	82604	3938
FBgn0085802	CR41548	100029	100012	13876
FBgn0085737	CR40502	39134	87148	402
FBgn0085767	CR40685	23564	36117	22197
FBgn0085772	CR40743	23503	35637	1074
FBgn0250731	5.8SrRNA:CR40454	23359	35803	44
FBgn0085826	CR41620	23276	35703	388
FBgn0085827	CR41621	5352	9367	7148
FBgn0053394	5SrRNA:CR33394	2638	3443	60
FBgn0053435	5SrRNA:CR33435	2606	3258	54
FBgn0053358	5SrRNA:CR33358	2595	3264	66
FBgn0053388	5SrRNA:CR33388	2595	3230	62
FBgn0053426	5SrRNA:CR33426	2557	3183	67
FBgn0053373	5SrRNA:CR33373	2556	3189	64
FBgn0053425	5SrRNA:CR33425	2554	3281	60
FBgn0053382	5SrRNA:CR33382	2546	2981	59
FBgn0053362	5SrRNA:CR33362	2545	3272	79
FBgn0053415	5SrRNA:CR33415	2519	2979	57
FBgn0053376	5SrRNA:CR33376	2505	3163	69
FBgn0053377	5SrRNA:CR33377	2494	3198	59

Table 7S. II. List of the ten most abundant transcripts in human and *Drosophila* EV samples following rRNA subtraction

HepG2 exRNA (*H.s.*)			A431 exRNA (*H.s.*)		
Gene_ID	Name	Biotype	Gene_ID	Name	Biotype
ENSG00000265150	RN7SL2	misc_RNA	ENSG00000265150	RN7SL2	misc_RNA
ENSG00000258486	RN7SL1	misc_RNA	ENSG00000258486	RN7SL1	misc_RNA
ENSG00000201098	RNY1	misc_RNA	ENSG00000259001	RPPH1	lincRNA
ENSG00000258486	RN7SL1	misc_RNA	ENSG00000269900	RMRP	lincRNA
ENSG00000259001	RPPH1	lincRNA	ENSG00000199990	VTRNA1-1	misc_RNA
ENSG00000266037	RN7SL3	misc_RNA	ENSG00000222328	RNU2-2P	snRNA
ENSG00000198886	MT-ND4	mRNA	ENSG00000200156	RNU5B-1	snRNA
ENSG00000198712	MT-CO2	mRNA	ENSG00000202538	RNU4-2	snRNA
ENSG00000202198	RN7SK	misc_RNA	ENSG00000202198	RN7SK	misc_RNA
ENSG00000207142	Y_RNA	misc_RNA	ENSG00000263934	SNORD3A	snoRNA

D17 exRNA (*D.m.*)			S2R+ exRNA (*D.m.*)		
Gene_ID	Name	Biotype	Gene_ID	Name	Biotype
FBgn0003938	snRNA:U5:63BC	snRNA	FBgn0082973	snoRNA:Psi28S-3308	snoRNA
FBgn0085364	CR34335	lincRNA	FBgn0085364	CR34335	lincRNA
FBgn0058469	CR40469	lincRNA	FBgn0086659	snoRNA:Psi18S-176	snoRNA
FBgn0085742	CR40546	lincRNA	FBgn0086667	snoRNA:Psi28S-3342	snoRNA
FBgn0086659	snoRNA:Psi18S-176	snoRNA	FBgn0002622	RpS3	mRNA
FBgn0086667	snoRNA:Psi28S-3342	snoRNA	FBgn0083057	snoRNA:Psi18S-110	snoRNA
FBgn0065098	RNaseMRP:RNA	lincRNA	FBgn0000556	Ef1alpha48D	mRNA
FBgn0082973	snoRNA:Psi28S-3308	snoRNA	FBgn0058469	CR40469	lincRNA
FBgn0083057	snoRNA:Psi18S-110	snoRNA	FBgn0261602	RpL8	mRNA
FBgn0083040	snoRNA:Psi18S-1854c	snoRNA	FBgn0066084	RpL41	mRNA

Table 7S. III. List of the ten most abundant repeat element-derived RNAs in human and *Drosophila* EVs

HepG2 exRNA (*H.s.*)		A431 exRNA (*H.s.*)		D17 exRNA (*D.m.*)		S2R+ exRNA (*D.m.*)	
Name	Coordinates	Name	Coordinates	Name	Coordinates	Name	Coordinates
THE1C-int7700	chr5: 71,145,492-71,146,744	THE1C-int7700	chr5: 71,145,492-71,146,744	TART	chrX: 15,069-26,164	Copia_I-int39252	chr2L: 22,201,683-22,206,275
HERVK-int980	chr2: 230,045,354-230,045,487	L1PA227648	chrX: 108,297,820-108,302,024	(AT)n	chrX: 3,439,872-3,439,921	DIVER2_I-int16214	chr3R: 1,283,864-1,286,948
7SLRNA2870	chr14: 50,329,267-50,329,567	HERVK-int980	chr2: 230,045,354-230,045,487	(CAAT)n	chr2R: 10,245,859-10,245,899	TART57386	chrX: 15,069-26,164
7SLRNA2827	chr14: 50,053,297-50,053,597	HERVK-int25623	chr19: 36,063,339-36,066,505	(T)n	chr3R: 19,555,771-19,555,805	Gypsy2-LTR_DM969	chr2L: 22,427,715-22,428,019
L1PA227648	chrX: 108,297,820-108,302,024	7SLRNA2827	chr14: 50,053,297-50,053,597	Copia_I-int	chr3L: 2,108,581-2,113,175	TC1-2_DM11645	chr2R: 1,650,689-1,652,229
tRNA-Lys-AAG558	chr19: 36,066,749-36,066,825	tRNA-Lys-AAG558	chr19: 36,066,749-36,066,825	Copia_I-int	chr2L: 20,159,291-20,163,993	(CAG)n318	chr3R: 26,032,306-26,032,414
MER21C370	chr1: 91,852,712-91,852,785	7SLRNA2870	chr14: 50,329,267-50,329,567	(CA)n	chrX: 16,147,995-16,148,061	DM297_I-int35675	chr3R: 11,066,820-11,070,826
L1PA218775	chr5: 71,146,942-71,150,001	L1PA101524	chr21: 9,827,578-9,827,850	Copia_I-int	chr2L: 11,563,967-11,568,561	SAR_DM2740	chr3R: 25,142,562-25,144,328
L1PA101524	chr21: 9,827,578-9,827,850	L1PA218775	chr5: 71,146,942-71,150,001	Copia_I-int	chr2L: 17,820,252-17,824,847	(AT)n18	chrX: 3,439,872-3,439,921
L1PA325725	chr1: 145,277,249-145,277,491	MER21C370	chr1: 91,852,712-91,852,785	Copia_I-int	chr2L: 14,804,210-14,808,803	OSVALDO_I-int617	chr3R: 1,451,474-1,451,728

Table 7S. IV. List of the ten most abundant mRNAs in human and *Drosophila* EVs

HepG2 exRNA (*H.s.*)		A431 exRNA (*H.s.*)	
Transcript ID	Name	Transcript ID	Name
ENSG00000198886	MT-ND4	ENSG00000198938	MT-CO3
ENSG00000198712	MT-CO2	ENSG00000198886	MT-ND4
ENSG00000198804	MT-CO1	ENSG00000198712	MT-CO2
ENSG00000198938	MT-CO3	ENSG00000212907	MT-ND4L
ENSG00000087086	FTL	ENSG00000198899	MT-ATP6
ENSG00000212907	MT-ND4L	ENSG00000198804	MT-CO1
ENSG00000198899	MT-ATP6	ENSG00000198888	MT-ND1
ENSG00000198888	MT-ND1	ENSG00000228253	MT-ATP8
ENSG00000228253	MT-ATP8	ENSG00000198727	MT-CYB
ENSG00000198786	MT-ND5	ENSG00000034510	TMSB10

D17 exRNA (*D.m.*)		S2R+ exRNA (*D.m.*)	
Transcript ID	Name	Transcript ID	Name
FBgn0033926	Arc1	FBgn0002622	RpS3
FBgn0066084	RpL41	FBgn0000556	Ef1alpha48D
FBgn0261619	pAbp	FBgn0261602	RpL8
FBgn0003274	RpLP2	FBgn0066084	RpL41
FBgn0000042	Act5C	FBgn0039757	RpS7
FBgn0032987	RpL21	FBgn0032987	RpL21
FBgn0013674	mt:CoI	FBgn0000559	Ef2b
FBgn0013675	mt:CoII	FBgn0035422	RpL28
FBgn0001233	Hsp83	FBgn0026372	RpL23A
FBgn0000556	Ef1alpha48D	FBgn0023170	RpL39

Table 7S. V. List of the ten most significantly enriched gene ontology (GO) terms associated with human and *Drosophila* EV mRNAs.

HepG2 exRNA (*H.s.*)		A431 exRNA (*H.s.*)	
GO term	FDR q-value	GO term	FDR q-value
translational elongation	1.10E-69	translational elongation	4.00E-81
cytosolic ribosome	1.30E-56	cytosolic ribosome	1.30E-68
structural constituent of ribosome	1.70E-49	ribosomal subunit	1.80E-60
ribosomal subunit	4.30E-48	structural constituent of ribosome	5.20E-59
cytosolic part	2.60E-46	ribosome	1.80E-57
ribosome	1.00E-45	cytosolic part	2.20E-55
translation	1.30E-42	translation	4.00E-50
ribonucleoprotein complex	7.20E-42	ribonucleoprotein complex	2.70E-45
cytosol	1.20E-36	cytosol	1.40E-43
cytosolic small ribosomal subunit	2.10E-28	structural molecule activity	2.30E-42

D17 exRNA (*D.m.*)		S2R+ exRNA (*D.m.*)	
GO term	FDR q-value	GO term	FDR q-value
cytosolic ribosome	7.90E-64	cytosolic ribosome	1.70E-63
cytosolic part	5.20E-57	cytosolic part	2.90E-59
structural constituent of ribosome	6.00E-55	mitotic spindle elongation	9.10E-47
ribosomal subunit	2.10E-47	spindle elongation	1.40E-46
ribosome	2.90E-45	structural constituent of ribosome	1.50E-45
mitotic spindle elongation	7.40E-43	cytosolic large ribosomal subunit	1.20E-44
spindle elongation	1.00E-42	ribosomal subunit	1.60E-36
cytosolic large ribosomal subunit	9.60E-41	ribosome	1.20E-33
lipid particle	8.40E-39	lipid particle	1.00E-32
ribonucleoprotein complex	2.60E-37	mitotic spindle organization	2.10E-31

Table 7S. VI. Exhaustive list of "EV-exclusive" mRNAs in human HepG2 and *Drosophila* D17 cells.

HepG2	
Gene ID	Name
ENSG00000213741	RPS29
ENSG00000156508	EEF1A1
ENSG00000163631	ALB
ENSG00000084674	APOB
ENSG00000130167	TSPAN16

D17			
Gene ID	Name	Gene ID	Name
FBgn0034770	Obp58d	FBgn0052774	Muc4B
FBgn0005640	Eip63E	FBgn0038919	CG17843
FBgn0011288	Snap25	FBgn0000721	for
FBgn0262098	CG42851	FBgn0038508	CG5866
FBgn0052214	CG32214	FBgn0030926	psh
FBgn0040950	Muc26B	FBgn0035436	CG12016
FBgn0032313	CG14070	FBgn0038449	CG17562
FBgn0032851	CG13970	FBgn0036226	CG7252
FBgn0038078	CG14391	FBgn0038979	CG7046
FBgn0085488	CG34459	FBgn0063485	Lasp
FBgn0053554	Nipped-A	FBgn0263780	CG17684
FBgn0038733	CG11407		
FBgn0264395	CG43843		

References

1 Tsui, N. B., Ng, E. K. & Lo, Y. M. Stability of endogenous and added RNA in blood specimens, serum, and plasma. *Clin Chem* **48**, 1647-1653 (2002).

2 Skog, J. *et al.* Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat Cell Biol* **10**, 1470-1476; DOI:ncb1800 10.1038/ncb1800 (2008).

3 Valadi, H. *et al.* Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol* **9**, 654-659; DOI:ncb1596 10.1038/ncb1596 (2007).

4 Baj-Krzyworzeka, M. *et al.* Tumour-derived microvesicles carry several surface determinants and mRNA of tumour cells and transfer some of these determinants to monocytes. *Cancer Immunol Immunother* **55**, 808-818; DOI:10.1007/s00262-005-0075-9 (2006).

5 Raposo, G. & Stoorvogel, W. Extracellular vesicles: exosomes, microvesicles, and friends. *J Cell Biol* **200**, 373-383; DOI:jcb.201211138 10.1083/jcb.201211138 (2013).

6 Muralidharan-Chari, V., Clancy, J. W., Sedgwick, A. & D'Souza-Schorey, C. Microvesicles: mediators of extracellular communication during cancer progression. *J Cell Sci* **123**, 1603-1611; DOI:10.1242/jcs.064386 (2010).

7 Raposo, G. *et al.* B lymphocytes secrete antigen-presenting vesicles. *J Exp Med* **183**, 1161-1172 (1996).

8 Al-Nedawi, K. *et al.* Intercellular transfer of the oncogenic receptor EGFRvIII by microvesicles derived from tumour cells. *Nat Cell Biol* **10**, 619-624; DOI:ncb172510.1038/ncb1725 (2008).

9 Balaj, L. *et al.* Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. *Nat Commun* **2**, 180; DOI:10.1038/ncomms1180 (2011).

10 Crescitelli, R. *et al.* Distinct RNA profiles in subpopulations of extracellular vesicles: apoptotic bodies, microvesicles and exosomes. *J Extracell Vesicles* **2**; DOI:10.3402/jev.v2i0.20677 20677 (2013).

11 Lasser, C. *et al.* RNA-containing exosomes in human nasal secretions. *Am J Rhinol Allergy* **25**, 89-93; DOI:357310.2500/ajra.2011.25.3573 (2011).

12 Kogure, T., Yan, I. K., Lin, W. L. & Patel, T. Extracellular Vesicle-Mediated Transfer of a Novel Long Noncoding RNA TUC339: A Mechanism of Intercellular Signaling in Human Hepatocellular Cancer. *Genes Cancer* **4**, 261-272; DOI:10.1177/1947601913499020 (2013).

13 Zhang, L. *et al.* Microenvironment-induced PTEN loss by exosomal microRNA primes brain metastasis outgrowth. *Nature* **527**, 100-104; DOI:10.1038/nature15376 (2015).

14 Kogure, T., Lin, W. L., Yan, I. K., Braconi, C. & Patel, T. Intercellular nanovesicle-mediated microRNA transfer: a mechanism of environmental modulation of hepatocellular cancer cell growth. *Hepatology* **54**, 1237-1248; DOI:10.1002/hep.24504 (2011).

15 Bobrie, A. & Thery, C. Unraveling the physiological functions of exosome secretion by tumors. *Oncoimmunology* **2**, e22565; DOI:10.4161/onci.225652012ONCOIMM0322 (2013).

16 Deregibus, M. C. *et al.* Endothelial progenitor cell derived microvesicles activate an angiogenic program in endothelial cells by a horizontal transfer of mRNA. *Blood* **110**, 2440-2448; DOI:blood-2007-03-07870910.1182/blood-2007-03-078709 (2007).

17 Sjostrom, A. E., Sandblad, L., Uhlin, B. E. & Wai, S. N. Membrane vesicle-mediated release of bacterial RNA. *Sci Rep* **5**, 15329; DOI:10.1038/srep15329 (2015).

18 Bayer-Santos, E., Lima, F. M., Ruiz, J. C., Almeida, I. C. & da Silveira, J. F. Characterization of the small RNA content of Trypanosoma cruzi extracellular vesicles. *Mol Biochem Parasitol* **193**, 71-74; DOI:10.1016/j.molbiopara.2014.02.004 (2014).

19 Peres da Silva, R. *et al.* Extracellular vesicle-mediated export of fungal RNA. *Sci Rep* **5**, 7763; DOI:10.1038/srep07763 (2015).

20 Cody, N. A., Iampietro, C. & Lécuyer, E. The many functions of mRNA localization during normal development and disease: From pillar to post. *WIREs Dev Biol*; DOI: 10.1002/wdev.113 (2013).

21 Lecuyer, E. *et al.* Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* **131**, 174-187 (2007).

22 Beckett, K. *et al.* Drosophila S2 cells secrete wingless on exosome-like vesicles but the wingless gradient forms independently of exosomes. *Traffic* **14**, 82-96; DOI:10.1111/tra.12016 (2012).

23 Gross, J. C., Chaudhary, V., Bartscherer, K. & Boutros, M. Active Wnt proteins are secreted on exosomes. *Nat Cell Biol* **14**, 1036-1045; DOI:ncb257410.1038/ncb2574 (2012).

24 Koppen, T. *et al.* Proteomics analyses of microvesicles released by Drosophila Kc167 and S2 cells. *Proteomics* **11**, 4397-4410; DOI:10.1002/pmic.201000774 (2011).

25 Currie, J. D. & Rogers, S. L. Using the Drosophila melanogaster D17-c3 cell culture system to study cell motility. *Nat Protoc* **6**, 1632-1641; DOI:nprot.2011.39710.1038/nprot.2011.397 (2011).

26 Ui, K., Ueda, R. & Miyake, T. Cell lines from imaginal discs of Drosophila melanogaster. *In Vitro Cell Dev Biol* **23**, 707-711 (1987).

27 Yanagawa, S., Lee, J. S. & Ishimoto, A. Identification and characterization of a novel line of Drosophila Schneider S2 cells that respond to wingless signaling. *J Biol Chem* **273**, 32353-32359 (1998).

28 Cherbas, L. *et al.* The transcriptional diversity of 25 Drosophila cell lines. *Genome Res* **21**, 301-314; DOI:gr.112961.11010.1101/gr.112961.110 (2011).

29 Ullrich, A. *et al.* Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature* **309**, 418-425 (1984).

30 Ramaiahgari, S. C. *et al.* A 3D in vitro model of differentiated HepG2 cell spheroids with improved liver-like properties for repeated dose high-throughput toxicity studies. *Arch Toxicol* **88**, 1083-1095; DOI:10.1007/s00204-014-1215-9 (2014).

31 Fevrier, B. & Raposo, G. Exosomes: endosomal-derived vesicles shipping extracellular messages. *Curr Opin Cell Biol* **16**, 415-421; DOI:10.1016/j.ceb.2004.06.003 S0955067404000729 (2004).

32 Fertig, E. T., Gherghiceanu, M. & Popescu, L. M. Extracellular vesicles release by cardiac telocytes: electron microscopy and electron tomography. *J Cell Mol Med* **18**, 1938-1943; DOI:10.1111/jcmm.12436 (2014).

33 Chevillet, J. R. *et al.* Quantitative and stoichiometric analysis of the microRNA content of exosomes. *Proc Natl Acad Sci U S A* **111**, 14888-14893; DOI:10.1073/pnas.1408301111 (2014).

34 Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* **35**, W43-46; DOI:10.1093/nar/gkm234 (2007).

35 Johnson, L. F., Levis, R., Abelson, H. T., Green, H. & Penman, S. Changes in RNA in relation to growth of the fibroblast. IV. Alterations in theproduction and processing of mRNA and rRNA in resting and growing cells. *J Cell Biol* **71**, 933-938 (1976).

36 Li, M. *et al.* Analysis of the RNA content of the exosomes derived from blood serum and urine and its potential as biomarkers. *Philos Trans R Soc Lond B Biol Sci* **369**; DOI:10.1098/rstb.2013.0502 (2014).

37 Rosenbloom, K. R. *et al.* The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res* **43**, D670-681; DOI:10.1093/nar/gku1177 (2015).

38 Nair, M. & Saxena, P. Recent patents on mesenchymal stem cell mediated therapy in inflammatory diseases. *Recent Pat Inflamm Allergy Drug Discov* **7**, 105-113 (2013).

39 Lapidot, M. & Pilpel, Y. Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep* **7**, 1216-1222; DOI:10.1038/sj.embor.7400857 (2006).

40 Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811; DOI:10.1038/35888 (1998).

41 Mayer, C., Schmitz, K. M., Li, J., Grummt, I. & Santoro, R. Intergenic transcripts regulate the epigenetic state of rRNA genes. *Mol Cell* **22**, 351-361; DOI:10.1016/j.molcel.2006.03.028 (2006).

42 Xue, S. & Barna, M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat Rev Mol Cell Biol* **13**, 355-369; DOI:10.1038/nrm3359 (2012).

43 Roy, A. M. *et al.* Upstream flanking sequences and transcription of SINEs. *J Mol Biol* **302**, 17-25; DOI:10.1006/jmbi.2000.4027 (2000).

44 List, A. F. *et al.* Overexpression of the major vault transporter protein lung-resistance protein predicts treatment outcome in acute myeloid leukemia. *Blood* **87**, 2464-2469 (1996).

45 Scheffer, G. L. *et al.* The drug resistance-related protein LRP is the human major vault protein. *Nat Med* **1**, 578-582 (1995).

46 Mossink, M. H., van Zon, A., Scheper, R. J., Sonneveld, P. & Wiemer, E. A. Vaults: a ribonucleoprotein particle involved in drug resistance? *Oncogene* **22**, 7458-7467; DOI:10.1038/sj.onc.1206947 (2003).

47 Kedersha, N. L. & Rome, L. H. Isolation and characterization of a novel ribonucleoprotein particle: large structures contain a single species of small RNA. *J Cell Biol* **103**, 699-709 (1986).

48 Persson, H. *et al.* The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. *Nat Cell Biol* **11**, 1268-1271; DOI:10.1038/ncb1972 (2009).

49 McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res* **41**, W597-600; DOI:10.1093/nar/gkt376 (2013).

50 Altman, S. A view of RNase P. *Mol Biosyst* **3**, 604-607; DOI:10.1039/b707850c (2007).

51 Schmitt, M. E. & Clayton, D. A. Yeast site-specific ribonucleoprotein endoribonuclease MRP contains an RNA component homologous to mammalian RNase MRP RNA and essential for cell viability. *Genes Dev* **6**, 1975-1985 (1992).

52 Sauliere, J. *et al.* The exon junction complex differentially marks spliced junctions. *Nat Struct Mol Biol* **17**, 1269-1271; DOI:10.1038/nsmb.1890 (2010).

53 Ender, C. *et al.* A human snoRNA with microRNA-like functions. *Mol Cell* **32**, 519-528; DOI:10.1016/j.molcel.2008.10.017 (2008).

54 Saraiya, A. A. & Wang, C. C. snoRNA, a novel precursor of microRNA in Giardia lamblia. *PLoS Pathog* **4**, e1000224; DOI:10.1371/journal.ppat.1000224 (2008).

55 Martens-Uzunova, E. S., Olvedy, M. & Jenster, G. Beyond microRNA--novel RNAs derived from small non-coding RNA and their implication in cancer. *Cancer Lett* **340**, 201-211; DOI:10.1016/j.canlet.2012.11.058 (2013).

56 Sharma, U. *et al.* Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **351**, 391-396; DOI:10.1126/science.aad6780 (2016).

57 Vickers, K. C., Roteta, L. A., Hucheson-Dilks, H., Han, L. & Guo, Y. Mining diverse small RNA species in the deep transcriptome. *Trends Biochem Sci* **40**, 4-7; DOI:10.1016/j.tibs.2014.10.009 (2015).

58 Mittelbrunn, M. & Sanchez-Madrid, F. Intercellular communication: diverse structures for exchange of genetic information. *Nat Rev Mol Cell Biol* **13**, 328-335; DOI:10.1038/nrm3335 (2012).

59 Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* **Chapter 4**, Unit 4 10; DOI:10.1002/0471250953.bi0410s25 (2009).

60 Levis, R. W., Ganesan, R., Houtchens, K., Tolar, L. A. & Sheen, F. M. Transposons in place of telomeric repeats at a Drosophila telomere. *Cell* **75**, 1083-1093 (1993).

61 Wright, W. E., Tesmer, V. M., Huffman, K. E., Levene, S. D. & Shay, J. W. Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev* **11**, 2801-2809 (1997).

62 Ulaner, G. A., Hu, J. F., Vu, T. H., Giudice, L. C. & Hoffman, A. R. Telomerase activity in human development is regulated by human telomerase reverse transcriptase (hTERT) transcription and by alternate splicing of hTERT transcripts. *Cancer Res* **58**, 4168-4172 (1998).

63 Fukuda, M. *et al.* CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature* **390**, 308-311; DOI:10.1038/36894 (1997).

64 Pardue, M. L. & DeBaryshe, P. G. Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annu Rev Genet* **37**, 485-511; DOI:10.1146/annurev.genet.38.072902.093115 (2003).

65 Mason, J. M., Frydrychova, R. C. & Biessmann, H. Drosophila telomeres: an exception providing new insights. *Bioessays* **30**, 25-37; DOI:10.1002/bies.20688 (2008).

66 Mason, J. M. & Biessmann, H. The unusual telomeres of Drosophila. *Trends Genet* **11**, 58-62 (1995).

67 Shpiz, S., Kwon, D., Rozovsky, Y. & Kalmykova, A. rasiRNA pathway controls antisense expression of Drosophila telomeric retrotransposons in the nucleus. *Nucleic Acids Res* **37**, 268-278; DOI:10.1093/nar/gkn960 (2009).

68 Klenov, M. S. *et al.* Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the Drosophila melanogaster germline. *Nucleic Acids Res* **35**, 5430-5438; DOI:10.1093/nar/gkm576 (2007).

69 Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320-324; DOI:10.1126/science.1129333 (2006).

70 Saito, K. *et al.* Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes Dev* **20**, 2214-2222; DOI:10.1101/gad.1454806 (2006).

71 Bou-Abdallah, F. The iron redox and hydrolysis chemistry of the ferritins. *Biochim Biophys Acta* **1800**, 719-731; DOI:10.1016/j.bbagen.2010.03.021 (2010).

72 Choi, D. S. *et al.* Proteomic analysis of microvesicles derived from human colorectal cancer ascites. *Proteomics* **11**, 2745-2751; DOI:10.1002/pmic.201100022 (2011).

73 Conde-Vancells, J. *et al.* Characterization and comprehensive proteome profiling of exosomes secreted by hepatocytes. *J Proteome Res* **7**, 5157-5166 (2008).

74 Graner, M. W. *et al.* Proteomic and immunologic analyses of brain tumor exosomes. *FASEB J* **23**, 1541-1557; DOI:10.1096/fj.08-122184 (2009).

75 Pan, B. T., Teng, K., Wu, C., Adam, M. & Johnstone, R. M. Electron microscopic evidence for externalization of the transferrin receptor in vesicular form in sheep reticulocytes. *J Cell Biol* **101**, 942-948 (1985).

76 Johnstone, R. M., Mathew, A., Mason, A. B. & Teng, K. Exosome formation during maturation of mammalian and avian reticulocytes: evidence that exosome release is a major route for externalization of obsolete membrane proteins. *J Cell Physiol* **147**, 27-36; DOI:10.1002/jcp.1041470105 (1991).

77 Miguet, L. *et al.* Proteomic analysis of malignant lymphocyte membrane microparticles using double ionization coverage optimization. *Proteomics* **6**, 153-171; DOI:10.1002/pmic.200500133 (2006).

78 Kobayashi, H., Yamamoto, S., Maruo, T. & Murakami, F. Identification of a cis-acting element required for dendritic targeting of activity-regulated cytoskeleton-associated protein mRNA. *Eur J Neurosci* **22**, 2977-2984; DOI:10.1111/j.1460-9568.2005.04508.x (2005).

79 Kawashima, T. *et al.* Synaptic activity-responsive element in the Arc/Arg3.1 promoter essential for synapse-to-nucleus signaling in activated neurons. *Proc Natl Acad Sci U S A* **106**, 316-321; DOI:10.1073/pnas.0806518106 (2009).

80 Hu, Y. *et al.* An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* **12**, 357; DOI:10.1186/1471-2105-12-357 (2011).

81 Batagov, A. O., Kuznetsov, V. A. & Kurochkin, I. V. Identification of nucleotide patterns enriched in secreted RNAs as putative cis-acting elements targeting them to exosome nano-vesicles. *BMC Genomics* **12 Suppl 3**, S18; DOI:10.1186/1471-2164-12-S3-S18 (2011).

82 Tomasoni, R. *et al.* SNAP-25 regulates spine formation through postsynaptic binding to p140Cap. *Nat Commun* **4**, 2136; DOI:10.1038/ncomms3136 (2013).

83 Tweedie, S. *et al.* FlyBase: enhancing Drosophila Gene Ontology annotations. *Nucleic Acids Res* **37**, D555-559; DOI:10.1093/nar/gkn788 (2009).

84 Martin, K. C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719-730 (2009).

85 Binns, D. *et al.* QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045-3046; DOI:10.1093/bioinformatics/btp536 (2009).

86 Huntley, R. P. *et al.* The GOA database: gene Ontology annotation updates for 2015. *Nucleic Acids Res* **43**, D1057-1063; DOI:10.1093/nar/gku1113 (2015).

87 Greening, D. W., Xu, R., Ji, H., Tauro, B. J. & Simpson, R. J. A protocol for exosome isolation and characterization: evaluation of ultracentrifugation, density-gradient separation, and immunoaffinity capture methods. *Methods Mol Biol* **1295**, 179-209; DOI:10.1007/978-1-4939-2550-6_15 (2015).

88 Quail, M. A. *et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, 341; DOI:10.1186/1471-2164-13-341 (2012).

89 Thery, C., Amigorena, S., Raposo, G. & Clayton, A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Curr Protoc Cell Biol* **Chapter 3**, Unit 3 22; DOI:10.1002/0471143030.cb0322s30 (2006).

Préface au chapitre 8

Ce chapitre est présenté sous la forme d'un article de recherche, publié dans le *Journal of Molecular Biology* : (Ref : Lefebvre, F. A., Bouvrette, L. P. B., Bergalet, J., & Lécuyer, E. (2017). Biochemical fractionation of time-resolved drosophila embryos reveals similar transcriptomic alterations in replication checkpoint and histone mRNA processing mutants. Journal of molecular biology, 429(21), 3264-3279.).

L'article traite en premier lieu de l'application d'une technique de purification nucléocytoplasmique couplée à un profilage transcriptomique dans le contexte de l'embryogenèse de la Drosophile. Les transcrits associés à une forte asymétrie d'expression à la fois spatiale et temporale sont identifiés et leurs propriétés sont contrastées. L'impact de la déplétion de deux facteurs maternels, Chk1 et dSLBP, est ensuite évalué sur ces populations de transcrits asymétriques.

Dans le contexte de la présente thèse, les résultats exposés dans cet article font suite au déploiement chez l'embryon de mouche de techniques de fractionnement subcellulaire couplé à l'analyse transcriptomique pour mieux évaluer la localisation des ARNs.

J'ai écrit l'intégralité du manuscrit. J'ai effectué les expériences et les analyses liées aux vésicules extracellulaires ainsi que la préparation des figures présentées dans cette publication. Louis Philip Benoit Bouvrette, étudiant au PhD au laboratoire, a contribué aux analyses de séquençage à haut débit. Julie Bergalet, associée de recherche du laboratoire, a effectué les analyses de transcriptomique sur les ovaires de Drosophile. Dr Éric Lécuyer, mon superviseur, a contribué à l'étude en amassant les financements requis et en relisant le texte pour y suggérer diverses améliorations.

Il est à noter que les données supplémentaires mentionnées dans ce chapitre ont fait l'objet d'une publication indépendantes, qui a été exclue de cette thèse. Le données en question ne figurent donc pas dans le preésent document, mais peuvent être consultées en ligne, où elle sont

disponibles en licence libre de droits : (Ref : Lefebvre, F. A., Bouvrette, L. P. B., Bergalet, J., & Lécuyer, E. (2017). Data for the generation of RNA spatiotemporal distributions and interpretation of Chk1 and SLBP protein depletion phenotypes during Drosophila embryogenesis. Data in brief, 13, 28-31.)

Chapitre 8 : Biochemical Fractionation of Time-Resolved Drosophila Embryos Reveals Similar Transcriptomic Alterations in Replication Checkpoint and Histone mRNA Processing Mutants (Article #6)

Chapitre 8 : Article #6

Biochemical Fractionation of Time-Resolved *Drosophila* Embryos Reveals Similar Transcriptomic Alterations in Replication Checkpoint and Histone mRNA Processing Mutants

Fabio Alexis Lefebvre[1,2], Louis Philip Benoit Bouvrette[1,2], Julie Bergalet[1], Eric Lécuyer[1,2,3,4]

1- Institut de Recherches Cliniques de Montréal (IRCM)
Montréal, Québec, Canada

2- Département de Biochimie
Université de Montréal, Montréal, Québec, Canada

3- Division of Experimental Medicine
McGill University, Montréal, Québec, Canada

4- Address correspondences to:
Dr. Eric Lécuyer
IRCM, RNA Biology Laboratory
110 Avenue des Pins, Ouest
Montréal, Québec, Canada
H2W 1R7
Tel: 514-987-5646, Fax: 514-987-5752
Email: Eric.Lecuyer@ircm.qc.ca

Abstract

In higher eukaryotes, maternally provided gene products drive the initial stages of embryogenesis until the zygotic transcriptional program takes over, a developmental process called the midblastula transition (MBT). In addition to zygotic genome activation, the MBT involves alterations in cell-cycle length and the implementation of DNA damage/replication checkpoints that serve to monitor genome integrity. Previous work has shown that mutations affecting histone mRNA metabolism or DNA replication checkpoint factors severely impact developmental progression through the MBT, prompting us to characterize and contrast the transcriptomic impact of these genetic perturbations. In this study, we define gene expression profiles that mark early embryogenesis in *Drosophila* through transcriptomic analyses of developmentally staged (early syncytial vs late blastoderm) and biochemically fractionated (nuclear vs cytoplasmic) *wild type* embryos. We then compare the transcriptomic profiles of loss-of-function mutants of the Chk1/Grapes replication checkpoint kinase and the Stem Loop Binding Protein (SLBP), a key regulator of replication-dependent histone mRNAs. Our analysis of RNA spatial and temporal distribution during embryogenesis offers new insights into the dynamics of early embryogenesis. In addition, we find that *grp* and *Slbp* mutant embryos display profound and highly similar defects in gene expression, most strikingly in zygotic gene expression, compromising the transition from a maternal to a zygotic regulation of development.

