Identification of mRNA features which facilitate their long-range transfer in an *in vivo Drosophila* model



Claire Marie Hill

Balliol College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary Term 2021

This thesis is in memory of
my Granny Lily
who taught us that great accomplishments
are not beyond our reach.

Acknowledgements

For me, these are the most important pages of my thesis because without those mentioned here, none of the other pages would have been possible. There are many others beyond those explicitly mentioned here who have helped and supported me and for that, I am truly thankful.

Institutional

I am deeply grateful to my supervisors Alberto Baena-Lopez and Dave Carter for their continued support, both personally and professionally, throughout my PhD studies. I would like to thank them for having me as a member of their research teams. They have guided me at every stage of this research project and have helped me develop as a scientist. They have given of their time and energy to fully support my training, mentor me and provide me with a wealth of opportunities.

I would also like to thank the Oxford Doctoral Training Centre, particularly Esther Becker and Gail Preston, for giving me the opportunity to join the Interdisciplinary Bioscience Doctoral Training Programme and for the support and training I have received over the past four and a half years. I would like to extend my sincere thanks to the Biotechnology and Biological Sciences Research Council (BBSRC) as without their funding, none of this research would have been possible.

I would like to offer my special thanks to Elizabeth Dellar for her camaraderie throughout this project. We have worked closely on various aspects of this research and I am truly grateful for her wealth of insights, hard work and dedication, especially during the challenging times experienced during our final year.

Additionally, I would like to thank our collaborators, Monika Gullerova and Jana Terenova. Along with their contributions to this research, they provided me with invaluable direction and advice, inspiring me to learn new techniques and approaches.

My thanks go to Keith A. Gillow who developed the original Oxford Thesis Template for Latex, as well as to Sam Evans and John McManigle for their modifications to this template.

I would like to show my deep appreciation to all members of the Lopez and Carter labs, past and present, for their advice, encouragement, vision and most importantly, their friendship. I have had the pleasure of working between two wonderful lab groups and I have met many fantastic scientists. I would also like to thank all those who work at the Sir William Dunn School of Pathology and the Department of Biological and Medical Sciences at Oxford Brookes for building such friendly and engaging environments to work in.

I offer a special thanks to Alessia Galasso, Jana Alonso, Lewis Arthurton and Derek Xu who have been a strong support since my first day in the lab, personally and professionally. I also thank Sebastian Kittleman, Franz Wendler, Osamu Suyari, Ryan Pink, Li Wang and Amy Lovell for their support, as well as their contributions to this work. My PhD experience would not have been the same without these scientists and I thank them for being great colleagues, as well as wonderful friends.

Personal

Without the steadfast support of my parents, Celine and Kieran, I would not have had the courage, determination and motivation to complete this research project and for that, I am truly grateful. I thank them for the example of hard work and perseverance they set for me. During my PhD, I have had the pleasure of living in Oxford alongside my brother, Dónal and my sister-in-law, Mie. They have provided me with generous support and shared insights from their own experience which were incredibly valuable. I thank them for their continuous encouragement and guidance. Thanks also to my wider family circle for their ongoing support over the years to help me reach this point.

I want to thank the staff and my friends at Balliol College for providing a fun and friendly place to live and study. A special thanks to my fellow Interdisciplinary Bioscience students, especially Conor and Heather, who have been a great support and wonderful friends since day one.

I have greatly benefited from the welcoming environment of the Oxford Catholic Chaplaincy. My thanks go to all the chaplains, past and present, for the supportive and homely environment they have built. A special thanks to Carol, Megan, Mitch, Gabriel, James, Jean, Paul, John and many others, for making my time here in Oxford so enjoyable.

I would like to express my gratitude to my friends from home who, even from a distance, were always there to encourage and motivate me. Thank you for your visits to Oxford and for always providing fun and welcome distractions from my studies when I returned home for visits; this was always much appreciated.

I am deeply grateful to my Eglantine Avenue family, Nathan, Adam, Megan, James and Issy. I want to especially thank them for providing me with a safe home during a global pandemic and for the fun and laughter which kept me motivated whilst writing this thesis.

Last but certainly not least, I would like to show my great appreciation to my boyfriend, Jordan. I thank him for his unwavering support and understanding during my PhD studies and for providing me with boundless guidance and encouragement, during the good days and the tough days. I am particularly grateful for his belief in me and for the time he has taken to help me believe in myself.

Abstract

Introduction

Extracellular vesicles (EV) are membrane bound carriers released by cells to mediate important aspects of intercellular communication; however, the basic principles of EV-biogenesis, in particular the mechanisms regulating mRNA EV-loading, are poorly characterised. Furthermore, a limited repertoire of tools has thus far made these processes challenging to research.

Methods

To address these fundamental biological questions, an RNA-sequencing analysis of *Drosophila* EVs and the cells from which they derive was carried out. A bioinformatic pipeline was subsequently developed with the aim of identifying molecular features within the 3'UTR of mRNAs which may encourage their EV incorporation and long-range transfer. In addition, the role of these EV-enriched features was assessed *in vivo*, in complex *Drosophila* tissues, using an adapted version of the Cre-LoxP reporter to visualise long-range cell communication.