Highlights

- **Biochemical fractionation of time-resolved *wild type* embryo collections identifies transcripts exhibiting asymmetric spatiotemporal distributions**
- **Mutants of the histone mRNA processing factor SLBP mimic the transcriptomic alterations associated to a mutant of the replication checkpoint factor Chk1**
- **Chk1 and SLBP loss-of-function mutations compromise the expression of over 2,500 RNAs, mostly zygotic transcripts enriched in *wild type* blastoderm embryos**

Introduction

In most metazoans, embryogenesis is initiated by a succession of rapid and synchronous cell divisions termed cleavage cycles[1]. For the most part, the zygotic genome remains transcriptionally silent during cleavage cycles, which are thus driven by maternal proteins and transcripts deposited in the oocyte prior to fertilization. In *Drosophila*, where the zygote initially assumes syncytial architecture, the cleavage stage prevails during the 13 first nuclear divisions. Meanwhile, cell cycles occur in the absence of transcriptional input, replication surveillance or growth and involve rapid succession of S and M phases without intervening gap phases[2]. The mechanisms that determine MBT timing remain elusive and pioneer transcription factor activity likely plays an important role in the process[3]. In one prominent model, the MBT is triggered by a maternally inherited DNA replication factor, which is in excess during the first divisions, successfully concluding replication prior to mitotic entry. As cleavage cycles proceed, cytoplasmic concentration of the maternal regulator is titrated against the rapidly expanding mass of zygotic nuclei, eventually compromising timely completion of S-phase[4]. According to a recent study in *Xenopus*, the maternal proteins responsible for triggering the MBT are the canonical histones H3 and H4[5], which are broadly regulated by the Stem Loop Binding Protein (SLBP) at the post-transcriptional level[6-8]. By the 14^{th} nuclear mitosis, nuclei that have migrated to the embryo cortex undergo the process of cellularization, giving rise to the cellular blastoderm, a primordial epithelium[1]. Concomitant cell cycle lengthening takes place, allowing for robust transcriptional activation of the zygotic genome. Meanwhile, the clearance of many maternally deposited transcripts takes place, a process that effectively promotes MBT completion and the establishment of DNA damage/replication checkpoints[2].

DNA damage/replication checkpoints are a set of highly conserved eukaryotic signal transduction systems relying on the activity of ATM, ATR, Chk1 and Chk2 protein kinases. They safeguard genome integrity by delaying mitotic entry and initiating DNA repair or apoptosis in response to various genomic insults. Diverse mutations in genes encoding DNA damage response components are associated with cancer susceptibility syndromes, highlighting the contribution of surveillance pathways in safeguarding genome integrity [9, 10]. In addition, DNA damage response signalling plays crucial roles in embryogenesis: mutations in *Grapes*

(*grp*), which encodes *Drosophila* homolog of Chk1, cause a Chk2-dependent metaphase arrest at nuclear cycle 13 in *Drosophila,* halting development by preventing zygotic genome activation and leading to a continuation of the maternal cell-cycle program [11-15]. Recent work suggests that the DNA replication checkpoint is triggered by early zygotic transcription, which in turn contributes to remodelling cell cycle parameters during the MBT[16].

In absence of an effective DNA replication checkpoint (e.g. in *grp* or *mei-41/dATR* mutants) or in response to genotoxic stress, nuclei dividing during late cortical cleavage cycles are prone to the accumulation of genomic alterations and undergo mitotic catastrophe[11, 12]. This leads damaged nuclei to fall out of the cortical layer during cycle 10-13 and to accumulate in the underlying yolk area, where they are excluded from the pool of dividing nuclei and targeted for degradation[17]. Previous work has shown that fallout nuclei selectively retain diverse mRNAs encoding proteins required for nuclear viability, which effectively precludes translation of the retained messages and clears out damaged nuclei[18]. In particular, histone mRNA retention is a regulated mechanism wherein DNA damage triggers a Chk2-dependent phosphorylation of SLBP, thus perturbing the nuclear export and translation of histone messages. SLBP stably binds the 3' end of replication-dependent histone mRNAs, which lack a polyA tail and PolyA-binding protein (PABP) affinity. It intrinsically regulates 3'-end processing, post-transcriptional maturation, nuclear export and translation of canonical histone mRNAs[7]. The emerging role of SLBP in nuclear fallout bridges intracellular mRNA shuttling to genome integrity surveillance, and links histone mRNA processing to DNA damage response pathways[18]. Maternal histone proteins and mRNAs are abundantly deposited in the early embryo, in order to sustain the proliferative pace of the cleavage cycles. The function of SLBP is manifest in late cleavage cycles when zygotic histone mRNA processing becomes crucial as the clearance of the maternal histone pools proceeds[19].

Here, we sought to define the transcriptomic alterations resulting from loss-of-function mutations in the SLBP protein and compare them to the phenotype of Chk1 depletion during embryogenesis. These proteins are required for developmental passage through the MBT and we hypothesized that their alterations would selectively compromise zygotic transcript expression. We first defined subpopulations of transcripts displaying asymmetry in space and time during normal embryogenesis and validated our spatiotemporal distributions using established lists of ovarian and purely zygotic RNAs. Our findings indicate that RNAs enriched during the MBT

arise from shorter genes that encode less isoforms per gene than RNAs enriched at earlier stages, which are largely contributed maternally. Embryo fractionation showed that RNAs localized in the blastoderm cytoplasm are shorter and simpler than products associated with zygotic nuclei. Functionally, messages encoding infrastructural proteins involved in broad metabolic and translational processes prevailed in early syncytial embryos and in cytoplasmic extracts of blastoderm stage embryos. Transcripts enriched in nuclear extracts of blastoderm embryos were highly related to DNA binding and transcription factor activity, underlying the first wave of developmental genes expressed from zygotic nuclei. We show that the transcriptomic alterations observed in the *Slbp*$^{10/12}$ and the *Slbp*$^{10/15}$ mutants largely phenocopy the *grp*^{fs1} mutant. These profiles are consistent with a developmental halt in all three mutants wherein the levels of over 2,000 genes are severely compromised, most strikingly ~70% of zygotic genes normally enriched in blastoderm stage embryos.

Results

Generation of transcriptomic datasets from staged and fractionated embryos

We sought to compare the impact of Chk1 and SLBP protein depletion on the transcriptomic landscape of early embryogenesis via RNA sequencing (RNA-seq). We also aimed to contrast the impact of these depletions on RNAs enriched before and during the MBT in *wild type* (*wt*) Oregon R embryos. To define repertoires of RNAs enriched in the '**Early/E**' cleavage (0-45 minutes AEL, mitotic cycles 1-6) and '**Late/L**' blastoderm (90-180 min AEL; mitotic cycles 10-14) periods, we profiled staged *wt* embryo collections by RNA-seq (**Figure 8.1A**). To assess the spatial distribution of RNA, we also derived '**Nuclear/N**' and '**Cytoplasmic/C**' extracts from 90-180 min AEL blastoderm embryos using a subcellular fractionation approach[20].

Prior to RNA-seq, staged embryo collections were fixed, stained for DNA with DAPI and examined for morphological validation[21]. Embryos collected 0-45 min AEL displayed a preblastoderm identity, corresponding to mitotic cycles 1-6, while 90-180 min AEL embryos typically adopted a syncytial or cellularized blastoderm morphology, associated with mitotic cycles 10-14. We purified and deep-sequenced biological duplicates of a total of 7 RNA samples from each genotype (*wt*, grp^{fs1}, $Slbp^{10/12}$, $Slbp^{10/15}$, 0-180 min AEL), staged collection (0-45 min AEL and 90-180 min AEL) and subcellular extract (cytoplasmic and nuclear, 90-180 min AEL). We obtained between 24 and 41 million reads per condition, mapping over 7,000 genes that reached a threshold of ≥5 normalized reads per condition (**Table 8.I**).

Systematic survey of RNA spatiotemporal distributions in *Drosophila* embryogenesis

We first aimed to systematically define the spatiotemporal profile of transcripts expressed during embryogenesis, using staged and fractionated *wt* embryos. While most RNAs were identified jointly in the early (E), late cytoplasmic (C) and late nuclear (N) datasets, others were highly enriched or exclusively traced in either population, indicating that our approach can capture signatures displaying strong asymmetry (**Figure 8.1B**). In order to define RNAs enriched

in the early, late, nuclear and cytoplasmic populations, we derived ratios based on normalized read counts. For example, the cytoplasmic enrichment score C_y for a given RNA was defined as $C_y=2\times C_x/(E_x+C_x+N_x)$, where E_x, C_x and N_x are the number of normalized reads mapped in the early syncytial, cytoplasmic blastoderm and nuclear blastoderm datasets. Mapped genes were sorted based on this metric, yielding exponential distributions. We sought to define an unbiased cut-off reflecting the dispersion of each function and thus we selected enriched values exceeding $y_{min}\times(y_{max}/y_{min})^{0.6}$ for each exponential distribution[22, 23] (**Figure 8.2A**). This exponential approach enables the selection of genes responsible for the top 40% dispersion of the enrichment score distribution. We defined 2,821 nuclear RNAs, 1,187 cytosolic RNAs, 364 early syncytial and 360 late blastoderm RNAs. Using trigonometric constants, RNA enrichment values were then plotted as a triangular simplex graph[24] in which relative proximity to a corner of the triangle reflects gene-specific enrichment values (**Figure 8.2B**).

We reasoned that RNAs highly enriched during early cycles broadly correspond to maternal transcripts targeted for clearance at the MBT, while products highly enriched at later stages essentially result from zygotic transcription. In agreement with this view, we identified well-characterized developmentally regulated mRNAs in each list: *vasa*, *hoi-polloi*, *Bicaudal C* and *exuperantia*, for example, were all enriched in the early syncytial pool[21] and have been associated with maternal deposition and targeted clearance at the MBT[25, 26]. Among RNAs enriched in late blastoderm, we found diverse regulators tied to early zygotic transcription[27], such as *runt*, *snail*, *Kruppel* and *patched[21]*. Over two thirds of genes enriched in late blastoderm were also enriched in nuclei (**Figure 8.2C**), among which figured canonical zygotic products[28] such as *sog*, *patched* and *slit*[21]. Non-coding RNAs were prominent in the blastoderm cytoplasmic pool (**Figure 8.2D**). In addition, maternally provided mRNAs such as *slmo*, *gnu* and diverse messages of ribosomal proteins were enriched, likely reflecting the cytoplasmic stabilization of maternally deposited mRNAs[21], consistent with previous observations[28].

Zygotic genome activation in *Drosophila* has been intensively studied through diverse approaches and we took advantage of two established resources to contrast our list of temporally regulated RNAs with surveys of zygotic transcripts. *De Renzis et al.* developed a combination of

chromosomal deletions and staged collections to profile zygotic transcription using microarrays[29], while *Lécuyer et al.*, performed a large scale *in situ* hybridization screen[28] and identified nearly 300 protein-coding genes expressing nascent zygotic RNAs during mitotic cycles 10-14. The 235 products defined as "purely zygotic" by *De Renzis et al.* were strongly enriched in the left corner of our simplex graph of spatiotemporal RNA distributions, a profile indicative of strong enrichments in late blastoderm embryos. We observed a similar trend with the selection from *Lécuyer et al.*, which was broadly enriched in vicinity of the nuclear blastoderm corner of the graph (**Figure 8.3A**). In spite of methodological discrepancies, 37% of the "purely zygotic" genes identified by *De Renzis* figured in our selection of RNAs temporally enriched in late embryos, while 74% of the genes reported by *De Renzis* were spatially enriched in late nuclei. Although only 9% of our list of transcripts enriched in late blastoderm embryo was annotated as zygotic foci by *Lécuyer et al.*, 54% of the RNAs reported in the study were enriched in blastoderm nuclei (**Figure 8.3B**). Importantly, the assessment of *Lécuyer et al.* is based on micrographs of nascent zygotic RNAs, many of which are maternally deposited in addition to being transcribed in the zygote[28]. By contrast, the work of *De Renzis et al.* relies on timely resolved collections of embryos bearing chromosomal deletions to track transcripts exclusively provided by zygotic transcription and lacking maternal deposition, called "purely zygotic". The lists of transcripts provided by *De Renzis et al.* and *Lécuyer et al.* thus refer to two sets of conceptually different zygotic products and our data is a closer reflection of *De Renzis et al.* results. Consistently, our selection of RNAs enriched in early syncytial embryos and in the cytoplasmic fraction of late blastoderm embryos were poorly related to the two zygotic gene surveys (0-6% and 1-3%). This result suggests that RNAs enriched in early syncytial embryos and in the cytoplasmic pool of blastoderm embryos are largely deposited maternally.

Next, we compared a dataset of *wt* total ovary extracts with spatiotemporal distributions of RNA expressed during embryogenesis in an attempt to track the fate of maternal products as embryogenesis proceeds. Correlative approaches showed that early syncytial transcripts are a closer reflection of ovarian profiles (Pearson's r=0.74, Spearman's ρ=0.94 ; $P<10^{-4}$) than cytoplasmic (r=0.53 ; ρ=0.89 ; $P<10^{-4}$) and nuclear blastoderm extracts (r=0.69 ; ρ=0.88 ; $P<10^{-4}$) (**Figure 8.3D**). We selected the 1,000 most enriched and 1,000 most abundant ovarian RNAs and plotted their coordinates on a simplex graph of spatiotemporal distributions during embryogenesis (**Figure 8.3E**). We found that enriched ovarian transcripts were broadly

distributed but underrepresented in the left region of the simplex, an area associated with late blastoderm enrichments. Similarly, the 1,000 most abundant RNAs in ovaries were strictly excluded from the left end of the simplex and enriched in the central and right regions, underlying moderate to strong enrichments in early syncytial relative to late blastoderm stages. Together, these comparative efforts indicate that our approach of developmentally staged collections coupled to subcellular fractionation successfully enriches zygotic transcripts in the late blastoderm nuclear fraction. We also show that abundant ovarian RNAs are enriched during early embryogenesis but underrepresented in the syncytial and cellularized blastoderm. This conclusion is consistent with the dilution of the maternal pool by an increasing zygotic contribution and underlies a regulated maternal clearance process.

Properties of spatiotemporally restricted RNAs

Previous work in *Drosophila*, *Xenopus* and *Danio rerio* suggests that the cell-cycle length restrictions that prevail during the cleavage phase restrain early zygotic transcription to short and simple RNAs[30, 31]. In agreement with these findings, the median transcript length of RNAs enriched in MBT stage blastoderm embryos was markedly smaller than in early syncytial samples. In blastoderm embryos, cytoplasmic RNAs were shorter than products enriched in nuclei, likely reflecting the abundance of short non-coding species in the cytoplasm (**Figure 8.4A;** t-test $P<10^{-4}$). Our analysis of isoform counts per gene corroborated the trend: cytoplasmic RNAs exhibited lower transcriptional diversity than nuclear RNAs, and RNAs enriched in early embryogenesis were more diverse than RNAs enriched in blastoderm samples (**Figure 7.4B;** t-test $P<10^{-4}$). We found that cytoplasmic RNAs have a higher GC content than nuclear RNAs, possibly reflecting the prevalence of GC-rich infrastructural transcripts such as snRNAs in this group (**Figure 8.4C;** t-test $P<10^{-4}$) [32].

To glean insights into the transcriptional regulation of RNAs displaying spatial and temporal enrichments, we investigated known transcription factor motif enrichment in sequences up to 2000 bp upstream of the transcription start site (TSS) (**Figure 8.4D**). The most enriched motif among promoters of RNAs enriched in early syncytial embryos was the Caudal consensus sequence, associated with developmental genes involved in establishing posterior segment identity[33]. Interestingly, the sequence CTACCTG (CAGGTAG), which serves as the consensus sequence for a master regulator of the ZGA, the Zinc-finger protein Zelda, was the

most enriched motif among promoters of genes enriched in late blastoderm embryos[34, 35]. Motifs of other zygotic regulators, notably Snail and Hedgehog, were also identified among promoters of transcripts enriched in blastoderm embryos. Meanwhile, promoters of RNAs overrepresented in the cytoplasm of blastoderm embryos were most enriched for DRE (DNA replication-related element), a motif associated with the regulation of genes involved in DNA replication, protein metabolism and cell cycle regulation[36].

Enrichment analysis of *molecular function* ontology confirmed a sizeable overlap of the late blastoderm and nuclear blastoderm mRNA profiles, both enriched[21] with diverse terms related to transcriptional and DNA-binding activity (**Figure 8.4E**). The early syncytial and cytoplasmic blastoderm profiles were both enriched for infrastructural functions related to metabolic and housekeeping roles. "Structural constituent of cuticle" and "chitin binding" were the most enriched terms in the early syncytial pool, along with "polysaccharide binding" and "amino acid transmembrane transporter activity". Meanwhile, cytoplasmic terms were largely related to translation, notably "structural component of the ribosome", "rRNA binding" and "translation initiation activity".

Transcriptomic phenotype of *grp* and *Slbp* loss-of-function mutants

We then performed RNA-seq on collections of *wt*, grp^{fs1}, $Slbp^{10/12}$ and $Slbp^{10/15}$ embryos (0-180 min AEL), which were obtained through crosses of trans-heterozygous mutant females with *wt* OregonR males. We analyzed total embryos rather than subcellular extracts because the low number of embryo collected precluded biochemical fractionation. Read coverage at the *grp* and *Slbp* loci were severely hampered in the corresponding mutants, validating the loss-of-function models (**Figure 8.5A**). Consistent with the role of SLBP in histone mRNA regulation[37], we observed a dramatic reduction in read coverage across the repetitive histone gene cluster for both *Slbp* mutant allelic combinations, a phenotype that was not observed with grp^{fs1} mutants (**Figure 8.5A**). In addition, the levels of approximately 4,200 RNAs in $Slbp^{10/15}$ and nearly 2,200 RNAs in $Slbp^{10/12}$ were down-regulated in excess of 4-fold relative to *wt[21]*. This result is in accordance with previous descriptions by *Sullivan et al.*[37] indicating that the $Slbp^{10/15}$ allelic combination leads to a more severe phenotype than $Slbp^{10/12}$. We found a strong overlap when comparing the grp^{fs1} transcriptomic phenotype with SLBP loss-of-function models, especially in the $Slbp^{10/15}$ mutant (83% overlap; **Figure 8.5B**).

We sought to define the transcriptional regulation disrupted in these mutants and screened the promoters of genes exhibiting a 4-fold or greater decrease in grp^{fs1}, $Slbp^{10/12}$ and $Slbp^{10/15}$ embryos. The TATA-box, the Initiator element (Inr) and the Caudal consensus sequence were the most enriched motifs in promoters of genes affected by all three mutations (**Figure 8.5C**). We also identified motifs for the master zygotic gene regulator Zelda and the morphogen Dorsal, which contributes to dorsoventral axis establishment[38]. Among down-regulated RNAs, "transcription factor activity", "sequence-specific DNA binding" and "transcription regulator activity" were the most highly enriched *molecular function* ontology terms[21] for all three mutants (**Figure 8.5D**). These results reveal strong similarities in the transcriptomic alterations associated with grp^{fs1}, $Slbp^{10/12}$ and $Slbp^{10/15}$ embryos. The TATA-box and Initator elements are major and broad *cis*-regulators of transcription[39, 40]; our analyses thus suggest that severe failure of the gene expression program occurs in grp^{fs1}, $Slbp^{10/12}$ and $Slbp^{10/15}$ mutants.

We next sought to compare the transcripts compromised in grp^{fs1}, $Slbp^{10/12}$ and $Slbp^{10/15}$ mutants with our repertoires of transcripts enriched in *wt* early syncytial embryos, late blastoderm embryos and nuclear and cytoplasmic extracts of blastoderm samples. We first plotted the coordinates of transcripts compromised by the mutation on simplex graphs of *wt* RNA spatiotemporal distributions (**Figure 8.5E**). While transcripts distributed along the early syncyctial to cytoplasmic blastoderm spectrum were typically not affected, transcripts enriched in late nuclear extracts were heavily compromised. Indeed, a high proportion of transcripts enriched in the late blastoderm embryos showed a 4-fold or greater decrease in grp^{fs1} (45%), $Slbp^{10/12}$ (67%) and $Slbp^{10/15}$ (72%) mutants[21]. We then independently evaluated Spearman's (ρ) and Pearson's (r) coefficients across each mutant while selectively considering populations of RNAs enriched in early syncytial and late blastoderm embryos and nuclear and cytoplasmic extracts of blastoderm embryos[21]. The weakest correlations were observed for late blastoderm transcripts, most strikingly in $Slbp^{10/15}$.

To provide a quantitative assessment of these phenotypes, we determined the median fold change value resulting from each mutation for the subsets of RNA previously described: (i) RNAs most abundant in ovaries, (ii) RNAs most enriched in early syncytial samples, (iii) RNAs most enriched in the blastoderm cytoplasm, (iv) RNAs defined as purely zygotic by *De Renzis et al*., (v) RNAs most enriched in late blastoderm embryos and (vi) RNAs most enriched in nuclear extracts of blastoderm nuclei (**Figure 8.6**). To circumvent artifacts resulting from variations in

global expression levels, we selected the 100 transcripts displaying the lowest cumulative read count with a high threshold of 100 reads. Spatiotemporal categories associated with maternal deposition showed relatively high median fold change values. Indeed, abundant ovarian transcripts were poorly affected in all three mutants ($0.88 \leq$ median fold change ≤ 0.99), while transcripts enriched in early syncytial samples were moderately compromised ($0.27 \leq$ median fold change ≤ 0.60), a portrait similar to transcripts enriched in late blastoderm cytoplasmic extracts ($0.43 \leq$ median fold change ≤ 0.82). By contrast, spatiotemporal categories linked to zygotic genome transcription were heavily compromised in grp^{fs1}, $Slbp^{10/12}$ and $Slbp^{10/15}$ mutants. Indeed, we observed strong reductions in the levels of transcripts classified as purely zygotic by *de Renzis et al.* ($0.02 \leq$ median fold change ≤ 0.07), in transcripts enriched in late blastoderm embryos ($0.01 \leq$ median fold change ≤ 0.08) and in blastoderm nuclei ($0.02 \leq$ median fold change ≤ 0.05). Collectively, these results demonstrate that the grp^{fs1}, $Slbp^{10/12}$ and $Slbp^{10/15}$ mutations severely compromise zygotic genome activation.

Discussion

In this study, we aimed to define and compare the transcriptomic landscapes resulting from mutations in the histone maturation factor SLBP and the replication checkpoint factor *Grapes*/Chk1 during *Drosophila* embryogenesis. To interpret these phenotypes, we first studied the normal course of RNA dynamics during embryogenesis and tracked RNA distributions in time and space through nucleocytoplasmic separation of time-resolved collections. We investigated the functional profile, structural properties and transcriptional regulation of transcripts displaying strong spatiotemporal asymmetry. We then turned to grp^{fs1}, $Slbp^{10/12}$ and $Slbp^{10/15}$ mutants, which presented profound and largely synonymous alterations in the levels of over 2,500 transcripts. We focused on transcripts compromised in the mutants and showed that they largely correspond to those enriched in blastoderm embryo nuclei during normal embryogenesis. We conclude that mutations in the histone maturation factor SLBP and the replication checkpoint factor *Grapes*/Chk1 broadly compromises zygotic gene expression, consistent with aborted developmental progression through the MBT.

In metazoans, early embryogenesis commences with a period of high proliferation, which rapidly increases the number of cells that compose the embryo. In organisms such as Drosophila, Xenopus, and Zebrafish, this period involves abbreviated cell cycles with relaxed DNA replication checkpoints [3]. Seminal work has shown that the MBT is initiated when the embryo breaches a threshold nuclei-to-cytoplasm (N:C) ratio, a parameter that increases exponentially during cleavage cycle progression[41-43]. Upon reaching a critical N:C ratio tipping point, embryos undergo several molecular and cellular changes, including broad zygotic genome activation and alterations in cell-cycle regulation. This ensures a transition from maternal to zygotic control of embryogenesis and the proper coordination of downstream cell fate specification and morphogenetic events. The mechanisms that orchestrate the profound changes in gene expression dynamics, cell-cycle regulation, and morphological features associated with the MBT are complex and involve several interlinked processes [44]. In one prominent model, the MBT is triggered by maternally inherited DNA replication factors or transcriptional repressors, which are in excess during the cleavage cycles and respectively function to promote rapid cell cycling and to repress zygotic genome activation. As cleavage cycles proceed, the cytoplasmic concentration of these maternal regulators is titrated against the rapidly expanding mass of zygotic nuclei, eventually impeding the completion of rapid S-phases and offering a

favorable setting for transcriptional activation [45]. Such examples were recently characterized in *Xenopus*, where four replication factors (Tresli, Cut5, RecQ4, and Drf1) were found to become limiting at increasing N:C ratios, thus contributing to a slowing of the cell cycle during the MBT [46]. Moreover, in a recent study,

Amodeo et al. [43] used an in vitro assay with Xenopus egg extracts to show that canonical histones H3 and H4 act as concentration-dependent repressors of zygotic transcription, suggesting that free histone levels also modulate the timing of the MBT. In addition to this titration model involving the exhaustion of transcriptional repressors, zygotic genome activation is also known to be stimulated through the activity of pioneer transcription factors [34,47,48].

While defects in histone metabolism could be sensed and addressed as replicative stress by the zygote, this may also provoke a major impairment of genome organization and function. Indeed, independent of checkpoint activity, compromised histone levels resulting from SLBP depletion may lead to an unregulated and spontaneous collapse of nucleosome assembly and chromatin architecture that could directly compromise viability. As expected, we have found that histone mRNA levels are dramatically decreased in *Slbp* mutants, concordant with previous work in Drosophila showing the maternal deposition of histone protein and mRNA is hampered in these mutants [17,36]. According to a simple interpretation of the maternal factor titration model proposed by *Amodeo et al.* [43], a decrease in H3 and H4 cytoplasmic concentration during early embryogenesis should promote precocious zygotic genome activation, provided that H3 and H4 act as N:C ratio sensors in *Drosophila*. By contrast, the evidence we have gathered studying SLBP mutants suggests that zygotic gene expression is dramatically compromised, delayed or broadly aborted, rather than accelerated. Perhaps, this is not surprising, since it is known that either a shortage or excess of histones can severely impact chromatin function, leading to the disruption of transcription and genotoxic stress [49–52]. Indeed, several layers of regulation participate in coordinating histone levels with DNA replication to ensure the timely assembly of nucleosomes on newly replicated DNA strands while preventing the accumulation of toxic histone monomers when replication is complete or stalled due to DNA damage [53]. While this partly occurs through the activity of histone protein chaperones, the regulation of histone mRNA processing, export and stability during the cell cycle is also a critical control point in higher eukaryotes [17].

In **Figure 8.7**, we provide a general working model of how our results can be interpreted in the context of potential crosstalk occurring among the processes of ZGA, replication checkpoint control, and histone metabolism leading up to the MBT. Our data show that *slbp* and *grp*/Chk1 loss of functions lead to comparable transcriptomic alterations, revealing a strong reduction in the zygotic gene expression program. This is consistent with the observation that these mutants exhibit similar embryonic phenotypes, including excessive mitotic defects, catastrophic mitoses leading to nuclear fallout, a failure to cellularize, and embryonic lethality [6,13,36,54,55]. Chk1 function is required for delaying mitotic events until chromosomes are fully replicated [6,56], and it participates in the progressive cell-cycle lengthening, that occurs at the MBT [57]. While the mechanisms leading to checkpoint activation at the MBT have remained unclear, elegant work by *Blythe and Wieschaus* recently revealed that a primary triggering event is the process of zygotic transcription itself, which induces replicative stress and the eventual slowing down of the cell cycle [58]. If this genotoxic stress is too severe or left unrepaired, as is the case in chk1 mutants or following the exposure of embryos to DNA-damaging drugs, this triggers the activation of a developmental arrest pathway driven by the Chk2 kinase [12,13,57]. Indeed, the phenotypes observed in chk1 mutants are largely suppressed in chk1/chk2 double mutant embryos [57]. While *slbp* mutant embryos exhibit very similar phenotypes to chk1 mutants, our previous work revealed that *chk2* mutation fails to rescue *Slbp* phenotypes and that SBLP is actually a direct functional effector of Chk2 signaling [13]. Indeed, Chk2 directly phosphorylates SLBP on serine 118, leading to its destabilization and the disruption of its normal functions in histone mRNA nuclear export [13]. This process participates in a pathway through which Chk2 signaling inhibits the nuclear export of early zygotic mRNAs expressed prior to the MBT and which serves to trigger the elimination of subsets of nuclei that have accumulated genotoxic stress during the early cleavage cycles. In future work, it will be important to characterize additional players that operated downstream of Chk2 signaling in the developmental arrest pathway that it controls, which will help illuminate the mechanisms linking zygotic genome activation, replication checkpoint activity, and changes in cell behavior that occur at the MBT.

Material and method

Drosophila husbandry, imaging and fractionation

90-180 minutes AEL *wt* embryos were collected over 90 minutes and kept at 25°C for 90 additional minutes. Cytoplasmic and nuclear extracts were derived from 90-180 minutes AEL *wt* embryos as previously described[20]. Mutant embryos were obtained through crosses of trans-heterozygous mothers with *wt* OregonR males. For RNA-seq, embryos were dechorionated[49] and frozen in liquid nitrogen. For DAPI staining, embryos were fixed with formaldehyde as previously described[50] and imaged by epifluorescence on a Leica DM5500B microscope equipped with a QImaging ExiAqua camera (QImaging). Oregon R was used as *wt*.

RNA isolation and library generation for RNA-seq

Cytoplasmic, nuclear and total embryo extracts were homogenized in 1 mL of TRIZol™ (Ambion) using a pellet pestle and processed according to the manufacturer's instructions. RNA extracts were then purified using the RNA Clean & Concentrator™-5 system (Zymo) and on-column DNAse I treatment (New England BioLabs), washes and elution were performed as described by the manufacturer. Optical density was obtained using a NanoDrop2000c spectrophotometer to validate sample purity ($A_{260}/A_{280} \geq 2.0$)($2.00 \leq A_{260}/A_{280} \leq 2.25$). Sequencing libraries were prepared from 1.0 ug of RNA with the Illumina TruSeq Stranded RNA Kit as described by the manufacturer. Ribosomal RNA depletion was performed using the RiboMinus kit (Thermo Fisher). The Illumina HiSeq 2000 machine was used with the TruSeq PE Clusterkit v3-cBot-HS.

RNA-seq data analysis

We used FastQC v0.10.1 for read quality analysis and performed trimming with Trimmomatic when deemed necessary. Tophat v2.0.10 was used for read alignment on the *Drosophila melanogaster* BDGP5.78/dm3 genome. Read coverage was visualized with the UCSC genome browser using bigWig files derived from alignment BAM files. Deseq2 was used to derive

normalized read counts from BAM files and a threshold of 5 reads was applied in downstream analyses. Simplex graphs of RNA spatiotemporal distributions were derived from normalized read counts as previously described. Simplex coordinates were defined as $x=E\times\cos(11\pi/6)+C\times\cos(\pi/2)+N\times\cos(7\pi/6)$ and $y=E\times\sin(11\pi/6)+C\times\sin(\pi/2)+N\times\sin(7\pi/6)$, where E stands for the read count in 0-45 min AEL embryos, C is the read count in 90-180 min AEL cytosolic extracts and N is the read count in 90-180 min AEL nuclear extracts.

Motif enrichment analysis

We used HOMER[51] 4.8 to retrieve sequences upstream of all enriched and down-regulated genes through the promoter set fly-p 5.5, with parameters adjusted to encompass -2000 bp to 0 bp relative to the transcription start site (TSS).

Ontology enrichment analysis

DAVID[52] was used to retrieve enriched GO terms and associated statistical significance values from protein-coding gene selections, using the complete *Drosophila* protein-coding assembly as a background. Heatmaps were assembled with Microsoft Excel, using median-centered color scales.

Acknowledgments

We are grateful to the Theurkauf lab and the Duronio lab, which kindly provided the *grp*^{fs1} [48] stock and the *Slbp*^{10}, *Slbp*^{12} and *Slbp*^{15} stocks[37], respectively. We thank Odile Neyret, Alexis Blanchet-Cohen and the IRCM Molecular Biology Platform staff for RNA-seq advice and active contribution. F.A.L. is funded through a *Canadian Institute of Health Research* (CIHR) scholarship, J.B. benefits from a *Fond de Recherche en Santé du Québec* (FRSQ) scholarship. Funds allocated through a CIHR (MOP-111161) and *Natural Sciences and Engineering Research Council* of Canada (*NSERC* discovery grant #386644) were used in the pursuit of this study.

Author contributions

F.A.L. contributed to the study design, performed the experiments and analyses, assembled the figures and wrote the manuscript. L.P.B.B. performed motif enrichment analyses and contributed to manuscript revision. J.B. generated RNA-seq datasets of ovary extracts and contributed to manuscript revision. E.L. contributed to the study design and manuscript revision.

Figure and Table legends

Figure 8.1. Experimental pipeline and transcriptomic validation

(A) Early *wt* syncytial embryos were collected 0-45 minutes after egg laying (min AEL) and blastoderm embryos collected 90-180 min AEL were processed biochemically to prepare cytoplasmic and nuclear extracts. Embryos were also collected from mothers bearing grp^{fs1}, $Slbp^{10/12}$ and $Slbp^{10/15}$ mutations with *wt* controls (0-180 min AEL).

(B) Transcriptomic validation of embryo staging and subcellular fractionation. UCSC genome browser view showing read coverage at spatiotemporally restricted loci during normal embryogenesis: the *CG1663*, *Arc1*, *sog* and *Gapdh1* loci in early syncytial (top), blastoderm cytoplasmic (center) and blastoderm nuclear extracts (bottom).

Figure 8.2. Spatiotemporal dissection of the early *Drosophila* embryo transcriptome

(A) Datasets of early total syncytial embryos (E) were used with cytoplasmic (C) and nuclear (N) extracts of late blastoderm embryos to derive spatiotemporal enrichment scores, defined as the proportion of total cumulative reads mapped in a given fraction. For example, the cytoplasmic enrichment score C_y for a given RNA is defined as $C_y=2\times C_x/(E_x+C_x+N_x)$, where C_x, E_x and N_x are the number of normalized reads mapped in the cytoplasmic, early and nuclear datasets. Transcriptome-wide enrichment distributions are shown in light color tones. We selected enriched values exceeding $y_{min}\times(y_{max}/y_{min})^{0.6}$, displayed in dark tones.

(B) Simplex graphs representing transcriptome-wide spatiotemporal distributions during embryogenesis. Dark tones denote enrichment values higher than $v_{min}\times(v_{max}/v_{min})^{0.6}$. The number (n) of enriched RNAs in each group is indicated.

(C) Venn diagram of genes defined as early, late, nuclear and cytoplasmic.

(D) Pie charts of biotype distributions for all (left) and for the 100 most highly enriched RNAs (right) in each group.

Figure 8.3. Spatiotemporal distributions of zygotic and ovarian transcripts during embryogenesis

(A) Simplex graph showing the spatiotemporal distribution of 'purely zygotic genes' defined by *De Renzis et al.* (left) and genes identified as nascent zygotic transcripts during stage 4 and 5 by *Lécuyer et al.* (right). Zygotic genes identified in each study (blue) were overlaid on the transcriptome-wide distributions of embryogenesis (black). **(B)** Venn diagram of zygotic genes defined by *De Renzis et al*. and by *Lécuyer et al.* and RNAs enriched in early syncytial and late blastoderm embryos.

(C) Venn diagram of zygotic RNAs and RNAs enriched in the nuclear and cytoplasmic fractions of late blastoderm embryos.

(D) Correlation heatmap of RNA-seq datasets established from total ovary extracts, total early syncytial embryos and cytoplasmic and nuclear extracts of late blastoderm embryos. Pearson's (r) and Spearman's (ρ) correlation coefficients with associated *P*-values designated by asterisks are shown.

(E) Simplex graph of spatiotemporal distributions during embryogenesis are shown for the 1,000 most enriched (top) and most abundant (bottom) RNAs in ovary extracts. Ovary RNAs (beige) were overlaid on transcriptome-wide distributions of embryogenesis (black).

Figure 8.4. Properties of transcripts displaying asymmetric spatiotemporal profiles

(A) Box plot of RNA length distributions for enriched RNAs in each fraction. Median length ($\mu_{1/2}$) is indicated.

(B) Box plot of gene-specific isoform counts for genes enriched in each fraction. Median ($\mu_{1/2}$) and average values (μ) are indicated.

(C) Box plot of GC content for enriched RNAs in each fraction.

(D) Position weight matrix of known *Drosophila* motifs enriched upstream (-2000 bp to 0 bp relative to transcription start site) of RNAs enriched in each group. Associated transcription factors and *P*-values are indicated.

(E) Heatmap of 'molecular function' ontology enrichments associated with RNAs enriched in each fraction. Median-centered distributions of *P*-values (P) and Benjamini false discovery rate (F) are displayed. Color intensity is proportional to statistical significance.

Figure 8.5. *Slbp*$^{10/12}$ and *Slbp*$^{10/15}$ embryos phenocopy the transcriptomic alterations of *grp*^{fs1} mutants

(A) Transcriptomic validation of loss-of-function models. *UCSC* genome browser view of read coverage at the *grp*, *Slbp*, histone gene complex and *Gapdh1* loci in *wt*, *grp*^{fs1}, *Slbp*$^{10/12}$ and *Slbp*$^{10/15}$ embryos collected 0-180 min AEL.

(B) Venn diagram of genes exhibiting a 4-fold or greater increase (green) and decrease (red) in *Slbp*$^{10/12}$, *Slbp*$^{10/15}$ and *grp*^{fs1} 0-180 min AEL embryos relative to *wt*.

(C) Position weight matrix of consensus sequences of known *Drosophila* transcription factors enriched upstream of genes (-2000 bp to 0 bp relative to TSS) exhibiting a 4-fold or greater decrease in *grp*^{fs1}, *Slbp*$^{10/12}$ and *Slbp*$^{10/15}$ embryos relative to *wt.* Transcription factor names and associated *P*-values are indicated.

(D) Heatmap of 'molecular function' ontology enrichments associated with genes showing a 4-fold or greater decrease in *grp*^{fs1}, *Slbp*$^{10/12}$ and *Slbp*$^{10/15}$ mutant embryos relative to *wt*. Median-centered distributions of *P*-values (P) and Benjamini false discovery rate (F) are displayed.

(E) Simplex graph of the spatiotemporal distribution during embryogenesis for RNAs altered in *grp*^{fs1}, *Slbp*$^{10/12}$ and *Slbp*$^{10/15}$ embryos. Altered transcripts (4-fold decrease: dark red, 20-fold decrease: light red) were overlaid on the transcriptome-wide distributions (black).

Figure 8.6. Zygotic gene expression is impaired in *grp*^{fs1}, *Slbp*$^{10/12}$ and *Slbp*$^{10/15}$ embryos

(A) Boxplots of expression fold change in *grp*^{fs1} (top), *Slbp*$^{10/12}$ (middle) and *Slbp*$^{10/15}$ (bottom) mutant embryos relative to *wt* for groups of spatiotemporally asymmetric transcripts. Transcripts were selected from groups of abundant total ovary RNAs (1), RNAs enriched in total early syncyctial embryos (2), RNAs enriched in cytoplasmic extracts of blastoderm embryos (3), purely zygotic RNAs defined by *De Renzis et al*. (4), RNAs enriched in late blastoderm embryos (5) and RNAs enriched in nuclear extracts of blastoderm embryos (6). For each group, transcripts were ranked based on increasing cumulative read count and the 100 first transcripts breaching a threshold of 100 cumulative reads were selected. Median fold change values are displayed and global statistical differences of median fold change values were determined using a Kruskal-Wallis analysis of variance.

(B) Median-centered heatmap of expression fold change in *grp*^{fs1} (left column), *Slbp*$^{10/12}$ (center column) and *Slbp*$^{10/15}$ (right column) mutant embryos relative to *wt* the six groups of 100 RNAs defined above.

Fig. 8.7. Model of the relationship among zygotic genome activation, mRNA metabolism, and DNA replication checkpoint signaling during early Drosophila embryogenesis

During cleavage cycles, zygotic genome activation is influenced by pioneer transcription factors (e.g., Zelda) and by the titration of proteins that hinder transcription when in excess, such as histones potentially. The process of zygotic transcription induces replicative stress and activation of Chk1 signaling, which acts to slow down the cell cycle. If genotoxic stress is excessive, this triggers a developmental arrest controlled by Chk2, for which the effectors remain largely unknown. Chk2 signaling impairs robust zygotic genome activation and developmental progression. SLBP is phosphorylated by Chk2, a process that inhibits the normal function of SLBP in histone mRNA export. This mechanism links replication checkpoint signaling to the control of histone metabolism.

References

[1] O'Farrell PH, Stumpff J, Su TT. Embryonic cleavage cycles: how is a mouse like a fly? Curr Biol. 2004;14:R35-45.

[2] Yuan K, Seller CA, Shermoen AW, O'Farrell PH. Timing the Drosophila Mid-Blastula Transition: A Cell Cycle-Centered View. Trends in genetics : TIG. 2016;32:496-507.

[3] Leichsenring M, Maes J, Mossner R, Driever W, Onichtchouk D. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. Science. 2013;341:1005-9.

[4] Tadros W, Lipshitz HD. The maternal-to-zygotic transition: a play in two acts. Development. 2009;136:3033-42.

[5] Amodeo AA, Jukam D, Straight AF, Skotheim JM. Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition. Proceedings of the National Academy of Sciences of the United States of America. 2015;112:E1086-95.

[6] Sanchez R, Marzluff WF. The stem-loop binding protein is required for efficient translation of histone mRNA in vivo and in vitro. Mol Cell Biol. 2002;22:7093-104.

[7] Dominski Z, Yang XC, Raska CS, Santiago C, Borchers CH, Duronio RJ, et al. 3' end processing of Drosophila melanogaster histone pre-mRNAs: requirement for phosphorylated Drosophila stem-loop binding protein and coevolution of the histone pre-mRNA processing system. Mol Cell Biol. 2002;22:6648-60.

[8] Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet. 2008;9:843-54.

[9] Jackson SP, Bartek J. The DNA-damage response in human biology and disease. Nature. 2009;461:1071-8.

[10] Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. Mol Cell. 2010;40:179-204.

[11] Sibon OC, Stevenson VA, Theurkauf WE. DNA-replication checkpoint control at the Drosophila midblastula transition. Nature. 1997;388:93-7.

[12] Sibon OC, Laurencon A, Hawley R, Theurkauf WE. The Drosophila ATM homologue Mei-41 has an essential checkpoint function at the midblastula transition. Curr Biol. 1999;9:302-12.

[13] Takai H, Tominaga K, Motoyama N, Minamishima YA, Nagahama H, Tsukiyama T, et al. Aberrant cell cycle checkpoint function and early embryonic death in Chk1(-/-) mice. Genes Dev. 2000;14:1439-47.
[14] Liu Q, Guntuku S, Cui XS, Matsuoka S, Cortez D, Tamai K, et al. Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. Genes Dev. 2000;14:1448-59.
[15] Brown EJ, Baltimore D. ATR disruption leads to chromosomal fragmentation and early embryonic lethality. Genes & Development. 2000;14:397-402.
[16] Blythe SA, Wieschaus EF. Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. Cell. 2015;160:1169-81.
[17] Takada S, Kelkar A, Theurkauf WE. Drosophila checkpoint kinase 2 couples centrosome function and spindle assembly to genomic integrity. Cell. 2003;113:87-99.
[18] Iampietro C, Bergalet J, Wang X, Cody NA, Chin A, Lefebvre FA, et al. Developmentally regulated elimination of damaged nuclei involves a Chk2-dependent mechanism of mRNA nuclear retention. Dev Cell. 2014;29:468-81.
[19] Lanzotti DJ, Kaygun H, Yang X, Duronio RJ, Marzluff WF. Developmental control of histone mRNA and dSLBP synthesis during Drosophila embryogenesis and the role of dSLBP in histone mRNA 3' end processing in vivo. Mol Cell Biol. 2002;22:2267-82.
[20] Sullivan W, Ashburner M, Hawley RS. Drosophila protocols. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 2000.
[21] Lefebvre FA, Bouvrette LPB, Bergalet J, Lécuyer E. Data for the Generation of RNA Spatiotemporal Distributions and Interpretation of Chk1 and SLBP Protein Depletion Phenotypes during Drosophila Embryogenesis. Data in Brief. 2016;In press:In press.
[22] Glass HI, De Garreta AC. The quantitative limitations of exponential curve fitting. Phys Med Biol. 1971;16:119-30.
[23] Huang ZN. [Using serial regression to fit the exponential curve]. Zhonghua Yu Fang Yi Xue Za Zhi. 1988;22:200-2.
[24] Wang ET, Cody NA, Jog S, Biancolella M, Wang TT, Treacy DJ, et al. Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. Cell. 2012;150:710-24.

[25] Lazzaretti D, Veith K, Kramer K, Basquin C, Urlaub H, Irion U, et al. The bicoid mRNA localization factor Exuperantia is an RNA-binding pseudonuclease. Nat Struct Mol Biol. 2016;23:705-13.
[26] Song X, Wong MD, Kawase E, Xi R, Ding BC, McCarthy JJ, et al. Bmp signals from niche cells directly repress transcription of a differentiation-promoting gene, bag of marbles, in germline stem cells in the Drosophila ovary. Development. 2004;131:1353-64.
[27] Gergen JP, Butler BA. Isolation of the Drosophila segmentation gene runt and analysis of its expression during embryogenesis. Genes Dev. 1988;2:1179-93.
[28] Lecuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, et al. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell. 2007;131:174-87.
[29] De Renzis S, Elemento O, Tavazoie S, Wieschaus EF. Unmasking Activation of the Zygotic Genome Using Chromosomal Deletions in the Drosophila Embryo. PLoS Biol. 2007;5:e117.
[30] Shermoen AW, O'Farrell PH. Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. Cell. 1991;67:303-10.
[31] Heyn P, Kircher M, Dahl A, Kelso J, Tomancak P, Kalinka AT, et al. The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell reports. 2014;6:285-92.
[32] Wood V, Harris MA, McDowall MD, Rutherford K, Vaughan BW, Staines DM, et al. PomBase: a comprehensive online resource for fission yeast. Nucleic acids research. 2012;40:D695-9.
[33] Moreno E, Morata G. Caudal is the Hox gene that specifies the most posterior Drosophile segment. Nature. 1999;400:873-7.
[34] Liang HL, Nien CY, Liu HY, Metzstein MM, Kirov N, Rushlow C. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature. 2008;456:400-3.
[35] Nien CY, Liang HL, Butcher S, Sun Y, Fu S, Gocha T, et al. Temporal coordination of gene networks by Zelda in the early Drosophila embryo. PLoS genetics. 2011;7:e1002339.
[36] Ohno K, Hirose F, Sakaguchi K, Nishida Y, Matsukage A. Transcriptional regulation of the Drosophila CycA gene by the DNA replication-related element (DRE) and DRE binding factor (DREF). Nucleic acids research. 1996;24:3942-6.

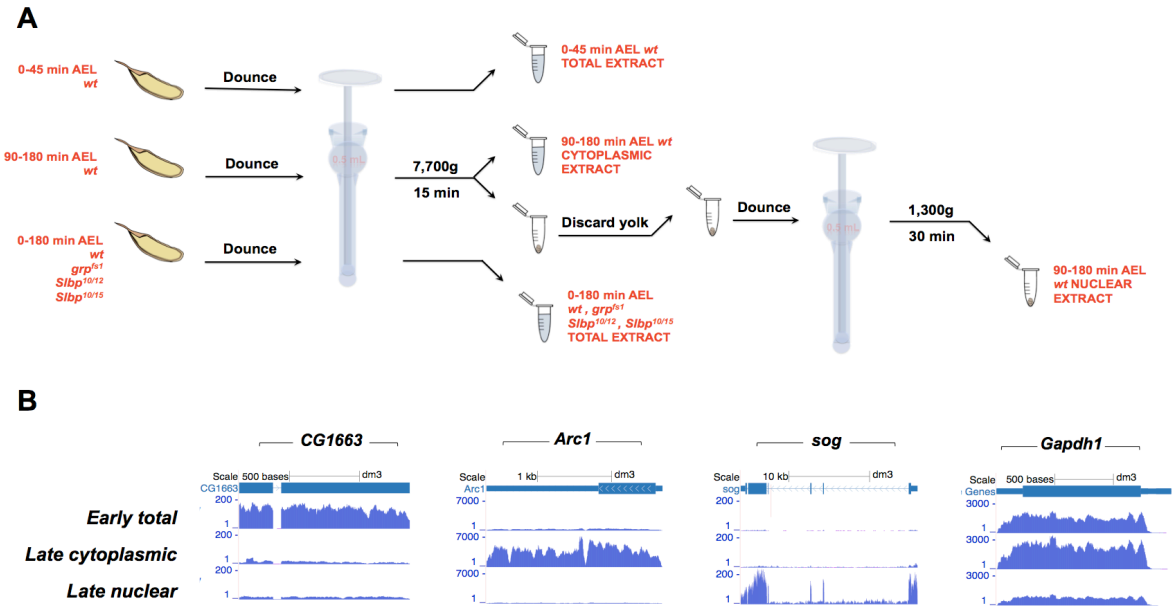
[37] Sullivan E, Santiago C, Parker ED, Dominski Z, Yang X, Lanzotti DJ, et al. Drosophila stem loop binding protein coordinates accumulation of mature histone mRNA with cell cycle progression. Genes Dev. 2001;15:173-87.
[38] Roth S, Stein D, Nusslein-Volhard C. A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the Drosophila embryo. Cell. 1989;59:1189-202.
[39] George AA, Sharma M, Singh BN, Sahoo NC, Rao KV. Transcription regulation from a TATA and INR-less promoter: spatial segregation of promoter function. The EMBO journal. 2006;25:811-21.
[40] De Gregorio E, Chiariotti L, Di Nocera PP. The overlap of Inr and TATA elements sets the use of alternative transcriptional start sites in the mouse galectin-1 gene promoter. Gene. 2001;268:215-23.
[41] Walter J, Newport JW. Regulation of replicon size in Xenopus egg extracts. Science. 1997;275:993-5.
[42] Amodeo AA, Skotheim JM. Cell-Size Control. Cold Spring Harbor perspectives in biology. 2016;8:a019083.
[43] Edgar BA, Schubiger G. Parameters controlling transcriptional activation during early Drosophila development. Cell. 1986;44:871-7.
[44] Sullivan KD, Mullen TE, Marzluff WF, Wagner EJ. Knockdown of SLBP results in nuclear retention of histone mRNA. Rna. 2009;15:459-72.
[45] de Vries HI, Uyetake L, Lemstra W, Brunsting JF, Su TT, Kampinga HH, et al. Grp/DChk1 is required for G2-M checkpoint activation in Drosophila S2 cells, whereas Dmnk/DChk2 is dispensable. J Cell Sci. 2005;118:1833-42.
[46] Canman CE. Replication checkpoint: preventing mitotic catastrophe. Curr Biol. 2001;11:R121-4.
[47] Gunesdogan U, Jackle H, Herzig A. A genetic system to assess in vivo the functions of histones and histone modifications in higher eukaryotes. EMBO Rep. 2010;11:772-6.
[48] Brodsky MH, Weinert BT, Tsang G, Rong YS, McGinnis NM, Golic KG, et al. Drosophila melanogaster MNK/Chk2 and p53 regulate multiple DNA repair and apoptotic pathways following DNA damage. Mol Cell Biol. 2004;24:1219-31.

[49] Legendre F, Cody N, Iampietro C, Bergalet J, Lefebvre FA, Moquin-Beaudry G, et al. Whole mount RNA fluorescent in situ hybridization of Drosophila embryos. J Vis Exp. 2013:e50057.

[50] Lecuyer E, Necakov AS, Caceres L, Krause HM. High-resolution fluorescent *in situ* hybridization of *Drosophila* embryos and tissues. CSH Protocols. 2008:doi:10.1101/pdb.prot5019.

[51] Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38:576-89.

[52] Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 2003;4:P3.



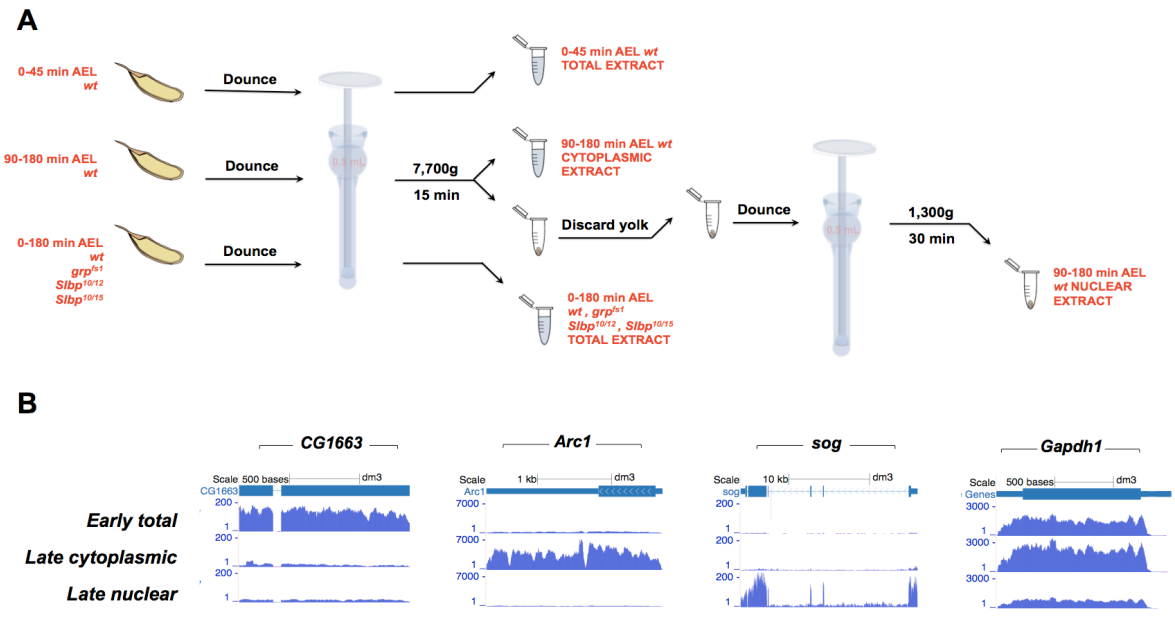


Figure 8. 1. Experimental pipeline and transcriptomic validation

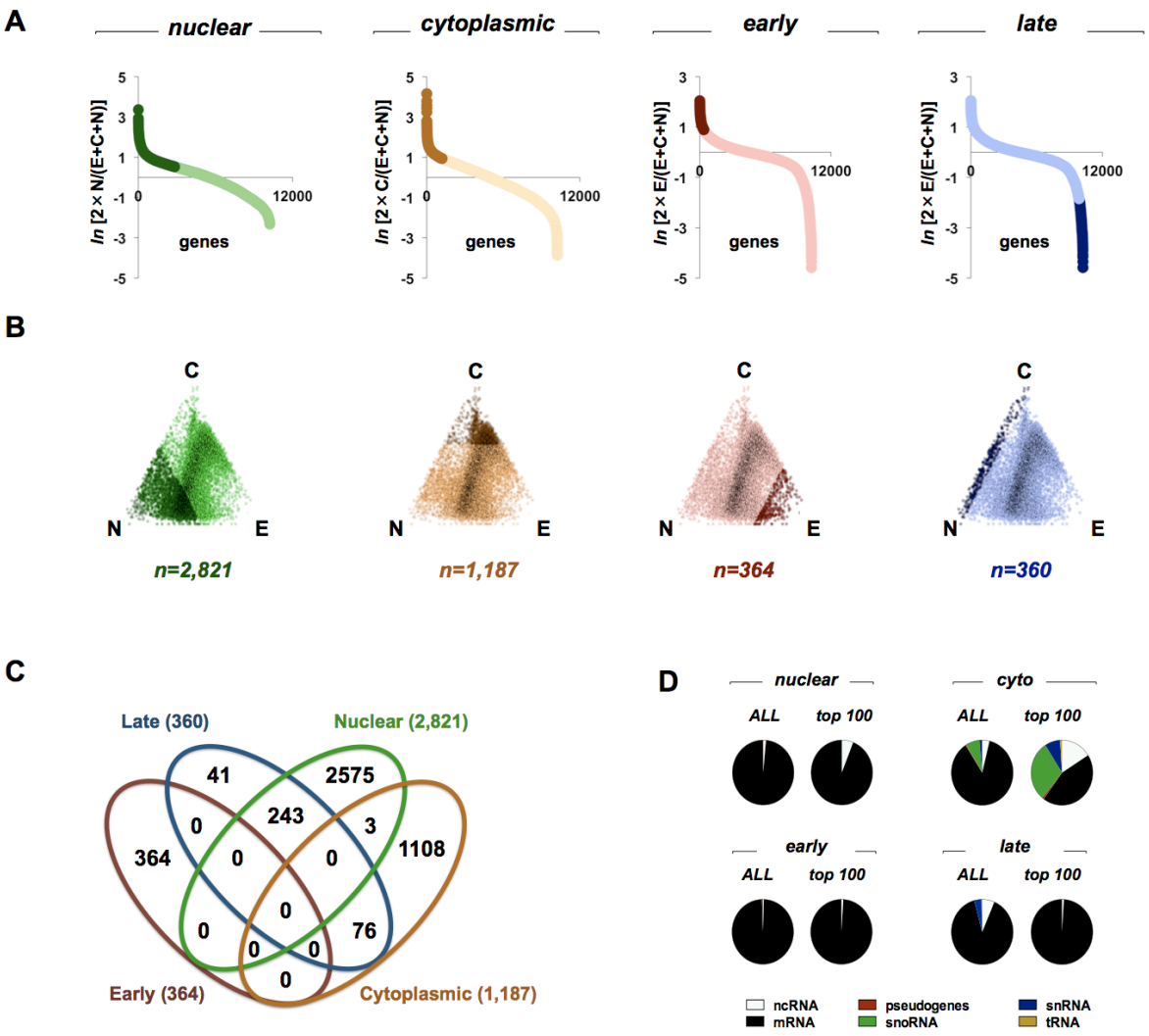


Figure 8. 2. Spatiotemporal dissection of the early *Drosophila* embryo transcriptome

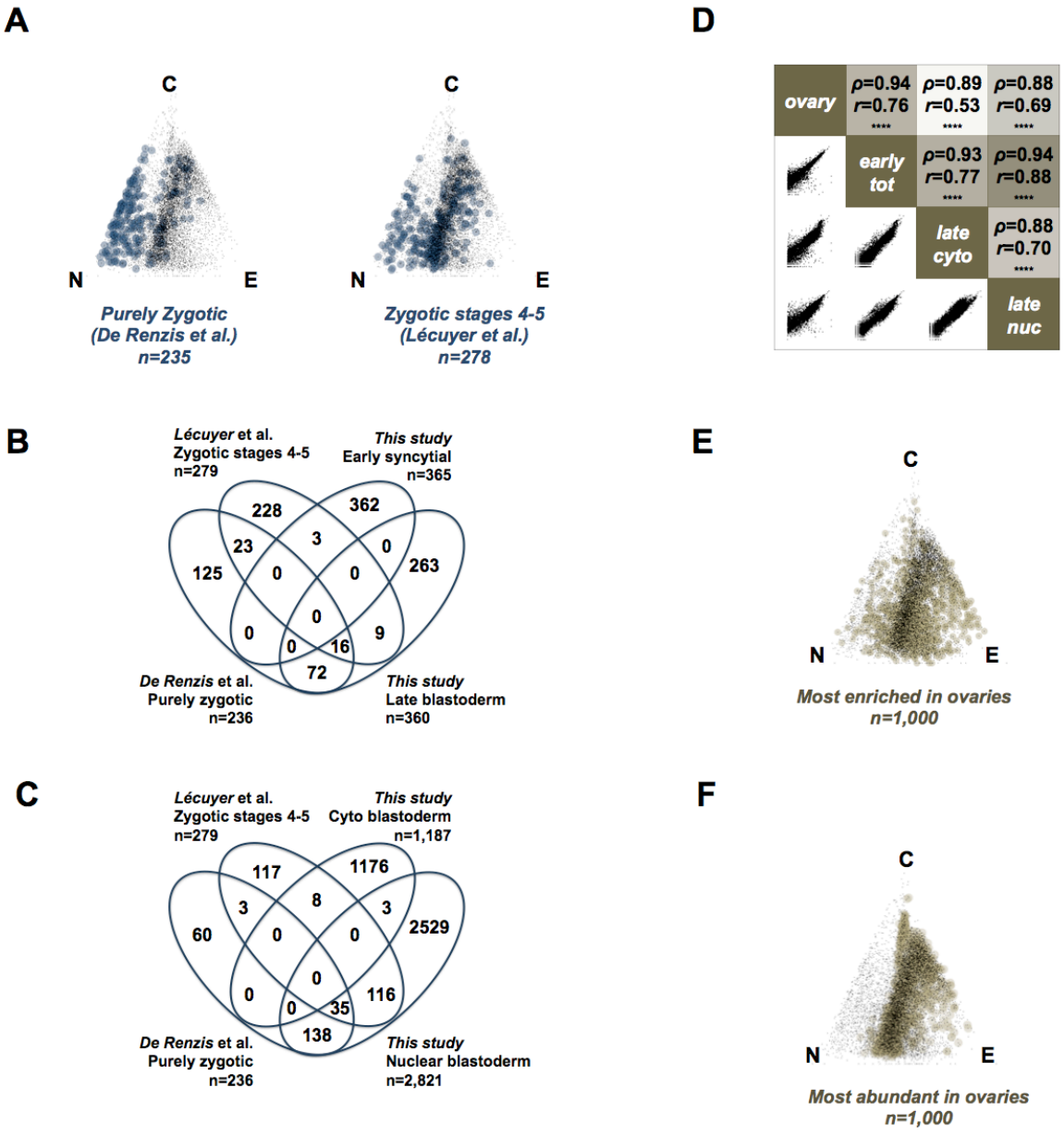


Figure 8. 3. Spatiotemporal distributions of zygotic and ovarian transcripts during embryogenesis

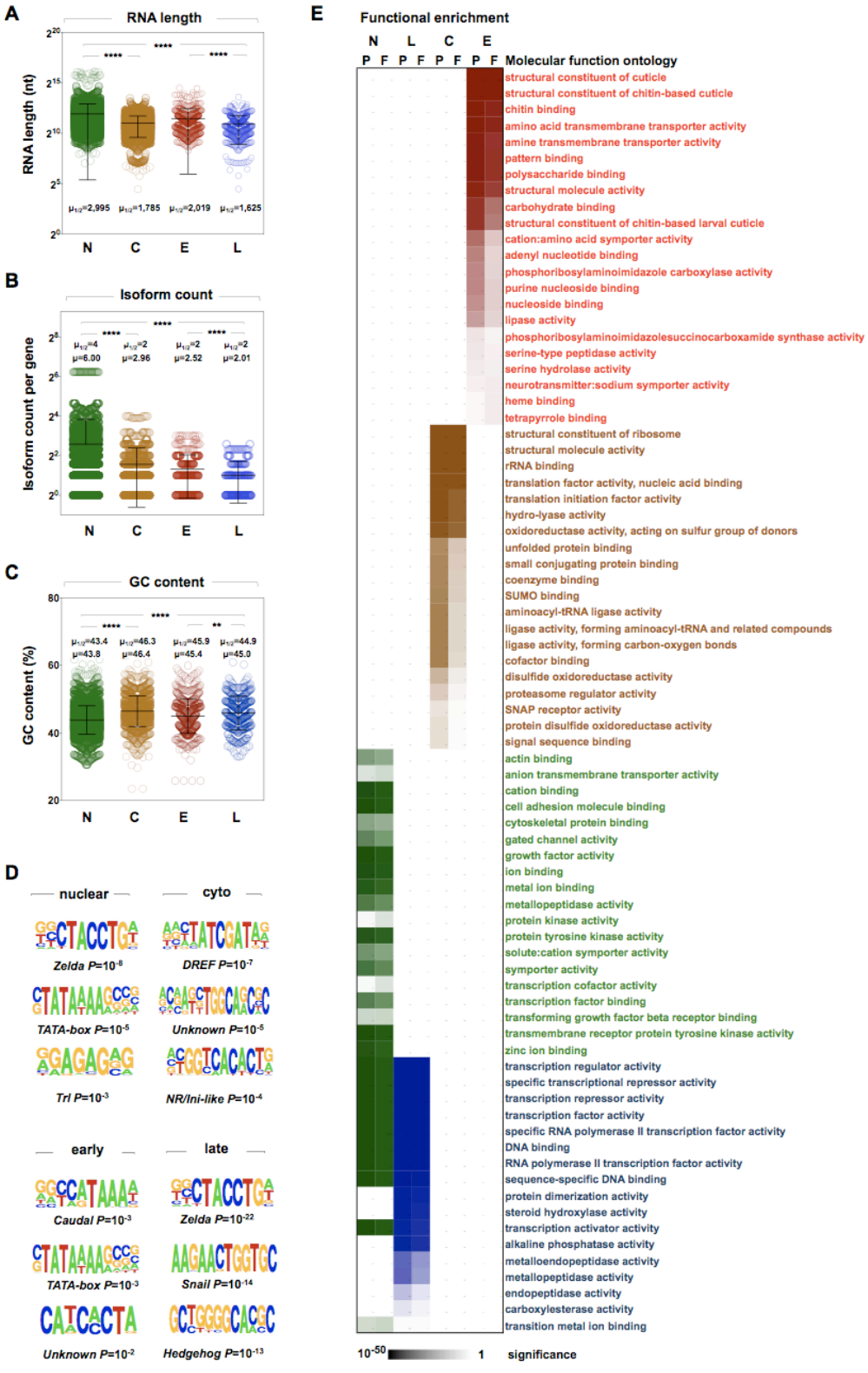


Figure 8. 4. Properties of transcripts displaying asymmetric spatiotemporal profiles

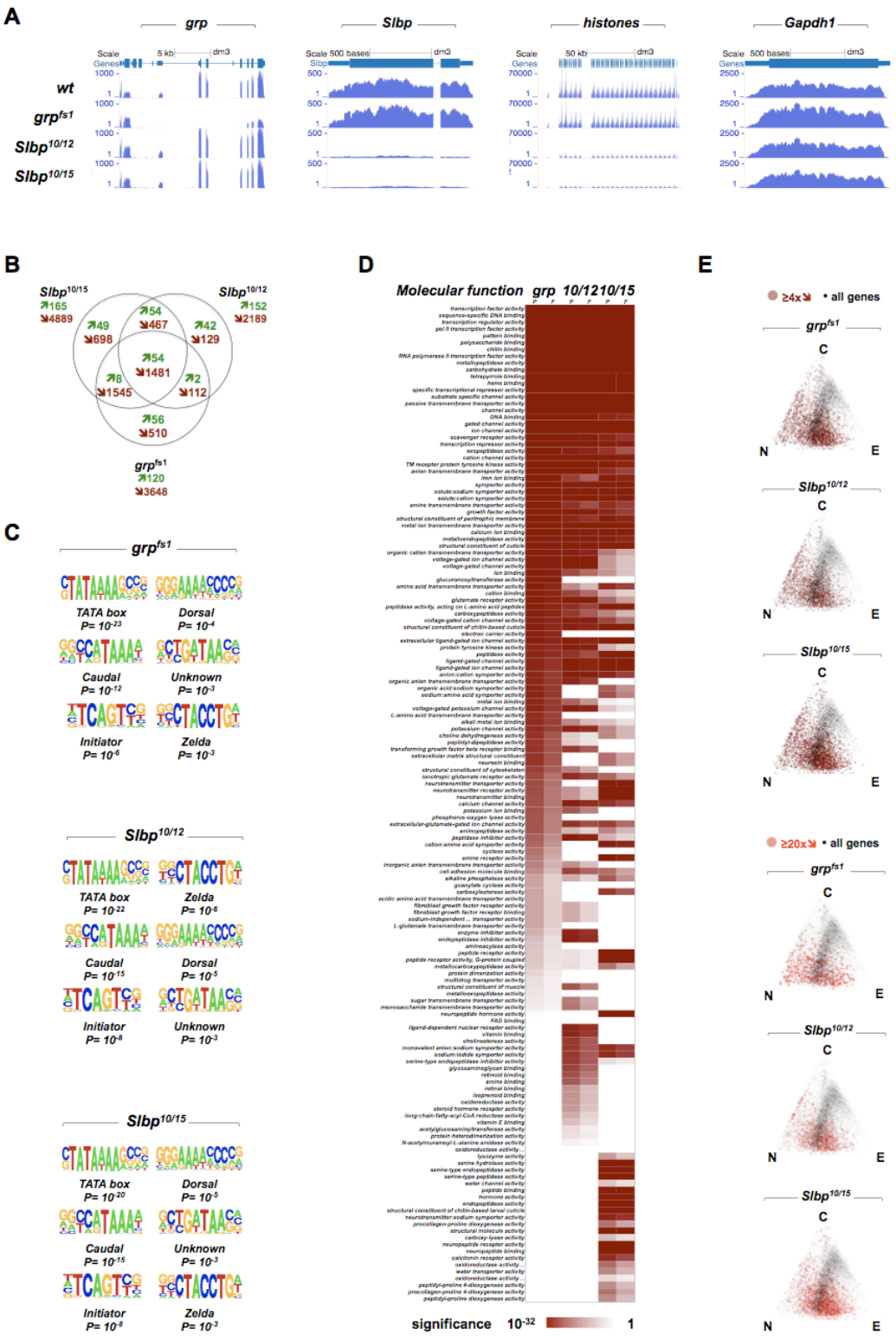


Figure 8. 5. ***Slbp*$^{10/12}$ and *Slbp*$^{10/15}$ embryos phenocopy the transcriptomic alterations of *grp*^{fs1} mutants**

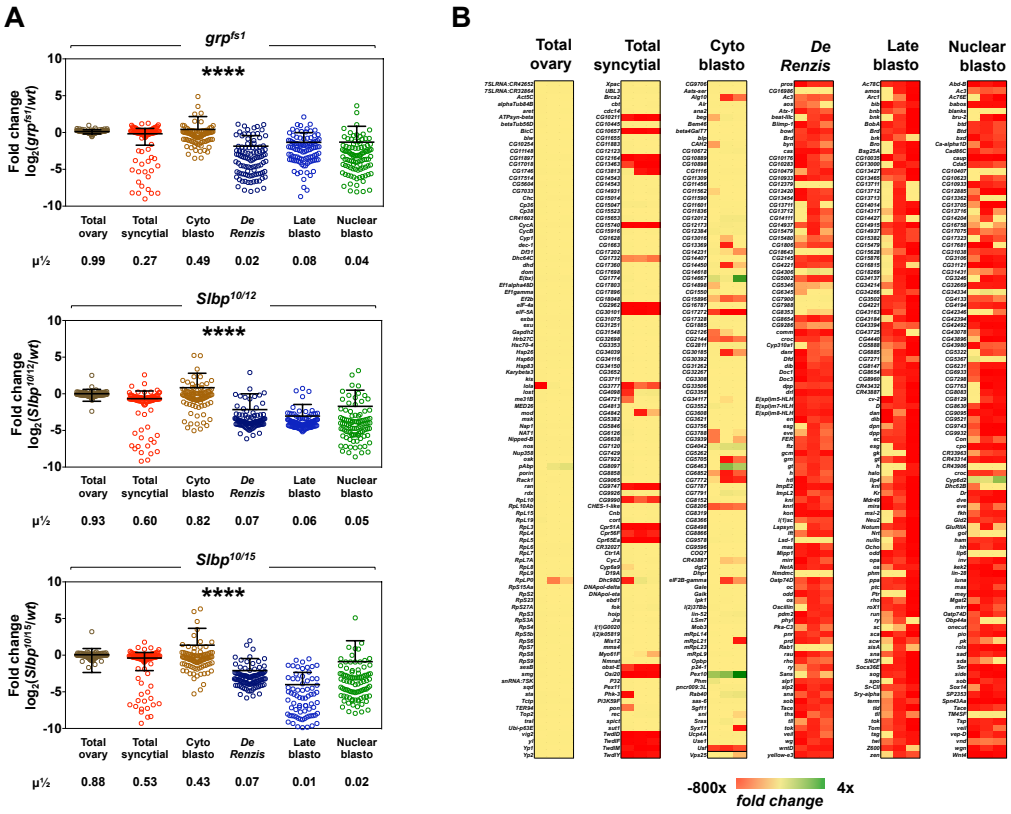


Figure 8. 6. Zygotic gene expression is impaired in *grp*^{fs1}, *Slbp*$^{10/12}$ and *Slbp*$^{10/15}$ embryos

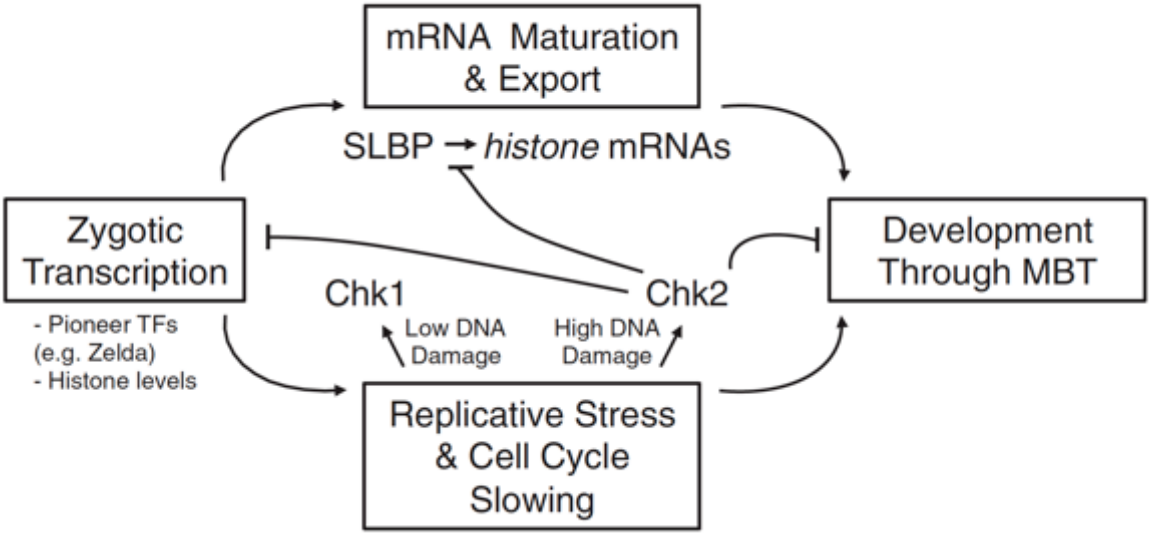


Figure 8. 7. Model of the relationship among zygotic genome activation, mRNA metabolism, and DNA replication checkpoint signaling during early Drosophila embryogenesis

Table 8. I. Read counts and properties

Sample	Reads ($\times 10^6$)	Duplicates (%)	Genes count (≥5 reads)
wt 0-45 min AEL	32.8	27	9,556
*w*t 0-45 min AEL	31.1	24	
wt 90-180 min AEL cyto	41.4	29	9,362
wt 90-180 min AEL cyto	36.2	27	
wt 90-180 min AEL nuclei	38.6	53	10,050
wt 90-180 min AEL nuclei	37.9	36	
wt 0-180 min AEL	24.3	25	9,673
wt 0-180 min AEL	26.4	27	
grp^{fs1} 0-180 min AEL	28.9	23	7,834
grp^{fs1} 0-180 min AEL	28.4	23	
$Slbp^{10/12}$ 0-180 min AEL	31.9	30	7,474
$Slbp^{10/12}$ 0-180 min AEL	30.3	22	
$Slbp^{10/15}$ 0-180 min AEL	25.4	24	7,562
$Slbp^{10/15}$ 0-180 min AEL	28.3	24	

Préface au chapitre 9

Ce chapitre est présenté sous la forme d'un article de recherche non-publié, qui sera soumis après le dépôt de la thèse au journal *Development*.

L'article documente via diverses techniques l'expression de courts transcrits antisens issus du locus des gènes d'histone pendant l'embryogenèse de la Drosophile. Nous démontrons que ces transcrits sont exprimés précocement par le génome zygotique d'une manière qui dépend du facteur SLBP. Chez l'embryon, ces transcrits sont exportés au cytoplasme et miment la distribution subcellulaire de leurs ARNm complémentaires, ce qui suggère qu'ils pourraient former ensemble des ARN double-brin, de potentiels précurseurs des petits ARN interférents endogènes. Dans la poursuite de cette hypothèse, nous démontrons que des petits ARN provenant du locus des histones sont associés à Argonaute-2, effecteur important des voies des petits ARNs. Finalement, nous démontrons que la perte d'Argonaure-2 s'accompagne d'une dérepression des ARNm d'histone. Ces observations débouchent sur un modèle : l'expression strictement zygotique des ARNs d'histone antisens contribue à la dégradation des ARNm d'histone de contribution maternelle via Argoaute-2.

Dans le contexte de la présente thèse, les résultats exposés dans cet article font suite au déploiement chez l'embryon de mouche de techniques de fractionnement subcellulaire couplé à l'analyse transcriptomique pour mieux évaluer la localisation des ARNs.

J'ai écrit l'intégralité du manuscrit. J'ai effectué les expériences et les analyses liées aux vésicules extracellulaires ainsi que la préparation des figures présentées dans cette publication. Louis Philip Benoit Bouvrette, étudiant au PhD au laboratoire, a contribué aux analyses de séquençage à haut débit. Dr Éric Lécuyer, mon superviseur, a contribué à l'étude en amassant les financements requis et en relisant le texte pour y suggérer diverses améliorations. Éric a également complété certaines des expérienced, dont certaines des hybridations *in situ*, pendant son stage post-doctoral.

Chapitre 9 : Developmentally regulated and SLBP-dependent antisense transcription of histone genes during Drosophila embryogenesis (Article #7)

Chapitre 9 : Article #7

Developmentally regulated and SLBP-dependent antisense transcription of *histone* genes during *Drosophila* embryogenesis

Fabio Alexis Lefebvre[1,2],
Louis Philip Benoit Bouvrette[1,2], Éric Lécuyer[1,2,3,4]

1- Institut de Recherches Cliniques de Montréal (IRCM)
Montréal, Québec, Canada

2- Département de Biochimie
Université de Montréal, Montréal, Québec, Canada

3- Division of Experimental Medicine
McGill University, Montréal, Québec, Canada

4- Address correspondences to: Dr. Eric Lécuyer
IRCM, RNA Biology Laboratory
110 Avenue des Pins, Ouest
Montréal, Québec, Canada
H2W 1R7
Tel: 514-987-5646, Fax: 514-987-5752
Email: Eric.Lecuyer@ircm.qc.ca

The datasets described here can be accessed at GEO under the accession numbers GSE89001 and at modENCODE using the accession numbers ENCSR283YJX, ENCSR053CWY, ENCSR473SBP, ENCSR709PEM, ENCSR302HSE, ENCSR432GTP, ENCSR197ZHM and ENCSR008VVB

Abstract

Natural antisense transcripts (NATs) form a widespread class of RNAs involved in the regulation of eukaryotic gene expression. In *Drosophila*, replication-dependent *histone* (*his*) mRNAs are encoded by a large complex consisting of tandemly arrayed gene units. *His* mRNAs lack a polyA tail and their processing, maturation and translation require the RNA-binding protein SLBP. Using *in situ* hybridization and deep-sequencing, we identified abundant and precocious zygotic expression of NATs mapped to all five *his* genes in early *Drosophila* embryos. Like *his* mRNAs, *his* NATs lack a polyA tail and their expression requires dSLBP, as shown through comparative transcriptomic analyses of biallelic mutant embryos. Through comparative transcriptomics, we show that *his* NATs are highly expressed in early embryos, but scarce in whole ovaries and in the larva cell line D17. A subcellular fractionation approach coupled to deep-sequencing revealed that *his* NATs mimic the nucleo-cytoplasmic distribution properties of their corresponding mRNAs in embryos. Consistent with cytoplasmic processing by Dicer, embryonic *his* NATs encompassed the genomic coordinates of several Argonaute 2 (AGO2)-associated small RNAs. In addition, AGO2 mutant embryos display a weak but significant increase in the levels of *his2a, his2b and his4* mRNAs, hinting at a contribution of the endo-siRNA pathway in the posttranscriptional silencing of *his* genes during early embryogenesis. Our work identifies a novel group of precociously expressed NATs, defines their transcriptional regulation and spatiotemporal landscape and suggests their prospective role as putative precursors of small RNAs. We propose a working model involving promoter-proximal downstream initiation that results in convergent transcription of *his* genes, possibly promoting the clearance of maternal *his* mRNAs via Ago2-dependent posttranscriptional silencing.

Highlights

- ***Histone* antisense transcripts are developmentally-regulated and strictly provided by the zygotic genome**
- **Antisense transcripts are SLBP-dependent and lack a polyA tail**
- **Small endogenous RNAs associated to AGO2 map to *histone* antisense signatures**
- **AGO2 depletion leads to a weak but significant upregulation of three *histone* mRNAs, suggestive of a role of RNAi pathways in modulating their expression**

Introduction

In the post-genomic era, pervasive transcription has emerged as a conserved trait among higher eukaryotes, challenging conventional views of narrowly delimited gene expression domains (1). *Cis*-natural antisense transcripts (*cis*-NATs) refer to overlapping transcriptional units encoded on the reverse strand of expressed loci. NATs have been identified in a large fraction of metazoan protein-coding genes, with estimates of ~30% of all human genes according to diverse sources (2, 3). NATs can regulate their conjugated sense transcripts through various mechanisms, including polymerase collision and via the formation of RNA duplexes upon complementary base annealing, which have been associated with post-transcriptional gene silencing and RNA editing (4-6).

During *Drosophila* oogenesis, gene products (RNAs and proteins) synthesized in polyploid nurse cells are trafficked through ring canals and deposited in the developing oocyte to sustain proliferation during embryonic cleavage cycles (7-9). Indeed, prior to the midblastula transition (MBT), which occurs at the 13th nuclear cycle (NC), the zygotic genome is largely transcriptionally silent and most RNA molecules found in the embryo result from maternal deposition (10-12). Prior to this stage, precocious transcription of certain genes is observed, such as *histone* (*his*) mRNAs and certain repetitive element loci (13) (14), however zygotic gene transcription is largely stochastic and frequently abortive (15). Canonical *his* genes are replication-dependent factors and their expression is tightly coordinated with cell cycle progression, which entails a set of specific transcriptional and posttranscriptional regulatory mechanisms. To enable the rapid expression of abundant *His* stores at the end of S-phase, canonical *his* genes have undergone extensive amplification through evolution. In *Drosophila*, *his1*, *his2a*, *his2b*, *his3* and *his4* are organized as a dense cluster of approximately 5 kb called the histone gene unit (*his*-GU) and approximately 100 copies of the *his*-GU are tandemly arrayed on chromosome 2L (16, 17). The core histones *his2a/b* and *his3/4* are encoded in a head-to-head orientation and share a single core promoter per pair of genes(18) (**Figure 9.1A**). In metazoans, canonical *his* mRNAs lack a 3'-end polyA tail and therefore require the regulation of specific factors including Sm-like proteins, the U7 small nuclear (sn)RNA and the Stem-Loop Binding

Protein/SLBP(19). Several of these proteins congregate at *his* loci, forming nuclear organelles called Histone Locus Bodies/HLBs(20). The emergence of HLBs during embryogenesis occurs around NC11 and broadly coincides with prominent zygotic *his* mRNA synthesis. Insertion of a single sequence found within the *his3-his4* dual promoter can direct the assembly of an ectopic proto-HLB and the bidirectional transcription of this sequence promotes HLB formation (20, 21). Non-coding transcription of *his* loci has also been documented: *HIS-1* NATs have been identified in the protozoan *Leishmania major*(22) and testis-specific *his3* NATs have been described in *Drosophila* (23).

Here, we investigated the spatiotemporal expression profile and regulatory features of *Drosophila his* NATs during development. Comparative FISH and transcriptomic studies revealed that all five canonical *his* genes encode NATs in *Drosophila* embryos. Using pharmacological inhibition of Polymerase II (Pol II), we demonstrate that *his4* NATs are products of early zygotic genome transcription. In addition, *his* NATs are developmentally regulated: they are undetected in whole ovaries, highly expressed during early embryogenesis and poorly represented in the larval haltere disc cell line Dm-D17-c3 (D17). To determine the regulatory determinants of *his* NAT expression, we considered mutant embryos derived from mothers bearing two distinct loss-of-function allelic combinations of the *Slbp* gene. This approach revealed a sharp decrease in both *his* mRNA and NAT expression upon *Slbp* loss-of-function. Comparative analysis of sequencing libraries established from D17 cells through polyA enrichment and ribosomal depletion showed that both *his* mRNA and *his* NATs lack a polyA tail. This result was further confirmed through Northern blotting of polyA$^+$ and polyA$^-$ RNA extracts derived from *wt* embryos. Through a biochemical subcellular fractionation approach coupled to RNA-seq (24-26), we show that *his* NATs closely mirror the spatial distribution profile of cognate mRNAs in blastoderm embryos. By contrast, in D17 cells, *his* NATs were enriched in nuclei relative to *his* mRNAs. In embryonic S2 cells, small RNAs derived from the *his*-GU associate with the RNA-induced silencing complex (RISC) catalytic factor, Argonaute 2/AGO2 (27). In addition, AGO2 depletion leads to increased expression levels of three *his* mRNAs and NATs. Together, this data suggests that *his* NATs result from downstream antisense transcription of core *his* genes, leading to a developmental regulation of *his* mRNAs through small RNA interference.

Results

Zygotic expression of *His2B* NAT precedes the midblastula transition

In previous work, we utilized highly-resolved fluorescence RNA *in situ* hybridization (FISH) to study *histone* mRNA expression properties in *Drosophila* embryos (13). To our surprise, when conducting control experiments using 'sense' coding-strand FISH probes, we observed strong and specific chromatin-associated punctate foci for all *histones* interrogated, hinting at the endogenous expression of *his* NATs. To follow up on this finding, we used FISH to perform side-by-side comparisons of the subcellular distributions of *his* mRNAs and NATs in staged 0-4h embryo collections, as shown for *his2b* (**Figure 9.1B**). Prior to the MBT (NC 1 to 13), a strong cytoplasmic signal was observed for *his2b* mRNA throughout the embryos, likely reflecting maternally deposited transcripts. In addition, weak chromatin-associated binuclear foci were noted on each nucleus. This pattern has been associated with HLBs (20, 28) and probably underlies nascent zygotic transcripts, which can be detected as punctate foci (29) (14). At the onset of cellularization, the 14th nuclear cycle, surface views showed intense HLB-like foci on cortical nuclei (**Figure 9.1B**). By contrast, the cytoplasmic signal intensity appeared weaker in NC 14 embryos than during earlier stages, consistent with the clearance of maternal mRNA populations, with persistent *his2b* mRNA expression in germline pole cells (**Figure 9.1B**) (13). By comparison, FISH of *his2b* NATs also identified strong HLB-like foci on syncytial (NC 7), yolk and cortical nuclei (NC 10), suggesting that the zygotic genome expresses *his2b* NATs before the MBT (**Figure 9.1C**). Notably, these foci were also observed on mitotic nuclei, indicating that their chromatin-associated binuclear presentation is maintained at all stages of the cell cycle. The intensity of the *his2b* NATs cytoplasmic signal was markedly weaker than for *his2b* mRNA, suggesting low maternal contribution. Similar chromatin-associated foci were observed when probing for *his1*, *his2a* and *his3* mRNAs and NATs by FISH in preblastoderm embryos (**Figure 9.1D**). Next, we used Northern blotting to further confirm the expression of *his* NATs, as shown for *his2b* in **Figure 9.1E** and **Figure 9.S1**. Increasing amounts of total RNA extracted from 0-4h *wt* embryo collections were loaded side by side to enable the detection of differentially expressed transcripts. Hybridization to an antisense *his2b* probe revealed a strong

signal at approximately 500 nt, corresponding to *his2b* mRNA. Hybridization to a sense *his2b* probe revealed a specific band of apparently similar electrophoretic mobility, consistent with a specific *his2b* NAT presenting substantial gene coverage of the *his2b* gene. To confirm that the signals observed by Northern blot and by FISH reflect the detection of antisense RNA and not the aberrant detection of intron-less *his* genomic copies, we submitted total RNA extracts from 0-2h embryos to RNAse A and DNAse I treatments followed by Northern blotting (**Figure 9.S2-9.S3**). The RNAse treatment lead to a complete loss of the *his4* mRNA imprint. By contrast, the DNAse treatment slightly dampened both the *his4* mRNA pattern and the rRNA band intensity visualized by ethidium bromide staining, probably reflecting non-specific nuclease activity. These results confirm that our probes detect the RNA population rather than the intron-less genomic *his* gene copies.

Next, we sought to confirm the zygotic origin of *his* NATs and treated *wt* embryo collections with the RNPII inhibitor α-amanitin to perform FISH targeting *his4* NATs (**Figure 9.1F**) (13). Chromatin-associated foci were absent in treated samples, consistent with a zygotic Polymerase II dependency, similar to our previous observations for *his* mRNA foci (Iamplietro 2014). By contrast, early embryos treated with the vehicle DMSO showed strong zygotic and polar body nuclear foci as early as stage 1, indicating that zygotic *his4* antisense transcription is initiated very early in embryogenesis, as we observed for *his* mRNAs.

His NATs are abundant in early embryos but poorly expressed in ovaries and D17 larval disc cells

We aimed to decipher the genomic coordinates of the transcriptional unit associated to *his2b* NAT and perform quantitative assessments of its expression. We used RNA-seq and analyzed strand-specific, paired-end datasets of rRNA-depleted extracts of *wt* embryos collected 0-180 min AEL (**Figure 9.2A-9.2B, Table 9.I**). Strong coding signatures were observed for all five *histone* genes, although *his2a*, *his2b* and *his4* were slightly more abundant than *his1* and *his3* (≅1.5-fold ; $P \leq 10^{-4}$). Consistent with FISH and Northern blot evidence, antisense read coverage extending approximately 150 bp was observed in the coding and 5' untranslated region (5'UTR) of the *his2b* gene, while much weaker coverage observed throughout the gene body. Moreover, we observed antisense expression of the other canonical *histone* genes. Indeed, a similar, albeit

weaker antisense signature was observed within the *his2a* gene, starting at the 5'-end of the gene and spanning approximately 75 bp. The *histone* duplex *his3-his4* displayed similar coverage, although the NATs traced at the 5'-end of these genes were less abundant and significantly longer, spanning approximately 200 bp ($P \leq 10^{-4}$). The linker *histone his1* gene showed evidence of an additional NAT of approximately 200 bp, mapped to the 3'-end of the gene (**Figure 9.2B**).

In absence of *his* NAT annotation, we relied on strand-specific maximal read coverage to perform comparative assessments of expression levels. *His2b* NAT (**Table 9.II**; 1,530 reads) had a significantly higher maximal coverage than, for example, *Gapdh1* mRNA (1,057 reads, t-test $P \leq 10^{-4}$). Of note, the "modENCODE temporal expression data" resource classifies *Gapdh1* expression level as "very high" in 0-2h embryos and as "high" in 2-4h embryos (30). To provide a global context for *his* NATs expression during development, we performed strand-specific RNA-seq on whole *Drosophila* ovary extracts and on the cell line Dm-D17-c3, which is derived from larval haltere discs (**Figure 9.2C-I** ; **Table 9.I**). While *his* mRNA levels were high all three models, *his* NATs were most abundant in embryos, scare in D17 cells and barely detected in ovaries (**Figure 9.2C-G**). We used maximal read counts to derive fold change values expressing antisense over coding levels for each *histone* gene in the ovarian, embryonic and larval models (**Figure 9.2G**). With over 1,000 reads in embryo libraries, maximal coverage of *his1* and *his2b* NATs respectively reached 10.0±0.2% and 9.5±0.1% of their complementary *his* mRNAs. By contrast, in D17 cells, NAT/mRNA read count ratios were much lower and averaged 1.7±0.3% for *his1* and 1.6±0.07% for *his2b*. In ovaries, NATs were even more severely under-represented: *his1* showed a NAT/mRNA ratio of 0.019±0.009%, while *his2b* NAT was undetected. In addition, we derived expression fold change across developmental tissues for both *his* mRNAs and *his* NATs (**Figure 9.2H-I**). While *his* mRNA levels were comparable in ovarian, embryonic and larval tissues (fold change < 5), embryonic levels of *his* NATs were much higher than their larval (8-70 fold) and ovarian counterparts (500-3,000 fold). Together, these comparative analyses strongly suggest that *his* NATs are developmentally regulated products that exhibit a peak in expression during early embryogenesis.

His NATs expression is SLBP-dependent

Canonical *his* mRNAs bear a 3'-end stem loop motif required for SLBP interaction, which promotes mRNA processing, cleavage, nuclear export and translation. Cross-linking immunoprecipitation assays coupled to RNA-seq (CLIP-seq) targeting SLBP have revealed its remarkable specificity for *his* mRNA binding (31, 32). Both maternal and zygotic expression of *Slbp* is crucial in *Drosophila* development and *Slbp* loss-of-function mutants carry a lethal phenotype, associated with a halt in development at the MBT (25). One hallmark of *Slbp* mutants is a dramatic loss of *his* mRNA expression (26). We sought to determine whether *his* NAT expression shows similar SLBP-dependency during *Drosophila* embryogenesis. We performed RNA-seq on 0-180 min AEL embryo collected from mothers bearing loss-of-function, trans-heterozygous *Slbp* mutations. To derive a robust portrait of the transcriptomic alterations resulting from *Slbp* loss-of-function, we profiled embryos from mothers carrying two distinct allelic combinations of mutations, $Slbp^{10/12}$ and $Slbp^{10/15}$. Consistent with a strong dSLBP loss-of-function, read coverage of the *Slbp* locus and the *his* gene complex was severely hampered in both mutant backgrounds, whereas *Gapdh1* expression was comparable in $Slbp^{10/12}$, $Slbp^{10/15}$ and *wt* embryos (**Figure 9.3A-C**). As reported in our previous work, *Slbp* mutations deeply compromised zygotic genome activation, leading to a 4-fold or greater down-regulation of over 2,000 RNAs (25, 26). In addition, *his* mRNAs and *his* NATs levels were strongly decreased and displayed a 6-fold to 50-fold drop in expression in $Slbp^{10/12}$ and $Slbp^{10/15}$ mutants (**Figure 9.3D-E, Table 9.III**). We conclude that *Slbp* function is required for *his* NATs expression, and that dSLBP depletion specifically compromises *his* NATs and *his* mRNA levels to a comparable extent.

A transcriptome-wide comparative approach reveals that *His* NATs lack a polyA tail

His mRNAs lack the polyA tail typical of other eukaryotic messengers and rely on SLBP binding for nuclear stability and translation initiation (19). Meanwhile, several RNPII-transcribed lncRNAs have been shown to contain a 3'-end polyA tail (33) and a polyA-dependent exosomal targeting mechanism has been linked to lncRNA turnover (34). We sought to determine whether *his* NATs are polyadenylated and developed a systematic transcriptomic approach to that end.

We assembled deep-sequencing libraries from total RNA extracts of D17 larva cells using two different methodologies: an affinity purification workflow based on polyA capture through an oligo(dT) primer and a ribosomal RNA depletion approach based on rRNA probe hybridization (24). Mapped transcripts were sorted as a function of their polyA index value, defined as the ratio of normalized read counts in the two libraries (polyA capture library read count/ribosomal depletion library read count). The resulting transcriptome-wide polyA index distribution can be described as an ordinal distribution of polyadenylation frequency.

To test our sorting approach, we generated histograms of polyA score by transcript biotype (**Figure 9.4A**). While mRNAs displayed a high polyA index (med=0.88 ± 0.61), small nuclear RNAs (med=0.0014 ± 0.009) and small nucleolar RNAs (med=0.0035 ± 0.02) showed lower figures, consistent with a rare polyadenylation. Collectively, *his* mRNA polyA index (med =0.0046 ± 0.007) was poor, comparable to small nucleolar RNA and in accordance with previous reports(19, 35). Meanwhile, the NATs mapped to *his2a*, *his2b* and *his4* were not traced in the polyA capture library and exclusively retrieved using the rRNA depletion methodology. PolyA index values were low for *his1* NAT (med =0.031 ± 0.002) and *his3* NAT (med =0.055 ± 0.08), suggesting that *his* NATs are not the targets of regulatory polyadenylation in D17 cells (**Figure 9.4B-C**). We sought to extend these findings in embryos using an alternative methodology. We submitted 5 µg of total RNA extracted from 0-180 min AEL *wt* embryos to a polyA affinity capture column system and retrieved both the $polyA^-$ and $polyA^+$ isolates. We then performed Northern blotting with probes targeting *his2b* mRNA and *his2b* NAT (**Figure 9.4D, Figure S9.4**). Both *his2b* mRNA and *his2b* NAT were detected in the total and $polyA^-$ samples, but not in the $polyA^+$ isolate. In accordance with the transcriptomic polyA index analysis performed in larva cells, these results indicate that *his* NATs are not polyadenylated in early *Drosophila* embryos.

His NATs mimic the subcellular distribution of cognate _His_ mRNAs in blastoderm embryos but not in larva cells

Diverse functional non-coding RNAs (lncRNAs) undergo nuclear export and are trafficked to specific cytoplasmic compartments, where they exert a wide range of biological functions (36-38). We reasoned that examining the subcellular distribution of *his* NATs may yield insights into potential biological functionality. Consequently, we adopted a subcellular fractionation approach coupled to RNA-seq to assess the spatial distribution of *his* mRNAs and *his* NATs in larva cells and in time-resolved embryo collections (24, 39) (**Figure 9.5**). The biochemical procedure we performed on D17 larva cells involves a hypotonic lysis, a centrifugation-based nuclear purification and a separation of cytosolic and endomembrane compartments (24, 40). RNA was extracted from each fraction and submitted to strand-specific RNA-seq analysis (24). We generated transcriptome-wide simplex representations of subcellular enrichment values based on normalized read counts in the nuclear (N), cytosolic (C) and endomembrane (M) fractions of D17 larval cells (40) (**Figure 9.5A**).

All five *his* mRNAs displayed similar subcellular distributions consisting of low nuclear occupancy and strong enrichments in the cytosolic and endomembrane compartments. By contrast, NATs mapped to *his1*, *his2a*, *his3* and *his4* were underrepresented in the endomembrane compartment and skewed towards the nuclear fraction. Interestingly, *his2b* NAT, the most abundant antisense *histone* transcript, showed a strong cytosolic pattern (3.8-fold enrichment), in sharp contrast to the other four *his* NATs we uncovered. We calculated transcriptome-wide cytosolic-to-nuclear enrichment ratios in D17 larva cells (**Figure 9.5B**). This analysis emphasized the cytosolic enrichment of all five *his* mRNAs and the contrasting nuclear accumulation of *his* NATs, except for *his2b* NAT, which mirrored the cytosolic-to-nuclear ratios of *His* mRNAs, possibly due to higher expression levels. We conclude that *his* NATs do not co-localize with cognate *his* mRNAs in larva cells and display nuclear enrichments, except for *his2b* NAT, which is predominantly cytosolic.

We next developed a tailored subcellular fractionation approach aimed at investigating the spatiotemporal distribution of RNAs in *Drosophila* embryos (24, 39). We performed a

nucleocytoplasmic separation protocol on collections of 90-180 min AEL *wt* blastoderm embryos. In addition, we analyzed our previously published total syncytial populations of early embryos collected 0-45 min AEL (24-26). RNA was extracted from each sample and submitted to strand-specific RNA-seq analysis. We generated transcriptome-wide simplex representations of spatiotemporal enrichment values based on normalized read counts in the nuclear blastoderm (N), cytoplasmic blastoderm (C) and total early syncytial (E) extracts of staged embryos (**Figure 9.5C**). All ten *his* mRNAs and *his* NATs were abundant in both early syncytial and late blastoderm embryos, in accordance with the results of FISH experiments. Interestingly, we noted strong variations in nucleocytoplasmic occupancy ratios among *his* mRNAs: *his3* was most abundant in nuclei, while *his1* showed a 3-fold cytoplasmic enrichment (**Figure 9.5D**). In addition, *his* NATs closely mirrored the nucleocytoplasmic occupancy ratio of their complementary mRNAs in blastoderm embryos. Therefore, our analyses show that *his* NATs undergo cytoplasmic export and are co-distributed with their cognate mRNAs in blastoderm embryos. These observations contrast with the subcellular distribution of *his* gene products in D17 larval cells, in which *his* NATs occupation appears essentially restricted to the nuclear compartment.

AGO2 associates with small RNAs overlapping with *His* NATs and regulates core *His* mRNA levels

Our findings of high cytoplasmic occupancy and co-segregation of *his* NATs and *his* mRNAs led us to hypothesize that *his* duplexes (dsRNA) may arise *in vivo*. *His* dsRNA could in turn feed into small RNA pathways, leading to Dicer recruitment and processing into small interfering RNAs, which have been termed “endogenous siRNAs” (esiRNAs). To explore this hypothesis, we set out to investigate the association of *his* transcripts with components of the RNA-induced silencing complex (RISC). We contrasted the genomic coordinates of *his* NATs with a previously published transcriptome-wide repertoire of AGO2-associated small RNAs determined in embryonic S2 cells by *Kawamura et al.* (27) (**Figure 9.6A**). We found that six distinct AGO2-associated small RNAs mapped to the *his*-GU, within the gene body of *his1*, *his2a*, *his4* and *his3*. In addition, we found that all six sequences overlapped with parts of the genes exhibiting bidirectional transcription in our analysis, notably near the 5’-end (*his1*, *his3*) or the 3’-end of the

genes (*his4*) (**Figure 9.6A**). This result suggests that *his* NATs may function as precursors of esiRNAs, potentially regulating *his* gene expression post-transcriptionally.

To evaluate the contribution of RNAi pathways on *his* mRNA levels, we used $Ago2^{51B}$ homozygous embryos, which bear a large deletion encompassing the start codon of the AGO2 gene. We also considered $Adar^{5G1}$ homozygous flies, which express amorphic alleles linked to a complete loss-of-function of the ADAR protein, a dsRBP involved in A-to-I editing. Homozygous $Ago2^{51B}$ and $Adar^{5G1}$ embryos were collected 0-180 min AEL, submitted to RNA extraction and a strand-specific deep sequencing analysis. Read coverage quantification at the *his*-GU revealed a small but significant up-regulation of *his2a* (P=0.05), *his2b* (P=0.02) and *his4* (P=0.02) mRNA levels in $Ago2^{51B}$ embryos relative to isochronic *wt* embryos (**Figure 9.6B**). Antisense read coverage was increased along the *his4* gene body (P=0.001), although the four other *his* NATs showed no significant change in $Ago2^{51B}$ mutants. By contrast, *his* mRNA and *his* NAT levels were not significantly affected in $Adar^{5G1}$ embryos, which showed sizeable cross-duplicate variations (**Figure 9.6C**). We used FISH to compare the subcellular distribution of *his4* mRNA in $Ago2^{51B}$ and *wt* embryos (**Figure 6D-E**). Surface view micrographs showed that *his4* coding transcripts form binuclear foci in NC10 $Ago2^{51B}$ nuclei, as described in *wt* embryos. This observation was confirmed in most-mitotic nuclei. In addition, the *his4* mRNA signal was prevalent in yolk nuclei of $Ago2^{51B}$ embryos, as reported in *wt* samples (**Figure 9.6E**). In conclusion, AGO2 depletion slightly increased the level of *his2a*, *his2b* and *his4* mRNAs without disrupting the localization properties of *his4* mRNA. Together, these results suggest that *his* NATs expression leads to the formation of esiRNAs that regulate the levels of *his2a*, *his2b* and *his4* mRNAs in an AGO2-dependent manner. In our analysis, these effects are relatively mild, possibly reflecting the modest decrease in AGO2 levels reported in of $Ago2^{51B}$ mutants (41-43), and it would be interesting to examine the levels of his mRNAs in a complete loss-of-function model of AGO2 and additional RNAi effectors.

Discussion

Here, we provide robust evidence of abundant bidirectional transcription of canonical *his* genes occurring during the first few syncytial cycles of embryonic development in *Drosophila*. *His2b* NATs form chromatin-associated foci which can be detected as early as the first nuclear cycle (NC 1) and persist in blastoderm and cellularized embryos. We demonstrate that zygotic antisense transcription occurs within all five canonical *his* genes, requires dSLBP and yields polyA-less products. Our tailored embryo fractionation approach coupled to RNA-seq showed that *his* NATs mimic the subcellular compartmentalization of their cognate mRNAs in embryos, but not in the larval haltere disc cells D17, where *his* NATs are largely retained in nuclei. Consistent with cytoplasmic processing by Dicer, we found that the genomic coordinates of *his* NATs encompass those of six embryonic AGO2-associated esiRNAs. Moreover, we observed a slight yet significant increase of *his2a*, *his2b* and *his4* mRNA levels in $Ago2^{51B}$ mutants, consistent with a role of esiRNA in the regulation of *his* gene expression. Our data suggests that SLBP-dependent, promoter-proximal downstream antisense transcription of core *his* genes occurs in early embryos and possibly contributes to the clearance of maternally deposited *his* mRNAs via AGO2-dependent RNA interference during the mother-to-zygotic transition (MZT) (**Figure 9.7**).

Downstream antisense transcription has been reported in diverse models and contexts. For example, promoter-proximal downstream antisense transcription is widespread in human breast cancer cells and has been documented within 2 kb of 28% of all transcription start sites (44). Similarly, a class of short NATs termed transcription start site-associated antisense RNAs (TSSasRNAs), has been described in mouse embryonic stem cells (mESCs). Much like the *his* NATs we describe, mESC TSSasRNAs lack a polyA tail and are associated with highly expressed RNPII-transcribed genes (45, 46). In *Drosophila* embryos, transcription surrounding the dual *his3-his4* promoter nucleates HLBs, subnuclear bodies specific to the *his* gene locus that contribute to the transcriptional regulation of *his* mRNAs. Indeed, a transgenic assay revealed

that ectopic insertions of the dual promoter was sufficient to drive the assembly of a proto-HLB, defined by the recruitment of the *his* mRNA processing factors FLASH and U7 snRNP (21). Interestingly, HLB maturation, surveyed through the engagement of the additional processing factors Mute and Lsm11, was observed ectopically exclusively after transcriptional activation of the zygotic genome. In turn, mature HLB assembly activated the expression of the proximal *his2a-his2b* module. Thus, activation of *his* mRNA zygotic expression is a step-wise process involving the recognition of a DNA element within the *his3-his4* promoter and an initial wave of transcription, which then recruits additional regulatory elements to *his*-GUs.

Our results indicate that *his* NATs levels are strongly compromised by dSLBP loss-of-function, which may appear surprising considering that *his* NATs don't encompass the 3'-end stem loop (SL) sequence of *his* mRNAs bound by SLBP. While SLBP binding is known to stabilize mature *his* mRNAs, SLBP is also a key component of HLBs and contributes to an endonucleolytic cleavage during *his* pre-mRNA processing, as demonstrated in human cells (47). Indeed, *his* pre-mRNAs are transcribed as long precursors that encompass both the SL sequence and a purine-rich *histone* downstream element (HDE), which anneals to a complementary sequence of U7 snRNA. Interestingly, the recognition of the HDE by the U7 snRNP requires the SLBP-dependent rearrangement of *his* mRNA 3'UTR that opens the HDE. The downregulation of *his* NATs following dSLBP loss-of-function thus suggests that *his* NATs are linked to the transcriptional activation of *his*-GUs and that their expression is tied to that of *his* pre-mRNAs. Since *his* NATs reads largely map to the 5'-end of core *his* genes, they don't appear to correspond to remnants of 3'-end pre-mRNA processing.

Three properties of the *his* NATs we identified suggest they could contribute to HLB maturation: (1) Unlike *his* mRNAs, *his* NATs are strictly zygotic products lacking a maternal contribution, and their transcription could modulate zygotic chromatin at *his*-GU ; (2) *his* NATs are developmentally regulated and their expression peaks in early embryogenesis, when HLBs are established ; (3) DNAse I sensitivity data suggests that the loci encoding *his* NATs exhibit high chromatin accessibility as early as NC5, which is consistent with enhancer-like regulatory features. Thus, *his* NATs could appear after the establishment of proto-HLBs in early embryogenesis, as the first zygotic *his* pre-mRNAs are being processed in an SLBP-dependent

manner. They could then serve as templates to recruit secondary processing factors such as Mute and Lsm11 during HLB maturation. Such hypothetic recruitment roles could be explored through a combination of *in vitro* RNA pulldowns and transgenic insertions of transcriptionally active and inactive *his* NATs sequences to score ectopic HLB nucleation.

In addition to putative roles in HLB assembly, our analysis of $Ago2^{51B}$ mutants suggests that *his* NATs could contribute to the regulated degradation of maternally deposited *his* mRNAs. Maternal clearance is a major event of the MZT that results in the selective destabilization of approximately 35% of maternally-deposited transcripts by the MBT (48). This process is multifaceted: it relies both on maternal pathways that operate through RBPs including Smaug (SMG) and Pumilio (PUM) and on zygotic inputs, involving the expression of ZGA-dependent regulators such as members of the miR-309 family (15). Although the prevalence of zygotic cues in guiding maternal clearance remains unclear in *Drosophila*, zygotic determinants are dominant in the destabilization of Zebrafish maternal transcripts, where miR-430 cumulates hundreds of maternal targets (49) (50). In *Drosophila*, esiRNAs derived from NATs can exert RNAi activity through AGO proteins (51). Interestingly, a recent study conducted in *Drosophila* cells provides evidence bridging esiRNA maturation with 3'-end processing factors that congregate at HLBs, including Symplekin, CPSF73 and CPSF100 (52). Depletion of these factors lead to a nuclear retention of retrotransposons and hairpin transcripts, which correlated with a reduction in mature esiRNA abundance. We found that AGO2 binds to *his* NATs fragments and that its depletion leads to weak but significant increases the levels of *his2a*, *his2b* and *his4* mRNA, together prompting speculations that *his* NATs may behave as esiRNA precursors. This model could be explored further by characterizing the molecular properties of putative *his* esiRNAs, such as 5' or 3' base preference and 3'-end modifications such as 2'-O-methylation. The biogenesis of putative *his* esiRNAs could be evaluated by assessing the impact of depleting small RNA effectors such as Dicer, Loquacious and R2D2 on *his* NATs and *his* esiRNAs levels and subcellular distributions.

In conclusion, the developmentally-regulated antisense transcripts we've described display properties that suggest a role in assisting HLB maturation at the co-transcriptional level and aiding maternal clearance post-transcriptionally (**Figure 9.7**). The abundance of *his*-GU repeats

and the overlap between *his* NATs and essential *his* mRNAs complicate functional interrogations through direct loss-of-function approaches. However, the involvement of *his* NATs in RNAi phenomena targeting the maternal *his* mRNA pool could be further investigated by evaluating the roles of diverse RNAi factors on *his* mRNA, *his* NATs and *his* esiRNAs expression profiles.

Material and method

Drosophila husbandry

OregonR flies were used as *wild type*. Nuclear and cytoplasmic extracts were prepared from time-resolved embryo collections as previously described(53). *Slbp*$^{10/12}$, *Slbp*$^{10/15}$ and *Ago*^{51B} embryos were obtained through crosses of trans-heterozygous mothers with *wt* males. Adar^{5G1} males were selected phenotypically and mated with *wt* females. Prior to deep-sequencing, time-resolved embryo collections were dechorionated with bleach (1%), frozen in liquid nitrogen and saved at -80 °C.

RNA *in situ* hybridization and Northern blotting

For imaging, embryos were collected, dechorionated with bleach (1%) and fixed with formaldehyde (4%) as described previously (54). Digoxigenin-labeled probes were prepared through PCR and *in vitro* transcription of previously released full-length *his2b* sequences cloned in the pGEM4 expression plasmid (13, 54, 55). Image acquisition was conducted on a Leica DM5500B microscope equipped with a QImaging ExiAqua camera (QImaging).

RNA isolation, quality control and Northern blotting

RNA was purified from embryonic, ovarian and cellular extracts using TRIZol™ reagent (Ambion) and processed according to the manufacturer's instructions. RNA preparations were further purified using the RNA Clean & Concentrator™-5 system (Zymo) with on-column DNAse I treatment (New England BioLabs) according to the manufacturer's instructions with minor modifications that were previously described (56). RNA quality control was assessed with Bioanalyzer electrophoretic profiles (Agilent) and optical density reading through a NanoDrop2000c spectrophotometer. Samples selected for RNA-seq had to meet stringent criteria were selected (RNA Integrity Number $\geq$9 ; A260/A280 $\geq$ 2.0; 2.00 $\leq$ A260/A280 $\leq$ 2.25). Northern blotting was executed as previously described (13) with the *his2b* digoxigenin-labeled probes used for FISH detection and using the Northern Starter Kit™ (Roche) as instructed by the manufacturer.

polyA+ enrichment, rRNA depletion and RNA-seq library preparation

RNA-seq analysis was conducted as previously described. Briefly, 1 μg of high quality RNA was used for library preparation with the TruSeq Stranded RNA Kit (Illumina). The RiboMinus kit (Thermo Fisher) was used for rRNA depletion and polyA$^+$ capture was performed with total *wt* embryo RNA using Dynabead™ (ThermoFisher Scientific) according to the manufacturer's instructions. The HiSeq 2000 machine (Illumina) was employed with TruSeq PE Clusterkit v3-cBot-HS.

RNAse A, DNAse I and α-amanitin treatments

To assess the specificity of *his2b* mRNA Northern blot pattern, 1 ug of total RNA purified from embryonic and ovarian extracts were incubated with recombinant RNAse A (Qiagen) and DNAse I (Sigma-Aldrich) according to the manufacturer's instructions. The nucleic acids were then precipitated overnight and submitted to Northern blotting. To assess the contribution of zygotic RNA Polymerase II to *his* NATs expression, *wt* embryos were collected 0-180 min AEL, dechoroionated with bleach (1%) and incubated for 15 min in Robb's medium containing either the Pol II inhibitor α-amanitin (100 ug/mL) or vehicle alone (DMSO). The embryos were then fixed with formaldehyde (4%) for FISH analyses.

Subcellular fractionation

Subcellular fractionation were performed as previously described on time-resolved embryo collections (25, 26) and D17 cells (24, 39).

***In silico* analyses**

FastQC v0.10.1 was used for read quality analysis and read trimming involved Trimmomatic. Tophat v2.0.10 was used for alignment to the *Drosophila melanogaster* BDGP5.78/dm3 genome. bigwig files derived from alignment BAM files were deposited on the *UCSC* genome browser tool for read coverage visualization. Normalized read counts were obtained from BAM files with DESeq2. Simplex graphs of RNA spatiotemporal distributions were derived from normalized read counts using previously described trigonometric transformations . AGO2-associated small RNAs were obtained from Haruhiko Siomi as a Bedgraph file and uploaded to

the UCSC genome browser to juxtapose the coordinates of *his* NATs and *his*-derived putative esiRNAs.

Acknowledgments

We thank members of the Lécuyer lab for productive discussions and suggestions, in particular Neal A.L. Cody and Julie Bergalet who generated the RNA-seq data for fractionated D17 cells and ovaries. The authors wish to thank the Duronio lab (UNC Chapel Hill University), who kindly provided the $Slbp^{10}$, $Slbp^{12}$, and $Slbp^{15}$ fly stocks, as well as the Li lab (Stanford University), who offered the $Adar^{5G1}$ flies, and Haruhiko Siomi, who generously shared a track of AGO2-associated small RNAs. We gratefully acknowledge Odile Neyret, Alexis Blanchet-Cohen and the IRCM Molecular Biology for their help with deep-sequencing. This study was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC; Discovery Grant #386644), the Canadian Institutes of Health Research (CIHR; MOP-111161) and a Junior 2 research scholar of the Fonds de Recherche du Québec - Santé (FRQS) to E.L. F.A.L. is funded by a Frederick Banting and Charles Best Canada Graduate Scholarship from CIHR.

Author contributions

E.L. and F.A.L contributed to study design and experimental work. F.A.L. conducted most analyses, assembled the figures and wrote the manuscript. L.P.B.B. performed all bioinformatic analyses. E.L. offered project supervision and revised the manuscript.

Competing interests

The authors declare no competing interests.

Figure and Table legends

Figure 9. 1. *His2B* NATs form amanitin-sensitive binuclear foci on chromatin during early embryogenesis

(A) Genomic architecture of the *Drosophila histone* gene complex (*his*-C) consisting of arrayed repeats of *histone* gene units (*his*-GU). The promoters are shown in green. **(B,C)** Micrographs of whole-mount fluorescence *in situ* hybridization (FISH) targeting coding (**b**) and antisense (**c**) *his2b* transcripts in time-resolved *wild type* embryos (RNA in green, DNA in red). Sagittal and surface view are displayed and the nuclear cycles are indicated (NC). **(D)** Micrographs of whole-mount FISH targeting coding (left) and antisense (right) *his1*, *his2a* and *his3* transcripts. **(E)** Increasing amounts of RNA (2.5 µg, 5 µg, 10 µg) extracted from wild type 0-4h embryo collections were blotted using probes targeting coding and antisense *his2b* transcripts by northern blotting. Ribosomal RNAs were revealed with methyl blue. The blots have been cropped; multiple exposures of the full film are provided in figure S2. **(F)** Micrographs of staged embryos exhibiting compromised levels of *his4* NAT upon treatment with the Pol II inhibitor α-amanitin. The polar body nucleus is labeled "p.b." Scale bars are indicated in the corner of representative micrographs.

Figure 9. 2. ***His* NATs coordinates and developmental regulation revealed by RNA-seq**

(A) *UCSC* genome browser screenshot of strand-specific read coverage of *his2b* in 0-180 min AEL embryos**. (B)** Screenshots of read coverage of a representative *his*-GU unit in total 0-180 min AEL *wt* embryos**.** Strand-specific tracks to the top show the levels of *his* mRNAs (scale=20,000 reads) and the scale of the bottom tracks are adjusted to indicate the levels of *his* NATs (scale=2,000 reads). The lower track displays stage-specific regions of high chromatin accessibility inferred from DNAse I experiments provided by the *Berkeley Drosophila Transcription Network Project*. *His* NATs are denoted by an asterisk. **(C, D)** Screenshots of read coverage of a representative *his*-GU unit in D17 larva cells **(c)** and total ovaries **(d).** Detectable *his* NATs are indicated by an asterisk**. (E,F)** Gene-specific maximal read count of coding (**e)** and antisense (**f**) *histone* genes in ovaries, D17 larva cells and 0-180 min AEL *wt* embryos. **(G)** Ratio of maximal read count mapped to each *histone* antisense transcript relative to the corresponding coding transcript (NAT/mRNA) in ovaries, D17 cells and 0-3h AEL embryos. **(H)** Ratio of

maximal read count mapped to *histone* mRNA (light red) and antisense RNA (dark red) in embryo extracts relative to D17 larva cells (*x-axis*). Scale bars reflect standard deviations for cross-library fold change values. Asterisks denote the significativity for differences between pairs of cognate *his* mRNAs and *his* NATs inferred through t-tests (****= $P<10^{-4}$). **(I)** Abundance ratio of maximal read count mapped to *histone* mRNA (light red) and antisense RNA (dark red) in embryos extracts relative to ovary extracts. Scale bars reflect standard deviations for cross-library fold change values. Asterisks denote the significativity for differences between pairs of cognate *his* mRNAs and *his* NATs inferred through t-tests (****= $P<10^{-4}$).

Figure 9. 3. ***His*** **NATs expression is SLBP-dependent**

(A,B,C) *UCSC* genome browser screenshot of read coverage of the *Slbp* locus (**A**), *his* gene complex (**B**) and *Gapdh1* locus (**C**) in *wt*, $Slbp^{10/12}$ and $Slbp^{10/15}$ 0-180 min AEL embryos. (**D,E**) Ratio of maximal read count mapped to *his* mRNAs (light red) and *his* NATs (dark red) in $Slbp^{10/12}$ (**d**) and $Slbp^{10/15}$ (**e**) embryos relative to *wt* embryos (*x-axis*). Scale bars reflect standard deviations for cross-library fold change values. In every case, the differences between pairs of cognate *his* mRNAs and *his* NATs inferred through t-tests were not significant (*ns* = $P>10^{-1}$).

Figure 9. 4. ***His*** **NATs lack a polyA-tail**

(**A**) Histogram showing polyA index distributions per biotype in D17 larva cells. The polyA index is defined as the ratio of read mapped in the polyA+ selection library divided by the reads mapped the rRNA-depletion library. PolyA index median values are indicated for each biotype group. (**B**) *UCSC* genome browser screenshot of strand-specific read coverage of a representative *his*-GU for deep-sequencing libraries established through rRNA-depletion and polyA selection of total D17 larva cells (polyA+/rRNA-depleted). Detectable *his* NATs are indicated by an asterisk**. (C)** Ratio of maximal read count mapped to *his* mRNAs (light red) and *his* NATs (dark red) in libraries established through rRNA-depletion and polyA selection of total D17 larva cells (polyA+/rRNA-depleted). Scale bars reflect standard deviations for cross-library fold change values. In every case, the differences between pairs of cognate *his* mRNAs and *his* NATs inferred through t-tests were not significant (*ns* = $P>10^{-1}$). **(D)** Northern blots performed on total RNA, $polyA^+$ and $polyA^-$ RNA extracts of 0-180 min AEL *wt* embryos targeting *his2b*

mRNA and *his2b* NAT. Ethidium bromide staining of ribosomal RNA is shown. The blots have been cropped; multiple exposures of the full images are provided in figure S4.

Figure 9.5. ***His*** **NATs and** ***his*** **mRNAs display similar nucleocytoplasmic occupancy in blastoderm embryos but not in D17 cells**

(A) Simplex representation of subcellular distributions of *his* mRNAs (left) and NATs (right) in larva cells D17. Transcriptome-wide distributions are displayed in black. Nuclear (N), cytosolic (C) and endomembrane (M) fractions are represented by corners of the triangle. **(B)** Ratio of maximal read count mapped to *histone* mRNAs (light red) and NATs (dark red) in cytosolic extracts relative to nuclear extracts of D17 cells (*x-axis*). Scale bars reflect standard deviations for cross-library fold change values. Asterisks denote the significativity for differences between pairs of cognate *his* mRNAs and *his* NATs inferred through t-tests (****= $P<10^{-4}$; ns= $P>10^{-1}$). **(C)** Simplex representation of spatiotemporal distribution of *histone* mRNAs (left) and NATs (right) in *wt* embryos. Transcriptome-wide distributions are displayed in black. Nuclear (N), cysotplasmic (C) extracts of blastoderm embryos (90-180 min AEL) and total early syncytial extracts (E; 0-45 min AEL) fractions are represented by corners of the triangle. **(D)** Ratio of maximal read count mapped to *his* mRNA (light red) and NATs (dark red) in cytosolic extracts relative to nuclear extracts of blastoderm embryos (*x-axis*). Scale bars reflect standard deviations for cross-library fold change values. In every case, the differences between pairs of cognate *his* mRNAs and *his* NATs inferred through t-tests were not significant (*ns* = $P>10^{-1}$).

Figure 9.6. **Argonaute-2 associates with** ***his*** **NATs fragments in S2 cells and regulates** ***his*** **mRNA levels during embryogenesis**

(A) *UCSC* genome browser screenshot of strand-specific read coverage of a representative *his*-GU in 0-180 min AEL embryos juxtaposed to a custom track showing the position of AGO2-associated small RNAs identified in S2 embryonic cells by *Kawamura et al.* (27). Detectable *his* NATs are indicated by an asterisk**.** (**B,C**) Ratio of maximal read count mapped to *his* mRNAs (light red) and *his* NATs (dark red) in $Ago2^{51B}$ (**b**) and $Adar^{5G1}$ (**c**) embryos relative to *wt* embryos (*x-axis*). Scale bars reflect standard deviations for cross-library fold change values. *P*-values are indicated for differences between *wt* and mutant conditions for *his* mRNAs (light red) and *his* NATs (dark red) inferred through t-tests (ns= $P>10^{-1}$). **(D)** Micrographs of whole-mount

fluorescence *in situ* hybridization (FISH) targeting *his4* mRNA in *Ago251B* mutants (left) and in *wt* embryos (right). **(E)** Surface view micrograph showing binuclear *his4* mRNA foci transcripts in post-mitotic $Ago2^{51B}$ mutant embryos (left). Sagittal view micrograph showing *his4* mRNA foci in yolk nuclei of $Ago2^{51B}$ mutant embryos (right). The green channel corresponds to the RNA FISH signal, the red signal corresponds to DNA (DAPI). The nuclear cycle (NC) and scale bars are indicated for each sample.

Figure 9.7. **Model of *his* NATs roles in early embryogenesis**

Between NC1 and NC6, *his* NATs expression could contribute to the maturation of HLBs, which orchestrate *his* pre-mRNA processing. *His* NATs could then anneal to maternal and zygotic *his* mRNAs to form *his* dsRNA recognized and processed by Dicer to generate AGO2-associated esiRNAs. These products could then contribute to the mother-to-zygotic transition by targeting maternal *his* mRNA for degradation by NC14.

Table 9. I. Read metrics of deep-sequencing libraries

Biological Sample	Library preparation	Reads (millions)	Duplicates (%)
Total Embryos *wt* 0-180 min AEL	rRNA depletion	24.3	25
Total Embryos *wt* 0-180 min AEL	rRNA depletion	26.4	27
Total Embryos *wt* 0-45 min AEL	rRNA depletion	32.8	27
Total Embryos *wt* 0-45 min AEL	rRNA depletion	31.1	24
Cytoplasmic extract of *wt* embryos 90-180 min AEL	rRNA depletion	41.4	29
Cytoplasmic extract of *wt* embryos 90-180 min AEL	rRNA depletion	36.2	27
Nuclear extract of *wt* embryos 90-180 min AEL nuclei	rRNA depletion	38.6	53
Nuclear extract of *wt* embryos 90-180 min AEL nuclei	rRNA depletion	37.9	36
Total Embryos *Slbp* $^{10/12}$ 0-180 min AEL	rRNA depletion	31.9	30
Total Embryos *Slbp* $^{10/12}$ 0-180 min AEL	rRNA depletion	30.3	22
Total Embryos *Slbp* $^{10/15}$ 0-180 min AEL	rRNA depletion	25.4	24
Total Embryos *Slbp* $^{10/15}$ 0-180 min AEL	rRNA depletion	28.3	24
Cytosolic extract of D17 larval haltere disc cells	rRNA depletion	20.7	53
Cytosolic extract of D17 larval haltere disc cells	rRNA depletion	24.3	55
Membrane extract of D17 larval haltere disc cells	rRNA depletion	21.8	17
Membrane extract of D17 larval haltere disc cells	rRNA depletion	21.1	18
Nuclear extract of D17 larval haltere disc cells	rRNA depletion	23.0	57
Nuclear extract of D17 larval haltere disc cells	rRNA depletion	24.9	53
Total extract of D17 larval haltere disc cells	rRNA depletion	24.0	31
Total extract of D17 larval haltere disc cells	rRNA depletion	22.6	32
Total extract of D17 larval haltere disc cells	polyA+ selection	20.5	42
Total extract of D17 larval haltere disc cells	polyA+ selection	20.7	42

Table 9. II. Maximal read counts for *His* genes and abundantly expressed control genes in total embryos, larva cells D17 and total ovaries

	Total *wt* embryo (0-180 min AEL)						Larva cells D17						Total *wt* Ovaries					
	Coding			Antisense			Coding			Antisense			Coding			Antisense		
	max read count			max read count			max read count			max read count			max read count			max read count		
His1	10,253	±	249	1,022	±	18	6,417	±	1,515	109	±	44	7,856	±	45	2	±	1
His2B	16,135	±	467	1,530	±	18	11,054	±	1,616	172	±	33	15,144	±	513	0	±	0
His2A	18,867	±	333	769	±	8	12,335	±	1,365	51	±	11	21,210	±	100	1	±	1
His4	19,761	±	753	699	±	53	7,158	±	1,111	22	±	5	19,694	±	536	1	±	1
His3	13,815	±	462	542	±	28	5,451	±	1,055	9	±	1	14,349	±	882	1	±	1

Table 9. III. Maximal read counts for *His* genes and NATs in total 0-180 min AEL embryos for *wt*, *Slbp*$^{10/12}$ and *Slbp*$^{10/15}$ genotypes

	His mRNA maximal read count					
	Total *wt* embryo (0-180 min AEL)		Total Slbp$^{10/12}$ embryo (0-180 min AEL)		Total Slbp$^{10/12}$ embryo (0-180 min AEL)	
	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Replicate 1	Replicate 2
His1	10077	10429	765	729	583	328
His2B	15804	16465	4201	3405	2556	3766
His2A	18631	19102	5654	5697	4952	3380
His4	19228	20293	2940	2687	2873	2400
His3	13124	13870	1883	1660	2061	1813
	His NAT maximal read count					
	Total *wt* embryo (0-180 min AEL)		Total Slbp$^{10/12}$ embryo (0-180 min AEL)		Total Slbp$^{10/12}$ embryo (0-180 min AEL)	
	Replicate 1	Replicate 2	Replicate 1	Replicate 2	Replicate 1	Replicate 2
His1	1034	1009	66	63	82	31
His2B	1517	1542	421	318	413	249
His2A	775	763	217	221	191	221
His4	389	368	55	39	70	57
His3	513	552	61	63	88	79

References

1. Clark MB, Amaral PP, Schlesinger FJ, Dinger ME, Taft RJ, Rinn JL, et al. The reality of pervasive transcription. PLoS Biol. 2011;9(7):e1000625; discussion e1102.
2. Lapidot M, Pilpel Y. Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. EMBO Rep. 2006;7(12):1216-22.
3. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW. The antisense transcriptomes of human cells. Science. 2008;322(5909):1855-7.
4. Osato N, Suzuki Y, Ikeo K, Gojobori T. Transcriptional interferences in cis natural antisense transcripts of humans and mice. Genetics. 2007;176(2):1299-306.
5. Nishikura K. Editor meets silencer: crosstalk between RNA editing and RNA interference. Nat Rev Mol Cell Bio. 2006;7(12):919-31.
6. DeCerbo J, Carmichael GG. Retention and repression: fates of hyperedited RNAs in the nucleus. Curr Opin Cell Biol. 2005;17(3):302-8.
7. St Johnston D, Nusslein-Volhard C. The origin of pattern and polarity in the Drosophila embryo. Cell. 1992;68(2):201-19.
8. Januschke J, Gervais L, Dass S, Kaltschmidt JA, Lopez-Schier H, St Johnston D, et al. Polar transport in the Drosophila oocyte requires Dynein and Kinesin I cooperation. Curr Biol. 2002;12(23):1971-81.
9. Yasuda GK, Baker J, Schubiger G. Temporal regulation of gene expression in the blastoderm Drosophila embryo. Genes Dev. 1991;5(10):1800-12.
10. Van Doren M, Williamson AL, Lehmann R. Regulation of zygotic gene expression in Drosophila primordial germ cells. Curr Biol. 1998;8(4):243-6.
11. Lu X, Li JM, Elemento O, Tavazoie S, Wieschaus EF. Coupling of zygotic transcription to mitotic control at the Drosophila mid-blastula transition. Development. 2009;136(12):2101-10.
12. Tadros W, Lipshitz HD. The maternal-to-zygotic transition: a play in two acts. Development. 2009;136(18):3033-42.
13. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. 2015;43(Database issue):D670-81.

14. Isogai Y, Keles S, Prestel M, Hochheimer A, Tjian R. Transcription of histone gene cluster by differential core-promoter factors. Gene Dev. 2007;21(22):2936-49.
15. Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet. 2008;9(11):843-54.
16. White AE, Leslie ME, Calvi BR, Marzluff WF, Duronio RJ. Developmental and cell cycle regulation of the Drosophila histone locus body. Mol Biol Cell. 2007;18(7):2491-502.
17. Salzler HR, Tatomer DC, Malek PY, McDaniel SL, Orlando AN, Marzluff WF, et al. A sequence in the Drosophila H3-H4 Promoter triggers histone locus body assembly and biosynthesis of replication-coupled histone mRNAs. Dev Cell. 2013;24(6):623-34.
18. Belli SI, Monnerat S, Schaff C, Masina S, Noll T, Myler PJ, et al. Sense and antisense transcripts in the histone H1 (HIS-1) locus of Leishmania major. Int J Parasitol. 2003;33(9):965-75.
19. Akhmanova A, Kremer H, Miedema K, Hennig W. Naturally occurring testis-specific histone H3 antisense transcripts in Drosophila. Mol Reprod Dev. 1997;48(4):413-20.
20. Wang ET, Cody NA, Jog S, Biancolella M, Wang TT, Treacy DJ, et al. Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. Cell. 2012;150(4):710-24.
21. Lefebvre FA, Benoit Bouvrette LP, Bergalet J, Lecuyer E. Biochemical Fractionation of Time-Resolved Drosophila Embryos Reveals Similar Transcriptomic Alterations in Replication Checkpoint and Histone mRNA Processing Mutants. J Mol Biol. 2017.
22. Lefebvre FA, Bouvrette LPB, Bergalet J, Lecuyer E. Data for the generation of RNA spatiotemporal distributions and interpretation of Chk1 and SLBP protein depletion phenotypes during Drosophila embryogenesis. Data Brief. 2017;13:28-31.
23. Lefebvre FA, Cody N, Bouvrette LPB, Bergalet J, Wang X, Lecuyer E. CeFra-seq: Systematic mapping of RNA subcellular distribution properties through cell fractionation coupled to deep-sequencing. Methods. 2017.
24. Lecuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, et al. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell. 2007;131(1):174-87.

25. Legendre F, Cody N, Iampietro C, Bergalet J, Lefebvre FA, Moquin-Beaudry G, et al. Whole mount RNA fluorescent in situ hybridization of Drosophila embryos. J Vis Exp. 2013(71):e50057.

26. Nizami Z, Deryusheva S, Gall JG. The Cajal Body and Histone Locus Body. Csh Perspect Biol. 2010;2(7).

27. Wilkie GS, Shermoen AW, O'Farrell PH, Davis I. Transcribed genes are localized according to chromosomal position within polarized Drosophila embryonic nuclei. Curr Biol. 1999;9(21):1263-6.

28. Arbeitman MN, Furlong EE, Imam F, Johnson E, Null BH, Baker BS, et al. Gene expression during the life cycle of Drosophila melanogaster. Science. 2002;297(5590):2270-5.

29. Edgar BA, Weir MP, Schubiger G, Kornberg T. Repression and turnover pattern fushi tarazu RNA in the early Drosophila embryo. Cell. 1986;47(5):747-54.

30. Graveley BR, Brooks AN, Carlson J, Duff MO, Landolin JM, Yang L, et al. The developmental transcriptome of Drosophila melanogaster. Nature. 2011;471(7339):473-9.

31. Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat Methods. 2016;13(6):508-14.

32. Sullivan E, Santiago C, Parker ED, Dominski Z, Yang X, Lanzotti DJ, et al. Drosophila stem loop binding protein coordinates accumulation of mature histone mRNA with cell cycle progression. Genes Dev. 2001;15(2):173-87.

33. Lanzotti DJ, Kaygun H, Yang X, Duronio RJ, Marzluff WF. Developmental control of histone mRNA and dSLBP synthesis during Drosophila embryogenesis and the role of dSLBP in histone mRNA 3' end processing in vivo. Mol Cell Biol. 2002;22(7):2267-82.

34. Nam JW, Bartel DP. Long noncoding RNAs in C. elegans. Genome Res. 2012;22(12):2529-40.

35. Beaulieu YB, Kleinman CL, Landry-Voyer AM, Majewski J, Bachand F. Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. PLoS Genet. 2012;8(11):e1003078.

36. Lécuyer FALLPBBJBÉ. Data for the Generation of RNA Spatiotemporal Distributions and Interpretation of Chk1 and SLBP Protein Depletion Phenotypes during Drosophila Embryogenesis. Data in Brief. 2016;In press(In press):In press.

37. Dominski Z, Yang XC, Raska CS, Santiago C, Borchers CH, Duronio RJ, et al. 3' end processing of Drosophila melanogaster histone pre-mRNAs: requirement for phosphorylated Drosophila stem-loop binding protein and coevolution of the histone pre-mRNA processing system. Mol Cell Biol. 2002;22(18):6648-60.

38. Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell. 2009;33(6):717-26.

39. Mehler MF, Mattick JS. Non-coding RNAs in the nervous system. J Physiol. 2006;575(Pt 2):333-41.

40. Chisholm KM, Wan Y, Li R, Montgomery KD, Chang HY, West RB. Detection of long non-coding RNA in archival tissue: correlation with polycomb protein expression in primary and metastatic breast carcinoma. PLoS One. 2012;7(10):e47998.

41. Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, et al. Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature. 2008;453(7196):793-7.

42. Lavender CA, Cannady KR, Hoffman JA, Trotter KW, Gilchrist DA, Bennett BD, et al. Downstream Antisense Transcription Predicts Genomic Features That Define the Specific Chromatin Environment at Mammalian Promoters. PLoS Genet. 2016;12(8):e1006224.

43. Flynn RA, Almada AE, Zamudio JR, Sharp PA. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. Proc Natl Acad Sci U S A. 2011;108(26):10460-5.

44. Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, et al. Divergent transcription from active promoters. Science. 2008;322(5909):1849-51.

45. Jaeger S, Martin F, Rudinger-Thirion J, Giege R, Eriani G. Binding of human SLBP on the 3'-UTR of histone precursor H4-12 mRNA induces structural rearrangements that enable U7 snRNA anchoring. Nucleic Acids Res. 2006;34(17):4987-95.

46. Tadros W, Westwood JT, Lipshitz HD. The mother-to-child transition. Dev Cell. 2007;12(6):847-9.

47. Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, et al. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science. 2006;312(5770):75-9.

48. Giraldez AJ. microRNAs, the cell's Nepenthe: clearing the past during the maternal-to-zygotic transition and cellular reprogramming. Curr Opin Genet Dev. 2010;20(4):369-75.
49. Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, et al. Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science. 2008;320(5879):1077-81.
50. Harrington AW, McKain MR, Michalski D, Bauer KM, Daugherty JM, Steiniger M. Drosophila melanogaster retrotransposon and inverted repeat-derived endogenous siRNAs are differentially processed in distinct cellular locations. BMC Genomics. 2017;18(1):304.
51. Sullivan W, Ashburner M, Hawley RS. Drosophila protocols. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 2000. xiv, 697 p. p.
52. Iampietro C, Bergalet J, Wang X, Cody NA, Chin A, Lefebvre FA, et al. Developmentally regulated elimination of damaged nuclei involves a Chk2-dependent mechanism of mRNA nuclear retention. Dev Cell. 2014;29(4):468-81.
53. Lecuyer E, Parthasarathy N, Krause HM. Fluorescent in situ hybridization protocols in Drosophila embryos and tissues. Methods Mol Biol. 2008;420:289-302.
54. Lefebvre FA, Benoit Bouvrette LP, Perras L, Blanchet-Cohen A, Garnier D, Rak J, et al. Comparative transcriptomic analysis of human and Drosophila extracellular vesicles. Sci Rep. 2016;6:27680.
55. Streit S, Michalski CW, Erkan M, Kleeff J, Friess H. Northern blot analysis for detection and quantification of RNA in pancreatic cancer cells and tissues. Nat Protoc. 2009;4(1):37-43.
56. Bierhoff H, Schmitz K, Maass F, Ye J, Grummt I. Noncoding transcripts in sense and antisense orientation regulate the epigenetic state of ribosomal RNA genes. Cold Spring Harb Symp Quant Biol. 2010;75:357-64.

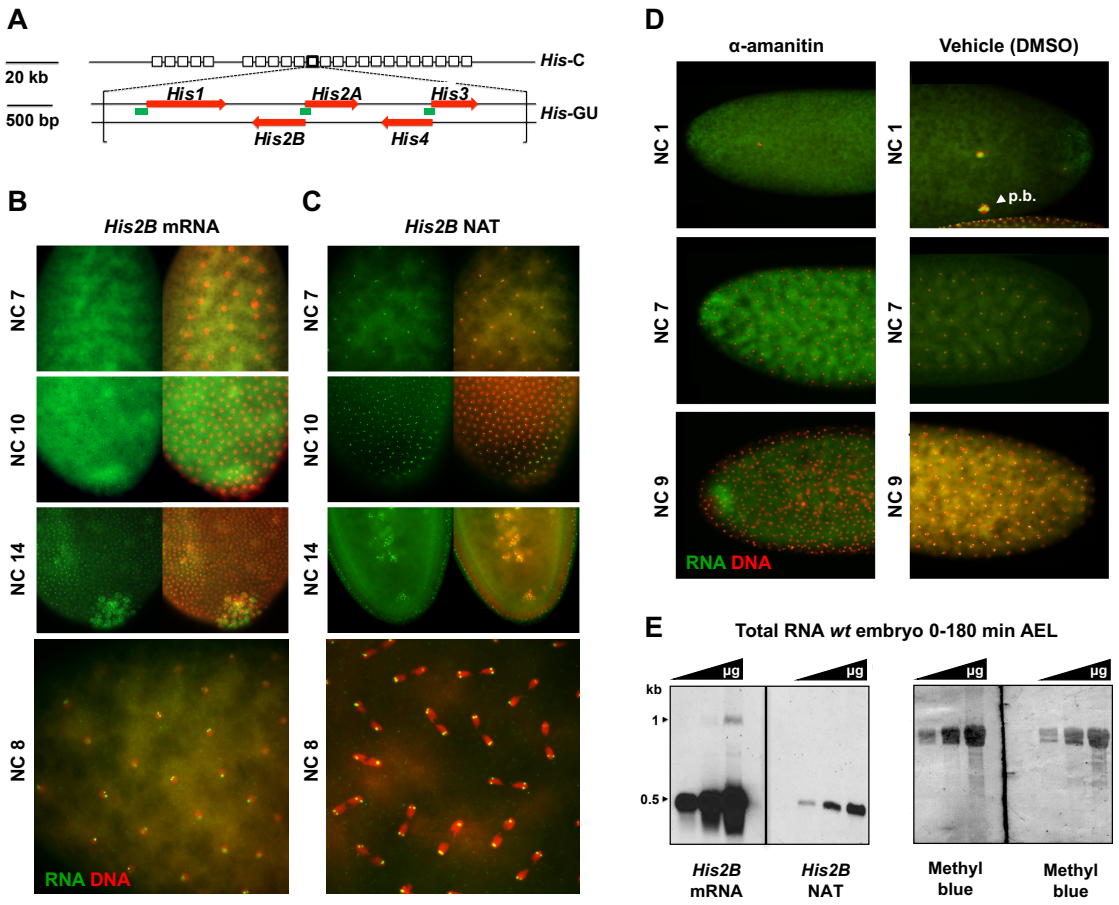


Figure 9. 1. ***His2B*** **NATs form amanitin-sensitive binuclear foci on chromatin during early embryogenesis**

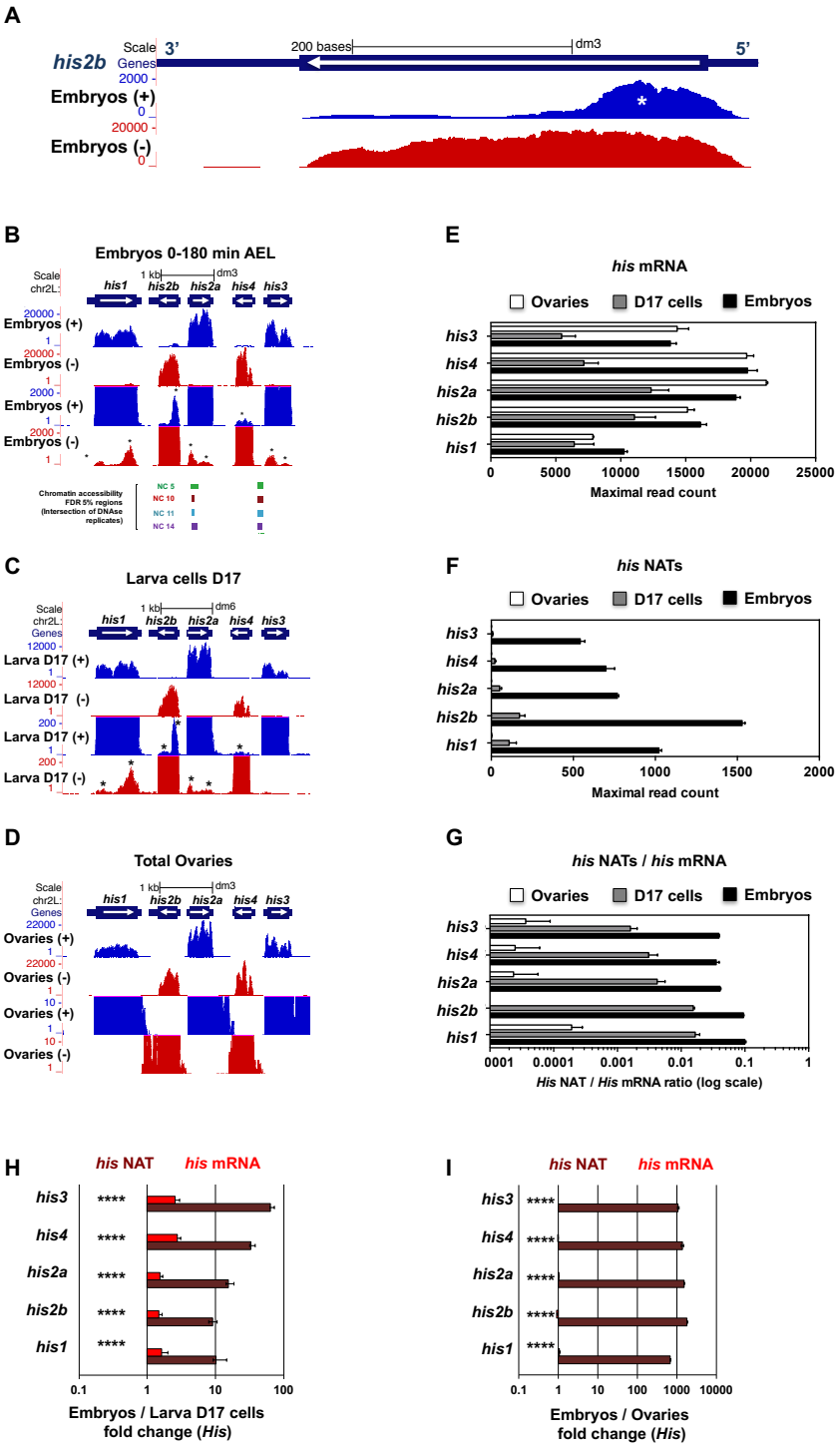


Figure 9. 2. ***His*** **NATs coordinates and developmental regulation revealed by RNA-seq**

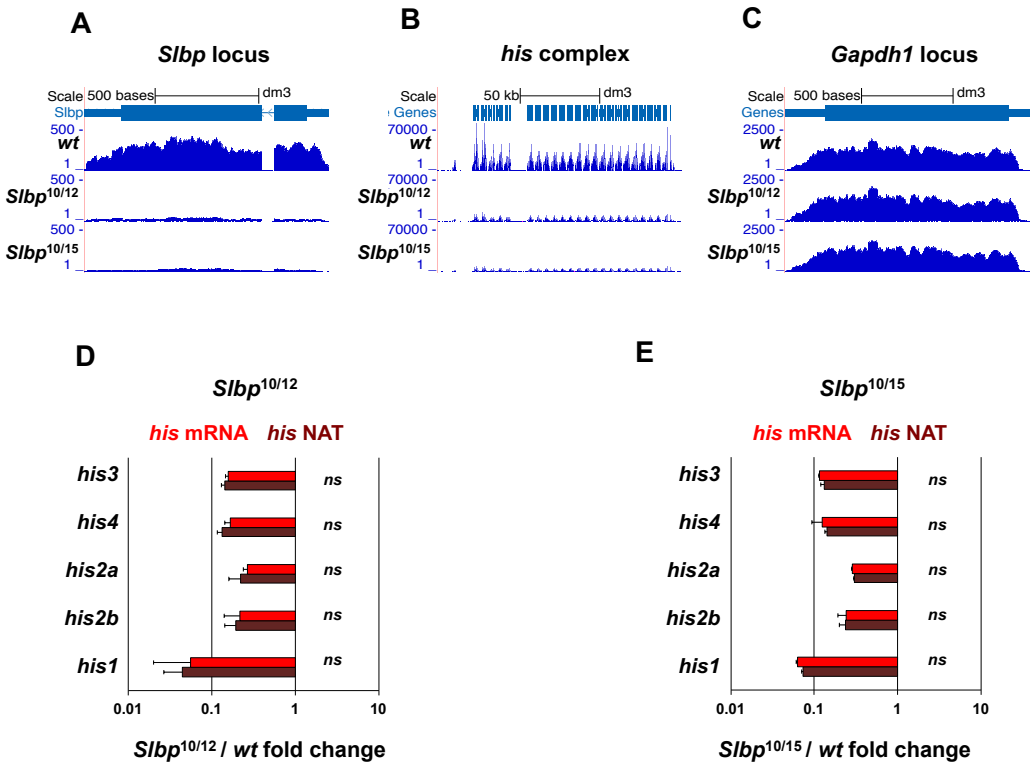


Figure 9. 3. ***His*** **NATs expression is SLBP-dependent**

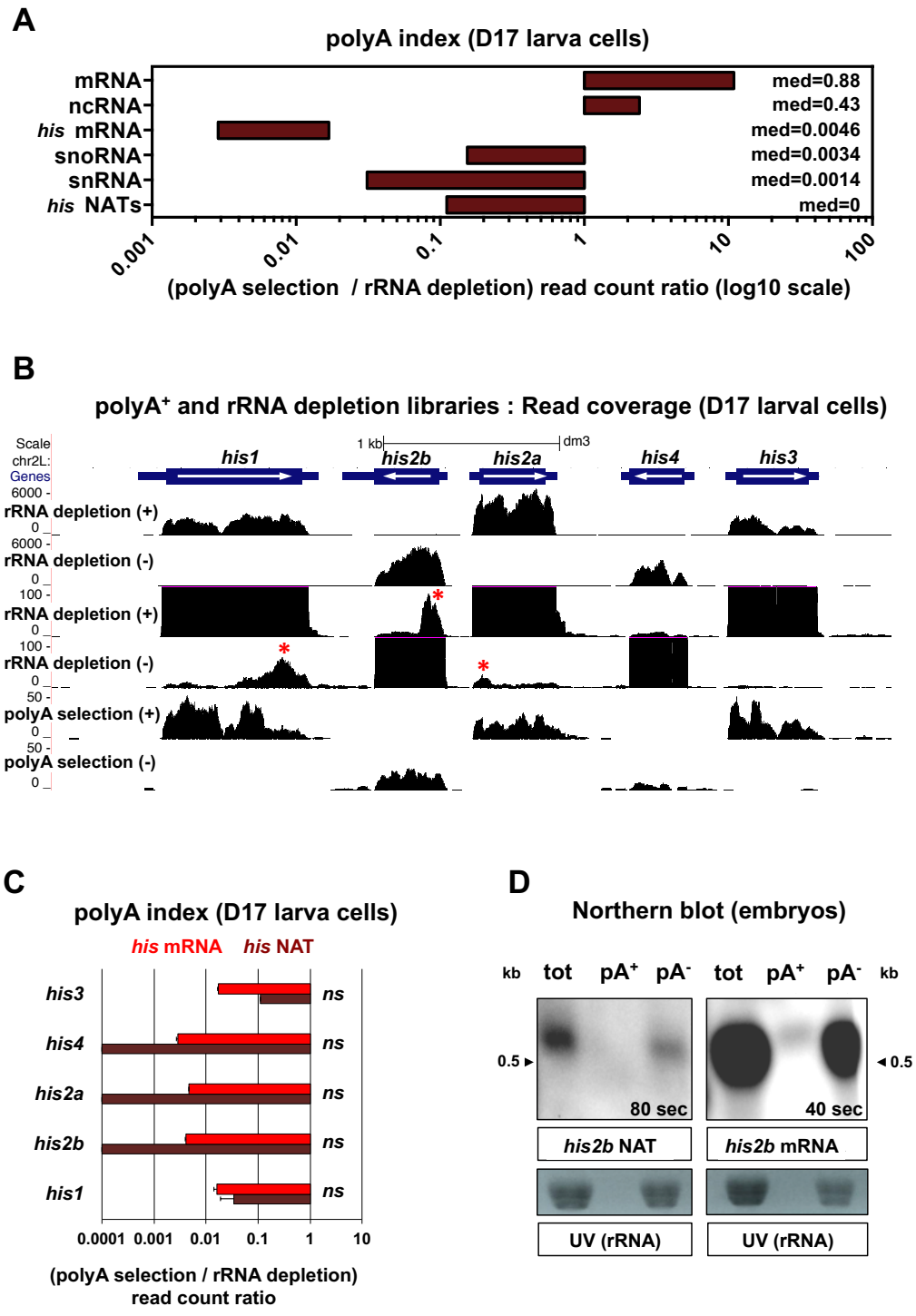


Figure 9. 4. ***His*** **NATs lack a polyA-tail**

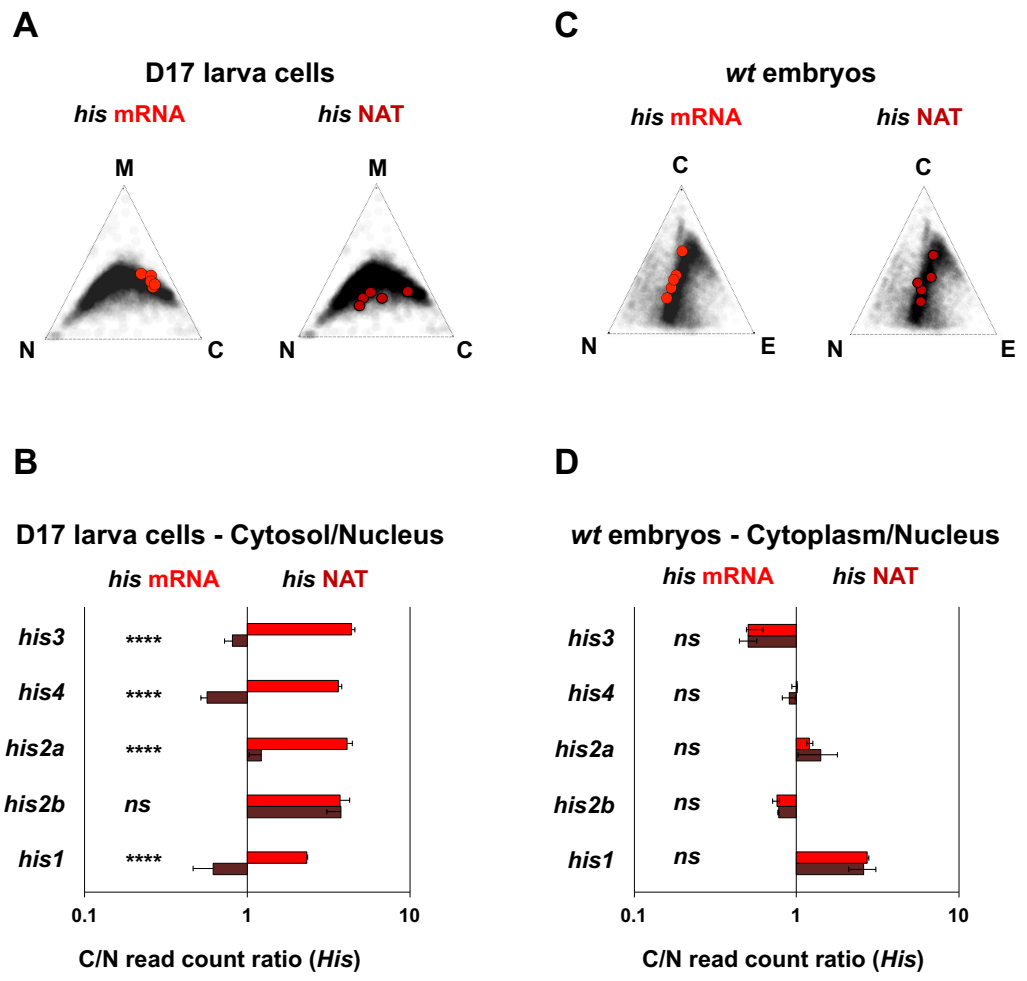


Figure 9.5. ***His*** **NATs and** ***his*** **mRNAs display similar nucleocytoplasmic occupancy in blastoderm embryos but not in D17 cells**

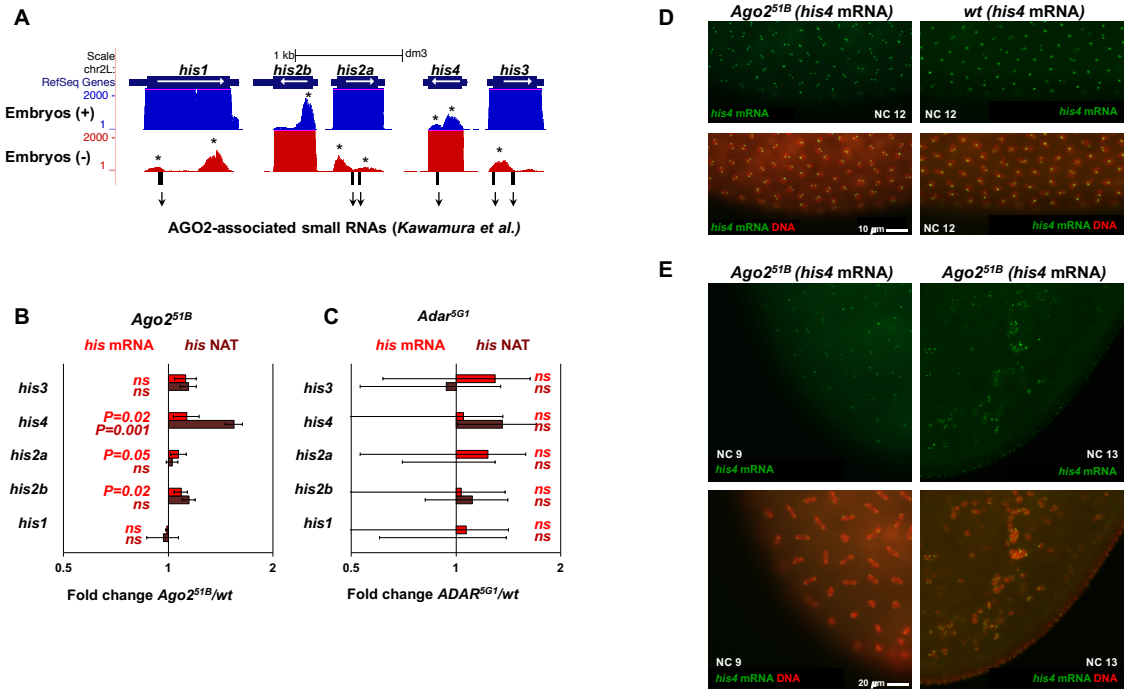


Figure 9. 6. Argonaute-2 associates with *his* NATs fragments in S2 cells and regulates *his* mRNA levels during embryogenesis

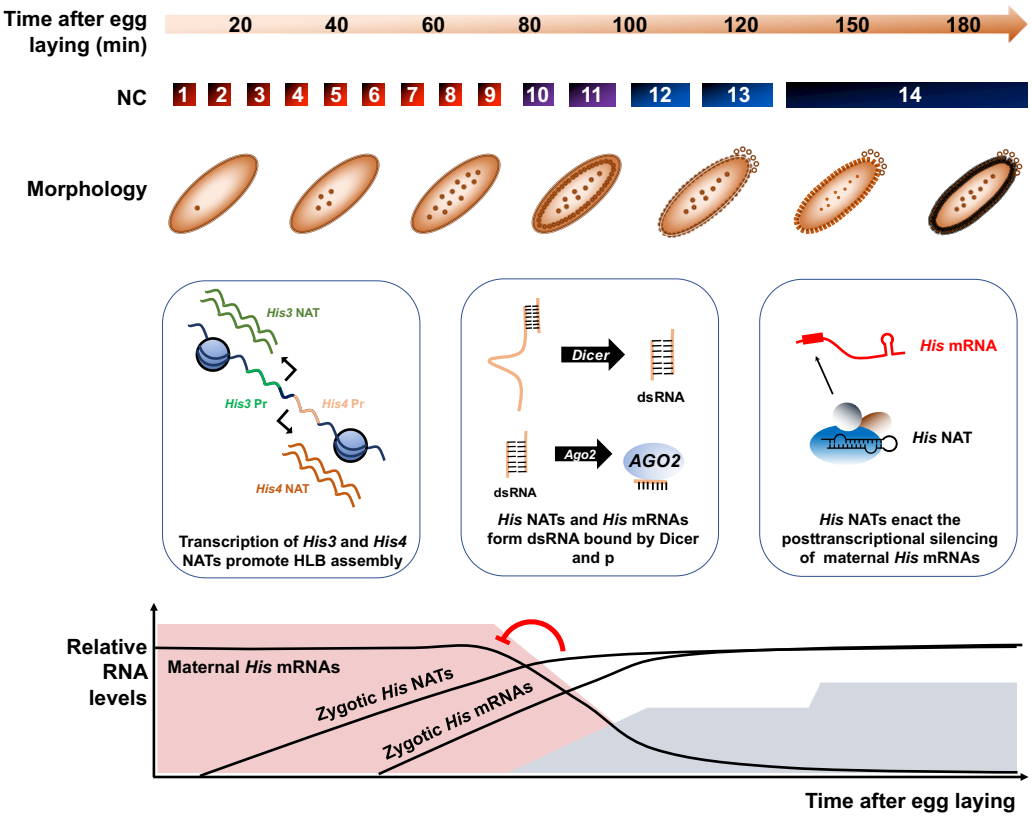


Figure 9. 7. Model of *His* NATs roles in early embryogenesis

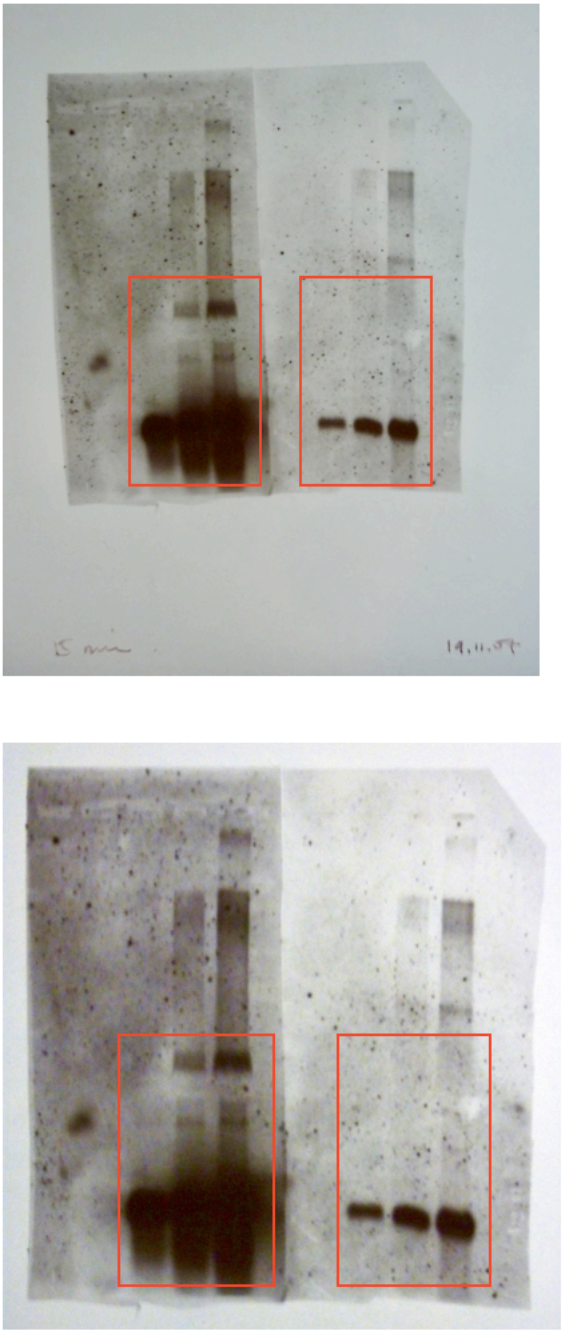


Figure 9S. 1. Full-length views of the *his2b* northern blot provided in figure 1E

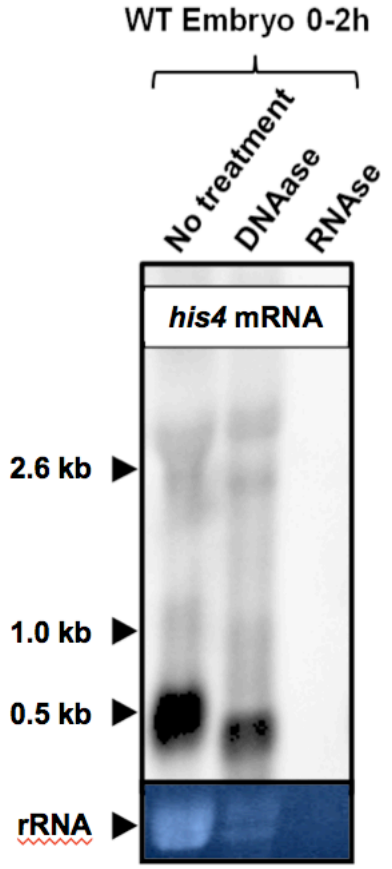
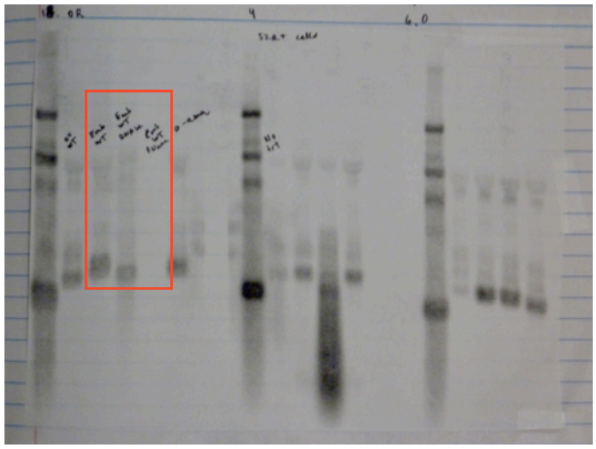


Figure 9S. 2. RNAse A treatment abolishes *his4* mRNA pattern by Northern blot but DNAse I does not



3 seconds – lab book photograph

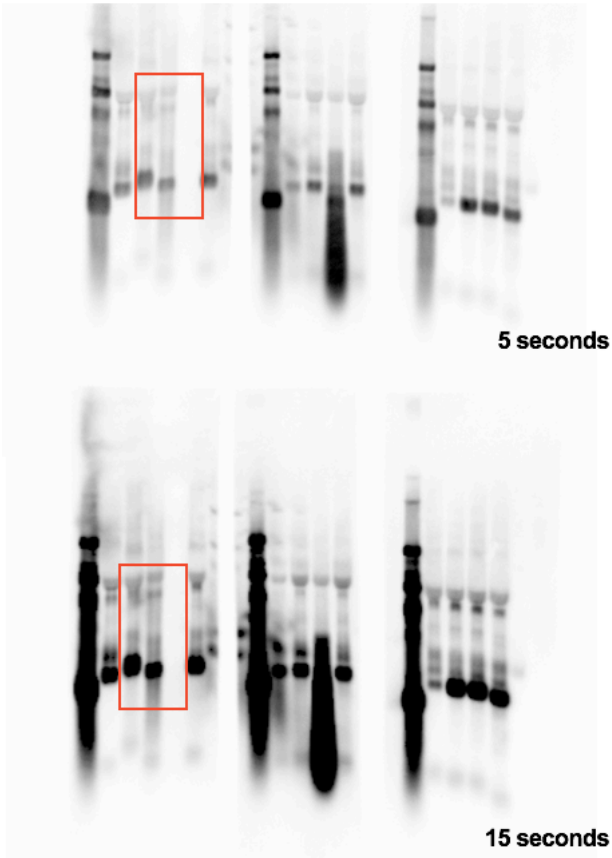


Figure 9S. 3. Multiple exposures and full-length views of the Northern blot provided in figure S2

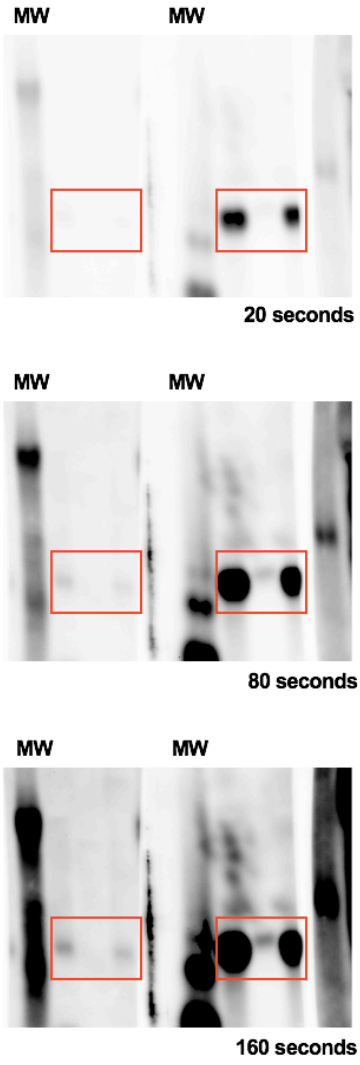


Figure 9S. 4. Multiple exposures and full-length views of the *his2b* Northern blot provided in figure 4D

Chapitre 10 : Discussion

10.1. Discussion en lien au chapitre 5 : l'approche CeFra-seq

10.1.1. Survol de l'approche CeFra-seq

Dans cette thèse, j'ai compilé les articles issus du développement et de l'application à divers systèmes biologiques d'une approche destinée à évaluer de manière systématique la distribution spatiale des ARNs. J'ai choisi le terme de « CeFra-seq » pour qualifier cette méthode, dont le protocole détaillé chez les systèmes cellulaires fait l'objet du chapitre 5. À la fois chez les cellules en culture et chez les embryons de Drosophile, le paradigme central de cette approche consiste à procéder à un fractionnement subcellulaire à froid en conditions isotoniques non-dénaturantes et en présence d'inhibiteurs de protéases et de ribonucléases. Les ARNs présents au sein de chaque échantillon subcellulaire et/ou extracellulaire sont ensuite extraits, purifiés et soumis à un séquençage à haut débit. En aval, des analyses informatiques permettent finalement de reconstituer de manière semi-quantitative la distribution des ARNs au sein des divers extraits subcellulaires. Le schéma ci-bas illustre les diverses applications de l'approche de CeFra-seq explorées pendant mon doctorat et recensées dans cette thèse (**Figure 10.1**).

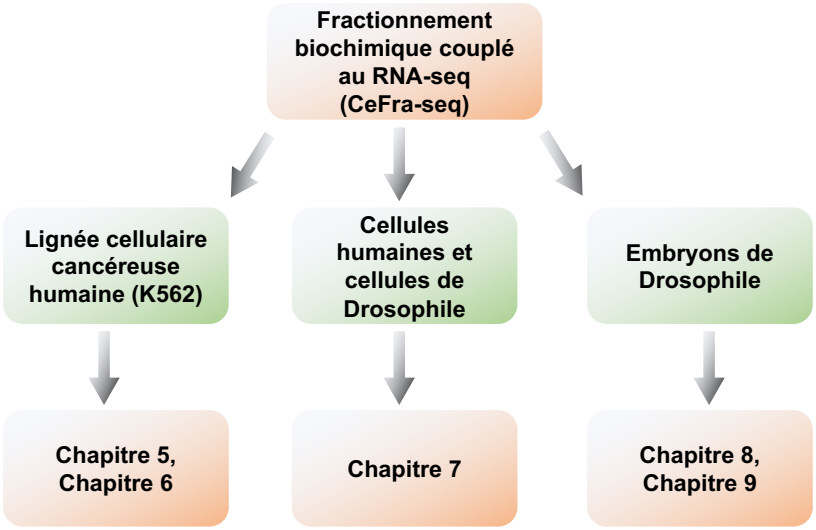


Figure 10. 1. Schéma récapitulatif des applications de la technique CeFra-seq au sein des différents chapitres de cette thèse.

10.1.2. Avantages et inconvénients de l'approche CeFra-seq

L'approche CeFra-seq est proposée comme un outil complémentaire aux méthodes d'imagerie des ARNs, tels l'hybridation *in situ* ou le système MS2, qui sont abordés dans les chapitres d'introduction [1] [2]. Ces méthodes permettant de visualiser les ARNs par microscopie offrent une résolution supérieure à CeFra-seq pour déterminer la distribution subcellulaire de ces molécules. Elles sont particulièrement utiles dans le contexte d'études fonctionnelles focalisées sur un transcrit ou quelques transcrits d'intérêt. En revanche, les méthodes d'imagerie s'avèrent rapidement fastidieuses quand vient le temps d'investiguer la distribution de centaines, voire de milliers d'ARNs distincts. Si quelques protocoles récents, comme le FISSEQ et des méthodes d'hybridation multiplexée, permettent d'apprécier la distribution de plusieurs transcrits sur un même échantillon, ces approches demeurent complexes et peu usitées [3]. CeFra-seq se présente donc comme une méthode avantageuse pour estimer la distribution spatiale du transcriptome dans son intégralité, bien que cette flexibilité soit acquise au prix d'une résolution moindre. Il importe également de noter que les fractions cytoplasmiques obtenues par l'approche CeFra-seq ne correspondent pas à des régions cytotopiques résolues de l'espace cytoplasmique et sont plutôt le reflet des coefficients de sédimentation de diverses structures subcellulaires, possiblement distales dans l'espace. Par exemple, mes analyses de protéomique ont révélé que la fraction cytoplasmique insoluble est enrichie en facteurs ribosomiques, en composantes du protéasome et en protéines du cytosquelette (Chapitre 5 et 6). Or, la biologie cellulaire nous enseigne que ces protéines ne sont pas nécessairement proximales dans l'espace cytoplasmique [4, 5]. Par association, les ARNs présentant des patrons de distribution analogues au sein des diverses fractions ne sont pas nécessairement co-localisées dans le sens strict du terme. Ainsi, les avantages de CeFra-seq se révèlent surtout dans le contexte d'une comparaison de la distribution du transcriptome suite à une perturbation donnée, par exemple une forme de stress ou la déplétion d'un facteur protéique qui contribue à la localisation des ARNs. Dans de tels contextes, ce sont les changements dans le patron de distribution plutôt que ce patron lui-même qui renseignent sur les implications post-transcriptionnelles de la perturbation à l'étude.

10.1.3. Validation de l'approche CeFra-seq par hybridation *in situ* et immunofluorescence

Une approche de validation détaillée par microscopie, reposant sur des immunofluorescences et des hybridations *in situ* visant les facteurs fortement enrichis dans les diverses fractions subcellulaires, permettrait de mieux définir ces-dernières. De tels efforts de validation par imagerie comptent parmi les plans de notre laboratoire pour raffiner l'approche CeFra-seq dans un avenir proche.

10.1.4. Fractionnement subnucléaire

Dans le même ordre d'idée, il serait intéressant d'étendre l'approche en incluant de nouvelles fractions subcellulaires, notamment des fractions subnucléaires. En effet, divers protocoles ont été proposés pour séparer jusqu'à huit fractions subnucléaires, respectivement associées au nucléoplasme, à l'euchromatine, aux nucléoles, à l'hétérochromatine et aux membranes nucléaires [6]. Soumettre ces divers échantillons à l'analyse transcriptomique de CeFra-seq pourrait révéler de nouvelles informations sur la maturation des ARNs et la composition transcriptomique de divers corps nucléaires, notamment les HLBs.

10.1.5. CeFra-seq, épissage alternatif et rétention des introns

L'épissage alternatif, phénomène par lequel certains exons sont sélectivement exclus de l'ARNm mature, contribue à la diversité du transcriptome et du protéome. De multiples études ont établi un rôle pour l'épissage alternatif dans la localisation différentielles de paires d'isoformes chez plusieurs organismes [7-14]. La contribution des processus de régulation post-transcriptionnelle à ces phénomènes demeure nébuleuse : les mécanismes de ces événements de localisation protéique différentielle s'exercent-ils sur les ARNm ou sur les protéines de ces isoformes? De même, certaines conditions mènent à la rétention de certains introns dans la séquence de l'ARNm mature [15-19]. Il serait intéressant d'utiliser l'approche CeFra-seq afin d'établir un inventaire systématique des isoformes présentant un patron de distribution distinct au niveau post-transcriptionel, ainsi que des cas de rétention intronique. J'ai tenté d'explorer l'épissage à partir des jeux de données exposés dans cette thèse en utilisant divers outils bio-informatiques, comme *LASSO* [20-22]. J'ai constaté que l'identification des événements

d'épissage alternatif par des analyses transcriptomiques requiert une profondeur de séquençage élevée, ce qui a compliqué ces efforts. Dans l'avenir, il pourrait s'avérer intéressant de répéter les expériences de CeFra-seq en séquençant les extraits d'ARN asymétriques avec une plus grande profondeur pour explorer de manière systématique les notions d'épissage alternatif et de rétention des introns.

10.1.6. Profilage des ribosomes

Afin d'élargir la palette de CeFra-seq, il serait intéressant de se servir de l'approche pour évaluer l'impact de diverses perturbations sur l'association des ARNs aux ribosomes. Au cours de la dernière décennie, l'étude des polysomes, ces complexes formés lors de la traduction par l'assemblage des ribosomes sur les ARNs en cours de traduction, a grandement bénéficié de l'accessibilité croissante des techniques de séquençage à haut débit. En effet, la technique de profilage des ribosomes vise à capturer les diverses formes de polysomes et à soumettre les transcrits associés à un séquençage en profondeur [23] [24]. L'application de cette méthode a notamment révélé que les lncRNAs s'associent aux ribosomes, bien qu'ils ne soient pas traduits [25]. Il serait possible d'étendre la procédure de fractionnement cytoplasmique afin d'isoler les polysomes et ainsi contraster l'impact d'une déplétion de facteurs protéiques spécifiques sur la localisation subcellulaire et l'efficacité de traduction.

10.2. Discussion en lien aux chapitres 6 et 7 : Ciblage des ARNs aux VEs

10.2.1. Ciblage des ARNs aux VEs

Les chapitres 6 et 7 sont alloués à l'exploration des mécanismes de ciblage des ARNs aux VEs. Les mécanismes connus, qui sont discutés dans le chapitre 2, suggèrent que le ciblage des ARNs aux VEs est un phénomène complexe reposant sur une multitude de voies de signalisation. En effet, des régulateurs protéiques spécifiques ont été impliqués dans la localisation des ARNs aux VEs, notamment hnRNPA2B1, Annexine A2 et KRAS [26-28] [29, 30]. Plusieurs travaux suggèrent également que le ciblage des ARNs aux VEs est influencé par certaines propriétés générales de ces transcrits, notamment leur taille [31]. En accord avec ce corpus d'évidences, mes résultats chez la lignée humaine K562 pointent vers une conclusion mitigée : le ciblage des ARNs aux VEs semble influencé à la fois par une dynamique déterministe, reposant sur des éléments de séquence, et sur une composante stochastique, par laquelle des milliers d'ARNs, surtout des ARNs de courte taille, sont incorporés aux VEs de manière non-spécifique (**Figure 10.2**).

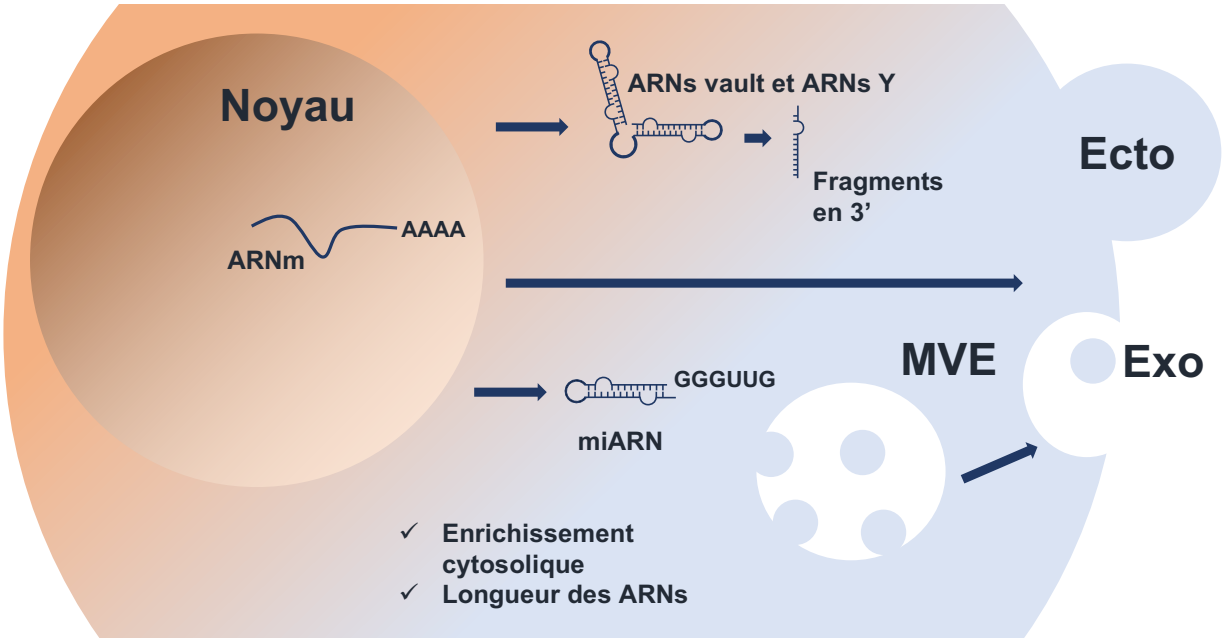


Figure 10. 2. Modèle récapitulatif des propriétés des ARNs associées à leur ciblage aux VEs, principales conclusions du chapitre 6.

La comparaison systématique des ARNs enrichis aux diverses fractions subcellulaires et extrtacellulaires (chapitre 6) a révélé diverses propriétés des transcrits enrichis aux VEs. Les populations cytosoliques et les populations extracellulaires présentent plusieurs propriétés communes : toutes deux sont majoritairement constituées d'ARNs courts (moins de 200 nt), surtout des ARNs non-codants, tels les ARNs vault et les ARNs Y. Ces ARNs sont présents à la fois sous leur forme intégrale et en tant que fragments en 3' d'environ 35 nt. Chez les ARNm, mes analyses ont révélé que l'enrichissement nucléaire corrèle avec l'enrichissement aux VEs. Parmi les miARNs enrichis aux VEs, la séquence GGGUUG est surreprésentée.

10.2.2. ARNs cytosoliques et *RNY4*

Le résultat le plus marquant du chapitre 6 est probablement le lien entre les populations d'ARNs cytosoliques et les populations d'ARNs extracellulaires. La similarité de ces populations s'est révélée avant même le séquençage, lors de l'étude des profils de taille des ARNs enrichis aux diverses fractions (puce *Bioanalyzer*). En effet, les ARNs cytosoliques comme les ARNs extracellulaires, sont surtout des ARNs de courte taille, notamment des ARNs non-codants. Parmi ces populations, plusieurs produits de la Polymérase III, notamment les ARNs vault et les ARNs Y, forment la majorité des transcrits extracellulaires. Cet enrichissement est si marqué que les fragments issus de *RNY4*, un ARN Y, constituaient 84% des séquences identifiées au sein des VEs dans la libraire de petits ARNs. Une telle abondance évoque un mécanisme hautement

spécifique de reconnaissance d'une séquence contenue chez *RNY4*. Dans le cadre d'une étude future, il serait intéressant d'explorer de tels mécanismes. Quel élément en *cis* contenu chez *RNY4* est responsable de ce ciblage? Existe-t-il un facteur protéique de reconnaissance des ARNs qui contribue à ce ciblage? Ces questions pourraient être abordées par une stratégie de mutagénèse de la séquence de *RNY4* et via la déplétion les facteurs associés à *RNY4* qui sont recensés dans la littérature, notamment Ro60 [32-34].

10.2.3. Ciblage aux VEs et demi-vie des ARNs

Une notion plus générale que je souhaiterais explorer en lien avec le ciblage aux VEs est la stabilité des ARNs. Mes travaux ont montré que les ARNm préférablement localisés aux VEs sont enrichis dans le noyau des cellules K562. En lien avec ce constat, je suggère une interprétation dans le chapitre 6 : ces ARNm présentent peut-être une demi-vie cytoplasmique courte. Il serait intéressant de valider cette hypothèse en mesurant la durée de leur transit dans le cytoplasme, ce qui pourrait être réalisé par des approches d'imagerie en cellules vivantes, tel le système MS2. De plus, identifier les déterminants de séquence de ce transit rapide représenterait une autre question à explorer par des approches de mutagenèse ou l'identification des facteurs protéiques associés à ces ARNs.

10.2.4. Ciblage aux VEs des ARNs d'éléments transposables

Dans le cadre d'une étude future, j'aimerais aussi explorer la dynamique du ciblage des éléments transposables aux VEs. En effet, plusieurs études ont montré que diverses séquences de rétrotransposons sont enrichies dans les VEs [35]. Dans le cadre de mon travail, cet aspect a été laissé de côté étant donné que les éléments transposables ne sont pas recensés dans les génomes de références utilisés dans nos analyses de séquençage. En revanche, il pourrait s'avérer révélateur de procéder à un ré-alignement ciblé de ces données pour obtenir un inventaire quantitatif des éléments transposables. Dans le contexte d'un projet plus ambitieux, qui pourrait s'intégrer à mon travail postdoctoral, j'aimerais explorer le transfert intercellulaire des séquences transposables via les VEs. En effet, puisque des séquences de transposon sont exportées au sein des VEs et que les VEs peuvent servir de véhicules dans la communication intercellulaire, il semble envisageable que les rétrotransposons issus d'un tissu donné intègrent le génome d'un

tissu éloigné ayant internalisé des VEs. Cette forme de communication conduisant à une restructuration du génotype même des cellules receveuses via les VEs pourrait représenter un nouveau paradigme élégant de la coordination tissulaire.

10.2.5. Ciblage des ARNs aux VEs chez *Homo sapiens* et *Drosophila melanogaster*

Le chapitre 7 est alloué à une analyse phylogénétique des propriétés morphologiques et transcriptomiques des VEs issues de deux systèmes biologiques, c'est-à-dire l'humain et la Drosophile. Ce travail a révélé divers traits convergents et divergents de ces structures chez les deux espèces. Ces traits sont illustrés ci-bas (**Figure 10.3**).

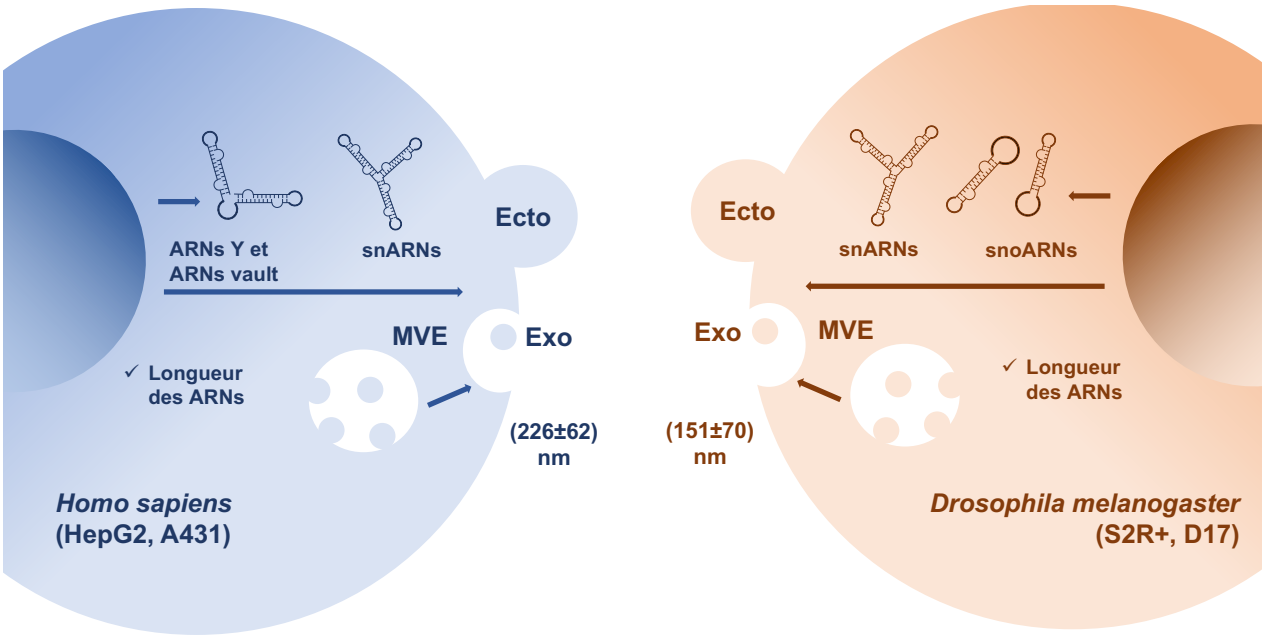


Figure 10. 3. **Modèle récapitulatif des propriétés morphologiques et transcriptomiques des VEs issues de l'humain et de la Drosophile.**

La taille moyenne des VEs chez les deux modèles étudiés pour chaque espèce est indiqué au centre de la figure. Les types d'ARN enrichis aux VEs de chaque espèce sont représentés dans le haut de la figure.

Le chapitre 7 a notamment révélé que l'enrichissement en ARNs courts documentés chez les VEs de mammifères prévaut également chez les VEs de Drosophile. Cette propriété évoque l'identification, dans le chapitre 6, des similarités entre les ARNs extracellulaires et les ARNs cytosoliques, qui sont eux aussi en moyenne plus courts que les ARN cellulaires totaux. Le corolaire de cette observation serait de vérifier, chez la Drosophile, si les ARN extracellulaires présentent des points communs avec les ARNs cytosoliques. Une telle hypothèse pourrait être

explorée en employant le protocole CeFra-seq sur les cellules de Drosophile utilisées dans le chapitre 7, c'est-à-dire les cellules D17 et S2R+.

10.2.6. Conservation de la biogenèse des VEs

Bien que notre compréhension des phénomènes de vésiculation et de biogenèse des VEs demeurent incomplets, des dizaines de travaux effectués chez les mammifères ont révélé certaines propriétés de ces phénomènes. Ainsi, un inhibiteur de la sphingomyélinase, GW4869, de même que la cytochalasine D ou le 5-ethyl-N-isopropyl amiloride bloquent la production d'exosomes chez diverses lignées cellulaires humaines [36, 37]. De plus, les petites GTPase Rab27a et Rab27b ont été impliquées dans la production des exosomes [38, 39]. De même, certaines composantes des complexes ESCRT (*endosomal sorting complexes required for transport*) sont requises pour la formation des exosomes [40, 41]. Afin de comprendre dans quelle mesure la biogenèse des VEs est un phénomène conservé, il pourrait s'avérer révélateur d'employer ces inhibiteurs et de procéder à la déplétion des orhtologues des facteurs Rab27 ou du complexe ESCRT chez la Drosophile. En aval, des analyses de NTA et de microscopie électronique effectuées après ces traitements pourraient révéler des changements dans le compte de VEs et dans la morphologie de ces structures. De telles expériences permettraient de vérifier dans quelle mesure les voies de sécrétion des VEs, caractérisées chez les mammifères, sont conservées chez les invertébrés.

10.2.7. Conservation des voies de ciblage hnRNPA2B1 et Annexine A2

S'il est vrai que plusieurs propriétés des VEs sont conservées de la Drosophile à l'humain, la question de la conservation des séquences d'ARN enrichies chez les deux systèmes est moins claire. En effet, dans le chapitre 7, une analyse corrélative des niveaux de milliers de transcrits orthologues chez l'humain et la Drosophile a montré que l'abondance de ces transcrits chez les VEs est moins bien corrélée que leur abondance cellulaire. L'interprétation intuitive de ce résultat est que collectivement, les voies de ciblage ne sont pas conservées entre les deux espèces. En revanche, face à la complexité et la diversité de ces voies, il semble judicieux de nuancer ce constat et de le confronter à de nouvelles évidences. À ce sujet, j'ai proposé une série d'expériences, qui ont été confiées à un nouvel étudiant du laboratoire. Nous savons, grâce à des

études précédentes, que les protéines hnRNPA2B1 et Annexine A2 sont impliquées dans le ciblage de certains miARNs aux VEs humaines [26, 30, 42]. Afin d'acquérir une meilleure compréhension des voies de ciblage, j'ai proposé de dépléter hnRNPA2B1, Annexine A2 et leurs orthologues chez la Drosophile, Hrb98DE et AnxB11. L'objectif serait ensuite de collecter les VEs produites par ces cellules et d'évaluer l'impact de ces déplétions protéiques sur les populations de petits ARNs ciblées aux VEs. La comparaison entre espèces des altérations associées à ces déplétions nous permettrait de déterminer s'il existe conservation des voies de ciblage de hnRNPA2B1 et d'Annexine A2.

10.2.8. Comparaisons phylogénétiques approfondies

Par ailleurs, recourir à des comparaisons phylogénétiques afin d'éclairer les mécanismes du ciblage des ARNs aux VEs me semble une approche prometteuse et je serais heureux d'avoir l'occasion de la pousser plus loin. L'intérêt de la communauté pour les VEs a rapidement progressé dans les dernières décennies, conduisant à l'apparition et à l'échange de jeux de données transcriptomiques des VEs issus plusieurs dizaines d'espèces éloignées [43-47] . En effet, les initiatives comme *Exocarta* et *Vesiclepedia* répertorient des analyses sur divers mammifères, mais également chez les Trypanosomes, les levures et les bactéries [48, 49]. Il serait intéressant d'établir un inventaire de ces jeux de données variés et de se prêter à un exercice bio-informatique visant à établir le degré de conservation des transcrits enrichis aux VEs des différentes espèces.

10.3. Discussion en lien aux chapitres 8 et 9 : Transcription des gènes d'histones dans l'embryogenèse précoce de la Drosophile

10.3.1. Extension de l'approche CeFra-seq à l'embryogenèse de la Drosophile

Le chapitre 8 propose d'étendre l'approche « CeFra-seq » à un contexte nouveau *in vivo*, le développement embryonnaire de *Drosophila melanogaster*. Pour ce faire, le protocole a été simplifié et adapté à la notion de changement de niveaux des transcrits dans le temps (Chapitre 8). Ici, CeFra-seq a permis d'identifier les ARNs enrichis chez l'embryon syncytial précoce, qui sont typiquement le fruit de la déposition maternelle. Il a également dévoilé les gènes strictement contribués par l'activation du génome zygotique, enrichis dans le noyau des blastodermes. Les gènes maternels stabilisés qui contournent la dégradation de la transition maternelle-zygotique sont plutôt enrichis dans le cytoplasme. Ce profilage a été suivi d'un effort de caractérisation, révélant que les ARNs zygotiques sont plus courts et simples que les ARNs maternels. Cette conclusion fait écho à un article récent démontrant cette tendance conservée chez diverses espèces [50]. Cet effort a également identifié un enrichissement de la séquence consensus de Zelda, un régulateur important de l'activation du génome zygotique, dans les promoteurs des gènes enrichis chez les noyaux blastodermes [51-55].

10.3.2. Le facteur SLBP est requis à l'activation du génome zygotique

J'ai ensuite utilisé le répertoire des gènes enrichis dans le contexte syncytial précoce, dans le noyau et dans le cytoplasme des blastodermes pour évaluer comment les niveaux d'expression de ces différents groupes d'ARNs répondent à la déplétion des facteurs SLBP et Chk1. Ce travail a montré que l'expression des gènes zygotiques est sélectivement compromise chez ces mutants (**Figure 10.4**).

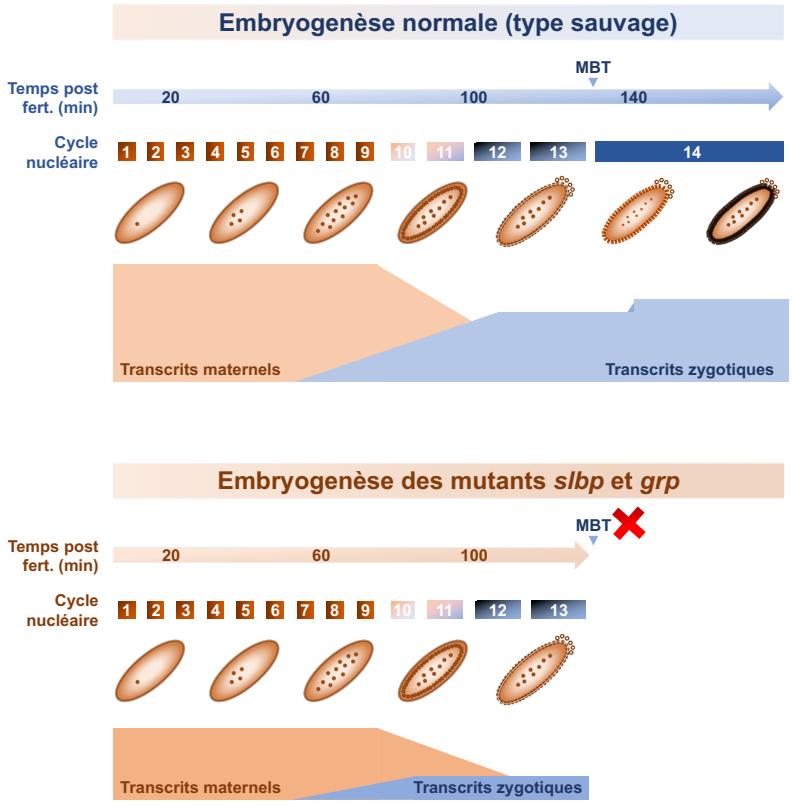


Figure 10. 4. Modèle récapitulatif de l'impact des mutations *slbp*$^{10/12}$, *slbp*$^{10/15}$ et *grp*^{fs1} sur le développement précoce [56] (Chapitre 7).

Dans l'embryogenèse normale (haut), les transcrits déposés maternellement font l'objet d'une dégradation régulée tandis que la transcription du génome zygotique fait l'objet d'une activation progressive, qui s'accompagne d'un allongement de la durée des cycles nucléaires. Ce processus culmine avec la transition du mi-blastula (MBT), où la structure syncytiale et la synchronie mitotique sont perdues et le contrôle du programme développemental devient strictement zygotique. Chez les mutants *slbp* et *grp*, le développement embryonnaire est interrompu avant la transition de mi-blastula et le génome zygotique n'est pas complètement activé.

10.3.3. Mécanismes de l'arrêt développemental des mutants *slbp*

Ainsi, l'adaptation de CeFra-seq au contexte développemental de la Drosophile a permis d'établir un inventaire des gènes zygotiques, dont l'expression s'est révélée sélectivement compromise par la déplétion des facteurs Chk1 et SLBP. En revanche, le chapitre 8 n'a pas permis d'identifier les causes de cet arrêt chez les deux mutants, question que j'aimerais explorer dans le contexte de travaux futurs. Tel que démontré dans des études précédentes et dans les chapitre 8 et 9, procéder à la déplétion de SLBP pendant l'embryogenèse compromet

l'expression des gènes d'histone, qui forment le cœur du nucléosome [57] [56, 58, 59]. Étudier la structure de la chromatine chez ces mutants pourrait éclairer les raisons de l'interruption de leur développement. En effet, la perte de la contribution zygotique des histones pourrait se traduire par une raréfaction des nucléosomes, menant à des défauts dans l'assemblage de la chromatine. Ces défauts pourraient conduire à des aberrations chromosomiques en mitose, débouchant par exemple sur des cas d'aneuploïdie, phénomène documenté chez les mutants *grp* [60, 61]. Des résultats de notre laboratoire ont montré que la perte du facteur Chk2 (*mnk*), impliqué dans la réponse aux dommages à l'ADN, permet d'atténuer certains phénotypes liés à la déplétion de SLBP [62]. Ce résultat suggère que la raréfaction des histones est interprétée comme une insulte génomique, activant une réponse régulée d'arrêt développemental. À ce sujet, il serait judicieux de comparer les altérations de la structure chromatinienne chez les mutants *slbp* et chez les doubles mutants *slbp mnk*. Une telle analyse pourrait révéler dans quelle mesure l'arrêt développemental est une conséquence directe de la perte d'expression des histones ou plutôt une conséquence de la réponse déclenchée par cette perturbation.

10.3.4. Applications additionnelles de CeFra-seq dans le contexte embryonnaire

La méthode CeFra-seq est un outil versatile et le développement embryonnaire se prête bien à l'utilisation de cette approche pour étudier d'autres questions biologiques qu'un arrêt développemental. Dans le but de mieux comprendre les voies de réponse aux dommages à l'ADN pendant l'embryogenèse, j'ai utilisé CeFra-seq sur des embryons préalablement soumis à une ronde d'irradiation aux rayons gamma, une source établie de lésions génotoxiques, notamment les bris double-brin [63-65]. Cette analyse, qui ne figure pas dans la présente thèse, a mis en évidence diverses altérations dans la distribution des ARNs suite à l'exposition aux rayons, notamment chez les ARNt. Dans les prochains mois, je compte revisiter ces données et tenter d'en extraire un manuscrit. J'aimerais également utiliser l'approche CeFra-seq afin de caractériser la dynamique post-transcriptionnelle des éléments transposables. Le développement précoce constitue une fenêtre temporelle importante pour ces parasites génétiques et pour la formation d'une défense contre la transposition par le génome zygotique [66-72]. De plus, dans son criblage visant à documenter l'étendue des phénomènes de localisation des ARNs pendant l'embryogenèse de la Drosophile, mon superviseur Éric a montré que les transcrits de divers éléments transposables sont exprimés chez l'embryon et adoptent des patrons de localisation

résolus, typiquement en association avec des domaines chromatiniens [73]. CeFra-seq pourrait contribuer à mieux comprendre le cycle de vie de ces transcrits, notamment chez des mutants des protéines Piwi ou d'autres facteurs associés à leur extinction transcriptionnelle.

10.3.5. Transcription antisens des gènes d'histones

Le chapitre 9 est alloué à la caractérisation et à l'exploration des rôles fonctionnels de transcrits antisens produits par le locus des histones dans le développement embryonnaire précoce de la Drosophile. Ce travail a mis en évidence la transcription antisens de séquences d'environ 80 nt situés en 5' et en 3' des gènes d'histone canoniques chez l'embryon. Cette expression est régulée au cours du développement, puisque les ovaires n'expriment pas ces produits et que l'inhibition de la Polymérase II chez le zygote la compromet. De plus, la perte du facteur SLBP est associée à une réduction massive de ces ARNs antisens non-polyadénylés. L'approche CeFra-seq a montré que ces transcrits co-ségrèguent étroitement avec leurs ARNm complémentaires, conduisant à l'hypothèse de la formation d'ARN double-brin, qui peut agir comme précurseur des petits ARNs interférents. Mon analyse d'une immunoprécipitation du facteur Argonaute-2 a révélé l'association de cet effecteur des voies d'interférence aux ARNs avec de petits transcrits issus du locus des histones. De plus, la déplétion du facteur Argonaute-2 chez l'embryon conduit à une dé-répression de trois ARNm d'histone, suggérant que les voies d'interférence aux ARNs sont impliquées dans la régulation de leur expression. Ensemble, ces évidences débouchent sur un modèle de contribution de la transcription antisens à la biogenèse de petits ARNs interférents, qui jouent un rôle dans la dégradation des populations maternelles d'ARNm (**Figure 10.5**).

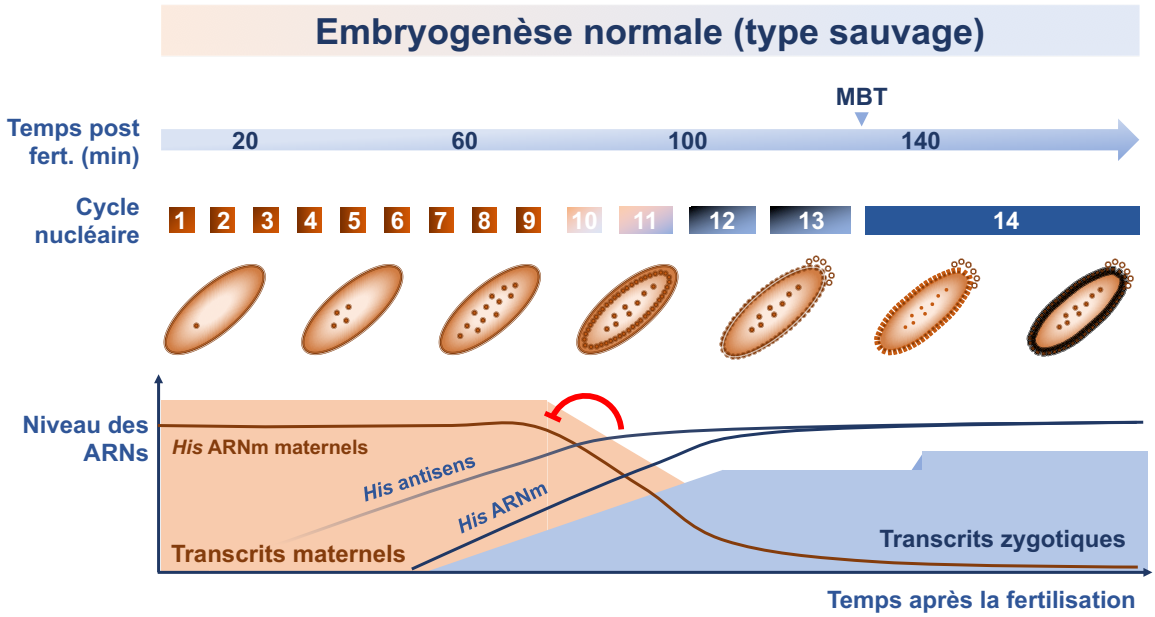


Figure 10. 5. Modèle récapitulatif de la dynamique et des rôles des transcrits d'histone pendant le développement embryonnaire de la Drosophile (Chapitre 9).

L'expression précoce par le zygote de séquences des gènes d'histone dans l'orientation antisens pourrait conduire à la formation d'ARN double-brin par hybridation aux ARNm endogènes. Ces produits pourraient ensuite faire l'objet d'une reconnaissance et d'une maturation par Dicer pour former de petits ARNs interférents, qui contribueraient à la dégradation des ARNm d'histone déposés maternellement, dans le contexte de la transition maternelle-zygotique.

10.3.6. Voies d'interférence aux ARNs dans la transition maternelle-zygotique

Plusieurs travaux effectués chez le poisson-zèbre et chez les mammifères ont mis en évidence la contribution de miARNs exprimés par le génome zygotique, notamment *miR-430*, dans la dégradation des transcrits maternels, facette importante de la transition maternelle-zygotique [74-76]. Ce paradigme est moins bien établi chez la Drosophile, bien que le facteur de transcription pionnier Zelda, régulateur crucial de l'activation du génome zygotique, ait été associé à l'expression précoce de la famille *miR-309* [54, 77, 78]. Examiner dans le détail les rôles des voies d'interférence aux ARNs dans la transition maternelle-zygotique pourrait s'intégrer à un profilage CeFra-seq de mutants Argonaute, qui pourrait révéler des observations évocatrices.

10.3.7. Supports additionnels au modèle du chapitre 9

Le chapitre 9 se termine sur la proposition du modèle d'une contribution de la transcription zygotique antisens à la dégradation des ARNm maternels via Argonaute-2. Avant de procéder à la soumission de ce manuscrit pour publication, j'aimerais étayer davantage ce modèle, qui m'apparaît séduisant, au moyen d'évidences expérimentales additionnelles. En effet, la déplétion d'Argonaute-2 ne mène qu'à une mince dé-répression de trois histones chez les embryons. Bien que ce résultat soit significatif, le modèle bénéficierait d'un support expérimental plus robuste. J'aimerais donc procéder à la déplétion d'autres protéines impliquées dans les voies d'interférence aux ARNs et évaluer l'impact de ces traitements sur les niveaux des ARNm maternels. Procéder à la déplétion des deux facteurs Dicer présents chez la Drosophile serait une expérience valable. Des travaux précédents ont associé Dicer-1 à la maturation des précurseurs de miARNs, tandis que Dicer-2 régule plutôt les petits ARNs interférents (siRNAs) [79]. Étudier les transcrits antisens chez ces mutants nous renseignerait donc sur la voie précise empruntée lors de la maturation des duplexes d'ARN issus du locus des histones. Par ailleurs, il serait intéressant de confronter le modèle en surexprimant les transcrits antisens, ce qui pourrait être réalisé au moyen d'une insertion chromosomique placée sous le contrôle d'un promoteur maternel fort. Une augmentation de l'expression des transcrits antisens devrait se traduire par une répression accrue des populations d'ARNm contribués maternellement.

Références

1. Hughes SC, Krause HM: **Single and double FISH protocols for Drosophila.** *Methods Mol Biol* 1999, **122:**93-101.
2. Daigle N, Ellenberg J: **LambdaN-GFP: an RNA reporter system for live-cell imaging.** *Nat Methods* 2007, **4:**633-636.
3. Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, Turczyk BM, Yang JL, Lee HS, Aach J, et al: **Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues.** *Nat Protoc* 2015, **10:**442-458.
4. Kwok R: **Cell biology: The new cell anatomy.** *Nature* 2011, **480:**26-28.
5. Paparella MM, Arnold W, Lim DJ, Jung TT, Doyle WJ, Eden AR, Uddman R, Goycoolea MV, Hawke WM, Ishii T, et al.: **Recent advances in otitis media. Anatomy, cell biology, and pathology.** *Ann Otol Rhinol Laryngol Suppl* 1989, **139:**19-22.
6. Tata JR, Baker B: **Sub-nuclear fractionation. I. Procedure and characterization of fractions.** *Exp Cell Res* 1974, **83:**111-124.
7. Delloye-Bourgeois C, Goldschneider D, Paradisi A, Therizols G, Belin S, Hacot S, Rosa-Calatrava M, Scoazec JY, Diaz JJ, Bernet A, Mehlen P: **Nucleolar localization of a netrin-1 isoform enhances tumor cell proliferation.** *Sci Signal* 2012, **5:**ra57.
8. Doorenweerd N, Mahfouz A, van Putten M, Kaliyaperumal R, PAC TH, Hendriksen JGM, Aartsma-Rus AM, Verschuuren J, Niks EH, Reinders MJT, et al: **Timing and localization of human dystrophin isoform expression provide insights into the cognitive phenotype of Duchenne muscular dystrophy.** *Sci Rep* 2017, **7:**12575.
9. Holt I, Duong NT, Zhang Q, Lam le T, Sewry CA, Mamchaoui K, Shanahan CM, Morris GE: **Specific localization of nesprin-1-alpha2, the short isoform of nesprin-1 with a KASH domain, in developing, fetal and regenerating muscle, using a new monoclonal antibody.** *BMC Cell Biol* 2016, **17:**26.
10. Ilouz R, Lev-Ram V, Bushong EA, Stiles TL, Friedmann-Morvinski D, Douglas C, Goldberg JL, Ellisman MH, Taylor SS: **Isoform-specific subcellular localization and**

function of protein kinase A identified by mosaic imaging of mouse brain. *Elife* 2017, **6**.

11. Kaneko T, Toshimori K, Iida H: **Subcellular localization of MS4A13 isoform 2 in mouse spermatozoa.** *Reproduction* 2017, **154:**843-857.
12. Manzo M, Wirz J, Ambrosi C, Villasenor R, Roschitzki B, Baubec T: **Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands.** *EMBO J* 2017, **36:**3421-3434.
13. Park SY, Grabau E: **Differential isoform expression and protein localization from alternatively spliced Apetala2 in peanut under drought stress.** *J Plant Physiol* 2016, **206:**98-102.
14. Thorne C, Eccles RL, Coulson JM, Urbe S, Clague MJ: **Isoform-specific localization of the deubiquitinase USP33 to the Golgi apparatus.** *Traffic* 2011, **12:**1563-1574.
15. Celebi JT, Wanner M, Ping XL, Zhang H, Peacocke M: **Association of splicing defects in PTEN leading to exon skipping or partial intron retention in Cowden syndrome.** *Hum Genet* 2000, **107:**234-238.
16. Hiller M, Huse K, Platzer M, Backofen R: **Non-EST based prediction of exon skipping and intron retention events using Pfam information.** *Nucleic Acids Res* 2005, **33:**5611-5621.
17. Kallabi F, Ben Rhouma B, Baklouti S, Ghorbel R, Felhi R, Keskes L, Kamoun H: **Splicing Defects in the AAAS Gene Leading to both Exon Skipping and Partial Intron Retention in a Tunisian Patient with Allgrove Syndrome.** *Horm Res Paediatr* 2016, **86:**90-93.
18. Ma L, Tan Z, Teng Y, Hoersch S, Horvitz HR: **In vivo effects on intron retention and exon skipping by the U2AF large subunit and SF1/BBP in the nematode Caenorhabditis elegans.** *RNA* 2011, **17:**2201-2211.
19. Verbeeren J, Verma B, Niemela EH, Yap K, Makeyev EV, Frilander MJ: **Alternative exon definition events control the choice between nuclear retention and cytoplasmic export of U11/U12-65K mRNA.** *PLoS Genet* 2017, **13:**e1006824.
20. Bitton DA, Rallis C, Jeffares DC, Smith GC, Chen YY, Codlin S, Marguerat S, Bahler J: **LaSSO, a strategy for genome-wide mapping of intronic lariats and branch points using RNA-seq.** *Genome Res* 2014, **24:**1169-1179.

21. Li W, Feng J, Jiang T: **IsoLasso: a LASSO regression approach to RNA-Seq based transcriptome assembly.** *J Comput Biol* 2011, **18:**1693-1707.
22. Zhang C, Zhang B, Lin LL, Zhao S: **Evaluation and comparison of computational tools for RNA-seq isoform quantification.** *BMC Genomics* 2017, **18:**583.
23. Sterne-Weiler T, Martinez-Nunez RT, Howard JM, Cvitovik I, Katzman S, Tariq MA, Pourmand N, Sanford JR: **Frac-seq reveals isoform-specific recruitment to polyribosomes.** *Genome Res* 2013, **23:**1615-1623.
24. Martinez-Nunez RT, Sanford JR: **Studying Isoform-Specific mRNA Recruitment to Polyribosomes with Frac-seq.** *Methods Mol Biol* 2016, **1358:**99-108.
25. van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E, Hao W, MacInnes AW, Cuppen E, Simonis M: **Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes.** *Genome Biol* 2014, **15:**R6.
26. Villarroya-Beltri C, Gutierrez-Vazquez C, Sanchez-Cabo F, Perez-Hernandez D, Vazquez J, Martin-Cofreces N, Martinez-Herrera DJ, Pascual-Montano A, Mittelbrunn M, Sanchez-Madrid F: **Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs.** *Nature Communications* 2013, **4**.
27. Cha DJ, Franklin JL, Dou Y, Liu Q, Higginbotham JN, Demory Beckler M, Weaver AM, Vickers K, Prasad N, Levy S, et al: **KRAS-dependent sorting of miRNA to exosomes.** *Elife* 2015, **4:**e07197.
28. McKenzie AJ, Hoshino D, Hong NH, Cha DJ, Franklin JL, Coffey RJ, Patton JG, Weaver AM: **KRAS-MEK Signaling Controls Ago2 Sorting into Exosomes.** *Cell Rep* 2016, **15:**978-987.
29. Filipenko NR, MacLeod TJ, Yoon CS, Waisman DM: **Annexin A2 is a novel RNA-binding protein.** *Journal of Biological Chemistry* 2004, **279:**8723-8731.
30. Hagiwara K, Katsuda T, Gailhouste L, Kosaka N, Ochiya T: **Commitment of Annexin A2 in recruitment of microRNAs into extracellular vesicles.** *Febs Letters* 2015, **589:**4071-4078.
31. Hung ME, Leonard JN: **A platform for actively loading cargo RNA to elucidate limiting steps in EV-mediated delivery.** *Journal of Extracellular Vesicles* 2016, **5**.
32. Simons FH, Rutjes SA, van Venrooij WJ, Pruijn GJ: **The interactions with Ro60 and La differentially affect nuclear export of hY1 RNA.** *RNA* 1996, **2:**264-273.

33. Reed JH, Gordon TP: **Autoimmunity: Ro60-associated RNA takes its toll on disease pathogenesis.** *Nat Rev Rheumatol* 2016, **12:**136-138.
34. Reed JH, Sim S, Wolin SL, Clancy RM, Buyon JP: **Ro60 requires Y3 RNA for cell surface exposure and inflammation associated with cardiac manifestations of neonatal lupus.** *J Immunol* 2013, **191:**110-116.
35. Balaj L, Lessard R, Dai L, Cho YJ, Pomeroy SL, Breakefield XO, Skog J: **Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences.** *Nat Commun* 2011, **2:**180.
36. Kosaka N, Iguchi H, Hagiwara K, Yoshioka Y, Takeshita F, Ochiya T: **Neutral Sphingomyelinase 2 (nSMase2)-dependent Exosomal Transfer of Angiogenic MicroRNAs Regulate Cancer Cell Metastasis.** *Journal of Biological Chemistry* 2013, **288:**10849-10859.
37. Khan S, Jutzy JM, Aspe JR, McGregor DW, Neidigh JW, Wall NR: **Survivin is released from cancer cells via exosomes.** *Apoptosis* 2011, **16:**1-12.
38. Bobrie A, Krumeich S, Reyal F, Recchi C, Moita LF, Seabra MC, Ostrowski M, Thery C: **Rab27a supports exosome-dependent and -independent mechanisms that modify the tumor microenvironment and can promote tumor progression.** *Cancer Res* 2012, **72:**4920-4930.
39. Ostrowski M, Carmo NB, Krumeich S, Fanget I, Raposo G, Savina A, Moita CF, Schauer K, Hume AN, Freitas RP, et al: **Rab27a and Rab27b control different steps of the exosome secretion pathway.** *Nat Cell Biol* 2009, **12:**19-30; sup pp 11-13.
40. Colombo M, Moita C, van Niel G, Kowal J, Vigneron J, Benaroch P, Manel N, Moita LF, Thery C, Raposo G: **Analysis of ESCRT functions in exosome biogenesis, composition and secretion highlights the heterogeneity of extracellular vesicles.** *J Cell Sci* 2013, **126:**5553-5565.
41. Tamai K, Tanaka N, Nakano T, Kakazu E, Kondo Y, Inoue J, Shiina M, Fukushima K, Hoshino T, Sano K, et al: **Exosome secretion of dendritic cells is regulated by Hrs, an ESCRT-0 protein.** *Biochem Biophys Res Commun* 2010, **399:**384-390.
42. Lefebvre FA, Lecuyer E: **Small Luggage for a Long Journey: Transfer of Vesicle-Enclosed Small RNA in Interspecies Communication.** *Front Microbiol* 2017, **8:**377.

43. Bayer-Santos E, Lima FM, Ruiz JC, Almeida IC, da Silveira JF: **Characterization of the small RNA content of Trypanosoma cruzi extracellular vesicles.** *Mol Biochem Parasitol* 2014, **193:**71-74.
44. Fernandez-Calero T, Garcia-Silva R, Pena A, Robello C, Persson H, Rovira C, Naya H, Cayota A: **Profiling of small RNA cargo of extracellular vesicles shed by Trypanosoma cruzi reveals a specific extracellular signature.** *Molecular and Biochemical Parasitology* 2015, **199:**19-28.
45. Peres da Silva R, Puccia R, Rodrigues ML, Oliveira DL, Joffe LS, Cesar GV, Nimrichter L, Goldenberg S, Alves LR: **Extracellular vesicle-mediated export of fungal RNA.** *Sci Rep* 2015, **5:**7763.
46. Lefebvre FA, Benoit Bouvrette LP, Perras L, Blanchet-Cohen A, Garnier D, Rak J, Lecuyer E: **Comparative transcriptomic analysis of human and Drosophila extracellular vesicles.** *Sci Rep* 2016, **6:**27680.
47. Lambertz U, Ovando MEO, Vasconcelos EJR, Unrau PJ, Myler PJ, Reiner NE: **Small RNAs derived from tRNAs and rRNAs are highly enriched in exosomes from both old and new world Leishmania providing evidence for conserved exosomal RNA Packaging.** *Bmc Genomics* 2015, **16**.
48. Keerthikumar S, Chisanga D, Ariyaratne D, Al Saffar H, Anand S, Zhao K, Samuel M, Pathan M, Jois M, Chilamkurti N, et al: **ExoCarta: A Web-Based Compendium of Exosomal Cargo.** *J Mol Biol* 2016, **428:**688-692.
49. Kalra H, Simpson RJ, Ji H, Aikawa E, Altevogt P, Askenase P, Bond VC, Borras FE, Breakefield X, Budnik V, et al: **Vesiclepedia: a compendium for extracellular vesicles with continuous community annotation.** *PLoS Biol* 2012, **10:**e1001450.
50. Heyn P, Kircher M, Dahl A, Kelso J, Tomancak P, Kalinka AT, Neugebauer KM: **The earliest transcribed zygotic genes are short, newly evolved, and different across species.** *Cell Rep* 2014, **6:**285-292.
51. Foo SM, Sun Y, Lim B, Ziukaite R, O'Brien K, Nien CY, Kirov N, Shvartsman SY, Rushlow CA: **Zelda potentiates morphogen activity by increasing chromatin accessibility.** *Curr Biol* 2014, **24:**1341-1346.
52. Hamm DC, Bondra ER, Harrison MM: **Transcriptional activation is a conserved feature of the early embryonic factor Zelda that requires a cluster of four zinc**

fingers for DNA binding and a low-complexity activation domain. *J Biol Chem* 2015, **290:**3508-3518.

53. Nien CY, Liang HL, Butcher S, Sun Y, Fu S, Gocha T, Kirov N, Manak JR, Rushlow C: **Temporal coordination of gene networks by Zelda in the early Drosophila embryo.** *PLoS Genet* 2011, **7:**e1002339.
54. Schulz KN, Bondra ER, Moshe A, Villalta JE, Lieb JD, Kaplan T, McKay DJ, Harrison MM: **Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo.** *Genome Res* 2015, **25:**1715-1726.
55. Sun Y, Nien CY, Chen K, Liu HY, Johnston J, Zeitlinger J, Rushlow C: **Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation.** *Genome Res* 2015, **25:**1703-1714.
56. Lefebvre FA, Benoit Bouvrette LP, Bergalet J, Lecuyer E: **Biochemical Fractionation of Time-Resolved Drosophila Embryos Reveals Similar Transcriptomic Alterations in Replication Checkpoint and Histone mRNA Processing Mutants.** *J Mol Biol* 2017.
57. Sullivan KD, Mullen TE, Marzluff WF, Wagner EJ: **Knockdown of SLBP results in nuclear retention of histone mRNA.** *Rna* 2009, **15:**459-472.
58. Dominski Z, Yang XC, Raska CS, Santiago C, Borchers CH, Duronio RJ, Marzluff WF: **3' end processing of Drosophila melanogaster histone pre-mRNAs: requirement for phosphorylated Drosophila stem-loop binding protein and coevolution of the histone pre-mRNA processing system.** *Mol Cell Biol* 2002, **22:**6648-6660.
59. Lanzotti DJ, Kaygun H, Yang X, Duronio RJ, Marzluff WF: **Developmental control of histone mRNA and dSLBP synthesis during Drosophila embryogenesis and the role of dSLBP in histone mRNA 3' end processing in vivo.** *Mol Cell Biol* 2002, **22:**2267-2282.
60. de Vries HI, Uyetake L, Lemstra W, Brunsting JF, Su TT, Kampinga HH, Sibon OC: **Grp/DChk1 is required for G2-M checkpoint activation in Drosophila S2 cells, whereas Dmnk/DChk2 is dispensable.** *J Cell Sci* 2005, **118:**1833-1842.
61. Takada S, Kwak S, Koppetsch BS, Theurkauf WE: **grp (chk1) replication-checkpoint mutations and DNA damage trigger a Chk2-dependent block at the Drosophila midblastula transition.** *Development* 2007, **134:**1737-1744.

62. Iampietro C, Bergalet J, Wang X, Cody NA, Chin A, Lefebvre FA, Douziech M, Krause HM, Lecuyer E: **Developmentally regulated elimination of damaged nuclei involves a Chk2-dependent mechanism of mRNA nuclear retention.** *Dev Cell* 2014, **29:**468-481.
63. Cassidy CL, Lemon JA, Boreham DR: **Impacts of low-dose gamma-radiation on genotoxic risk in aquatic ecosystems.** *Dose Response* 2007, **5:**323-332.
64. Graupner A, Eide DM, Brede DA, Ellender M, Lindbo Hansen E, Oughton DH, Bouffler SD, Brunborg G, Olsen AK: **Genotoxic effects of high dose rate X-ray and low dose rate gamma radiation in Apc(Min/+) mice.** *Environ Mol Mutagen* 2017, **58:**560-569.
65. Graupner A, Eide DM, Instanes C, Andersen JM, Brede DA, Dertinger SD, Lind OC, Brandt-Kjelsen A, Bjerke H, Salbu B, et al: **Gamma radiation at a human relevant low dose rate is genotoxic in mice.** *Sci Rep* 2016, **6:**32977.
66. Aravin AA, Hannon GJ, Brennecke J: **The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race.** *Science* 2007, **318:**761-764.
67. Chambeyron S, Popkova A, Payen-Groschene G, Brun C, Laouini D, Pelisson A, Bucheton A: **piRNA-mediated nuclear accumulation of retrotransposon transcripts in the Drosophila female germline.** *Proc Natl Acad Sci U S A* 2008, **105:**14964-14969.
68. Chueh AC, Northrop EL, Brettingham-Moore KH, Choo KH, Wong LH: **LINE retrotransposon RNA is an essential structural and functional epigenetic component of a core neocentromeric chromatin.** *PLoS Genet* 2009, **5:**e1000354.
69. Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z, Zamore PD: **Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells.** *Science* 2008, **320:**1077-1081.
70. Goodier JL, Kazazian HH, Jr.: **Retrotransposons revisited: the restraint and rehabilitation of parasites.** *Cell* 2008, **135:**23-35.
71. Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, Filippov DV, Blaser H, Raz E, Moens CB, et al: **A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish.** *Cell* 2007, **129:**69-82.
72. Kalmykova AI, Klenov MS, Gvozdev VA: **Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline.** *Nucleic Acids Res* 2005, **33:**2052-2059.

73. Lecuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, Hughes TR, Tomancak P, Krause HM: **Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function.** *Cell* 2007, **131:**174-187.

74. Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, Ma E, Mane S, Hannon GJ, Lawson ND, et al: **A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity.** *Science* 2010, **328:**1694-1698.

75. Giraldez AJ: **microRNAs, the cell's Nepenthe: clearing the past during the maternal-to-zygotic transition and cellular reprogramming.** *Curr Opin Genet Dev* 2010, **20:**369-375.

76. Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF: **Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs.** *Science* 2006, **312:**75-79.

77. Fu S, Nien CY, Liang HL, Rushlow C: **Co-activation of microRNAs by Zelda is essential for early Drosophila development.** *Development* 2014, **141:**2108-2118.

78. Harrison MM, Li XY, Kaplan T, Botchan MR, Eisen MB: **Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition.** *PLoS Genet* 2011, **7:**e1002266.

79. Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, Carthew RW: **Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways.** *Cell* 2004, **117:**69-81.

Chapitre 11 : Conclusion

Cette thèse relate le développement d'une approche novatrice dans l'étude systématique de la localisation des ARN, nommée CeFra-seq. Elle aborde ensuite quatre questions fonctionnelles chez deux systèmes biologiques distincts au moyen de cette approche. L'approche CeFra-seq est proposée comme une méthode versatile complémentaire aux outils d'imagerie dans l'analyse de perturbations larges du transcriptome. Tel que proposé dans le chapitre 10, il serait judicieux de procéder à des efforts de validation par imagerie afin de mieux définir l'identité des fractions subcellulaires obtenues par CeFra-seq. Dans un deuxième temps, l'approche pourrait être étendue de manière à inclure de nouvelles fractions subnucléaires ou en documentant l'association des transcrits aux polysomes. De telles modifications permettraient d'aborder de nouvelles questions biologiques et d'aborder la localisation de l'ARN en lien à la maturation nucléaire et à la régulation traductionnelle.

Dans les chapitres 6 et 7, la méthode CeFra-seq a été adaptée à l'étude des mécanismes de ciblage des ARNs aux VEs. Ces travaux ont révélé que les populations d'ARNs cytosoliques sont très similaires aux populations extracellulaires, toutes deux dominées par de courts transcrits non-codants. Ils ont aussi établi que le ciblage aux VEs est un phénomène complexe, reflétant une hétérogénéité marquée, ainsi que la contribution conjointe de facteurs déterministes et stochastiques. Dans le cadre de travaux futurs, il serait pertinent de se concentrer sur un cas précis d'ARN enrichi aux VEs, par exemple le transcrit *RNY4*. Ce-dernier pourrait faire l'objet d'expériences de mutagenèse afin d'identifier les éléments de séquence responsable de son enrichissement marqué. De telles expériences, couplées à des essais de perte de fonction, pourraient également conduire à l'identification de facteur en *trans* qui contribuent au ciblage des ARNs aux VEs.

Dans les chapitres 8 et 9, je propose une adaptation de la méthode CeFra-seq au contexte dynamique de l'embryogenèse de la Drosophile. Ces expériences ont mené à l'identification du répertoire de transcrits strictement exprimé par le génome zygote et mis en évidence la compromission sélective de leur expression chez les mutants *slbp* et *grp*. L'approche CeFra-seq a également révélé que les transcrits d'ARN, d'histones et leur transcrits antisens, caractérisés dans le chapitre 8, présentent une co-ségrégation dans l'embryon, menant à l'hypothèse de la formation d'ARN double-brin. Mes analyses d'immunoprécipitation et de déplétion d'Argonaute-2 ont conduit à un modèle selon lequel la transcription précoce et strictement

zygotique des ARNs antisens sert de précurseur à de petits ARNs interférents, qui contribuent à la transition maternlle-zygotique en dégradant les ARNm d'histones. Avant de procéder à la publication de ce manuscrit, je souhaiterais étayer ce modèle en étudiant l'impact de la déplétion d'autres effecteurs des voies des petits ARNs, dont Dicer-1 et Dicer-2, sur les niveaux des ARNm d'histone.

Collectivement, les travaux recensés dans cette thèse établissent l'approche de CeFra-seq comme un outil versatile dans l'étude de la distribution spatiale et temporelle du transcriptome. Encadrée d'efforts de validation en microscopie, cette méthode permet d'établir rapidement et de manière reproductible la distribution du transcriptome intégral, à la fois en culture et *in vivo*.