Results

These analyses revealed a distinct mRNA profile for *Drosophila* cells versus EVs and identified unique EV-enriched 3'UTR motifs. Motifs were preferentially arranged in distinct combinations and tended to occur in specific RNA secondary structures. These features may modulate their interaction with *trans*-acting factors, such as RNA binding proteins, with sequence similarity observed between these novel motifs and well-established RNA binding-protein binding sites. Subsequent *in vivo* analyses highlighted the potential for mRNA 3'UTR modifications to successfully modulate long-range mRNA transfer in complex tissues. EV-enriched motifs within these 3'UTR modifications, however, may only play a contributing role. Future work is needed to assess the effect of specific 3'UTR modifications on mRNA stability, length and GC content, as these factors have the potential to influence long-range mRNA transfer.

Conclusion

These investigations provide novel insights into the biological properties facilitating mRNA transfer between *Drosophila* cells without physical contact. Importantly, these biological features may be evolutionarily conserved, given that similar mRNA properties have been described in mammalian cells. Therefore, these findings could establish the basis to efficiently modulate mRNA transfer and cell function at a distance.

Impact of COVID-19

Before university buildings were closed to non-COVID-19 research in March 2020, I was generating the plasmids required for the development of *Drosophila* fly lines needed for the biological validation of the bioinformatically predicted EV motifs. Due to COVID-19, the delivery of these plasmids to the facility which generates fly lines was delayed and therefore, these new fly lines were not received until September 2020 (Cre-V5-paraWT and Cre-V5-paraA) or January 2021 (Cre-V5-paraC, Cre-V5-paraAC and Cre-V5-paraACD). Therefore, the *in vivo* biological validation of these EV motifs, described in **Chapter 5**, is currently ongoing and full details are not included in this thesis.

During this time, I was also completing the molecular biology required for the generation of a LexAop Cre RNAi fly line, needed to fully validate *in vivo cre* mRNA transfer and rule out the contribution of Cre protein transfer. These flies were also not received until January 2021 and therefore, experiments detailed in **Section 4.6.1** are currently ongoing.

In addition to the molecular biology carried out prior to university closures, I was also beginning to explore the contribution of caspase activity on reporter signalling and the long-range transfer of *cre* mRNA *in vivo*. Experiments targeting different levels of the caspase pathway (Hid, Reaper and Grim or Dronc) are currently ongoing and this area remains to be fully explored. Additional experiments described within **Section 5.5.4** aim to study the crosstalk between non-apoptotic caspase activity and long-range cell communication, more particularly in disease contexts.

Upon my return to wet-lab activities in late July 2020, limitations on lab access time hindered my progress. A major restriction during this time was my inability to continue working between the Dunn School of Pathology (where I primarily carried out fly related research) and Oxford Brookes (where much of my cell culture and EV extraction research was carried out). From July 2020 I was based fully in the Dunn School of Pathology and for this reason, much of the cell culture validation of the bioinformatically predicted EV motifs, described in **Chapter 5**, was carried out by, or in collaboration with, Elizabeth Dellar who was based at Oxford Brookes during this time.

Contributions

Footnotes throughout this thesis indicate where work has been carried out by, or in collaboration with, Elizabeth Dellar or Jana Terenova.

Contents

List of Figures xiii

List of Tables xviii

List of Abbreviations xx

1 Introduction 1

1.1 Intercellular communication is a fundamental aspect of multicellular life . . . 2

1.1.1 Soluble factor transfer as a means of paracrine or endocrine signalling . . . 3

1.1.2 Direct cell contacts mediate cell communication . . . 4

1.1.3 Cellular projections allow direct cell communication over long distances . . . 5

1.2 An introduction to extracellular vesicles . . . 6

1.2.1 Subtypes of extracellular vesicles . . . 7

1.2.2 Biogenesis pathways for extracellular vesicles . . . 10

1.2.3 Caspases and their non-apoptotic role in EV-biogenesis, loading and uptake . . . 15

1.2.4 Selective cargo loading into extracellular vesicles . . . 17

1.2.5 EV targeting to recipient cells . . . 26

1.2.6 EV uptake or signal transduction in recipient cells . . . 26

1.2.7 Functional effects of EV cargo in recipient cells . . . 27

1.2.8 Tools to study EV mediated communication and functional cargo transfer . . . 28

1.3 *Drosophila* as a model organism . . . 31

1.3.1 Imaginal wing disc structure . . . 34

1.3.2 Evolutionary conservation of EV-mediated cell communication in *Drosophila* . . . 35

1.4 Aims and objectives . . . 36

2 Methods 37

2.1 Generation of the Cre-LoxP reporter system . . . 38

2.1.1 Key molecular biology techniques . . . 38

2.1.2 Actin-loxP-myrGFP Reporter plasmid generation . . . 39

2.1.3 Cre-V5 and Cre-V5-25 plasmid generation . . . 39

2.1.4 Cre-V5-para construct generation to test the novel EV-enriched motifs . . . 40

2.2 Cell culture techniques . . . 41

2.2.1 Transfection . . . 41

2.2.2 Fixation and immunostaining 42
2.2.3 Reporter and Cre co-transfection experiments 43
2.2.4 Reporter and Cre Transwell assay 43
2.2.5 Optimisation of cell culture conditions for EV extraction . . 44
2.2.6 Preliminary EV extraction and characterisation 45
2.2.7 EV extraction from Cre transfected *Drosophila* S2R+ cells . 48
2.2.8 Cellular and EV RNA extraction 50
2.2.9 Quantitative Real-Time PCR 51
2.2.10 RNase protection assay 52
2.3 *Drosophila melanogaster* husbandry 53
2.3.1 Fly handling 53
2.3.2 Fly stocks 53
2.3.3 Experimental fly handling 54
2.3.4 Dissection, fixation and immunostaining 55
2.4 Microscopy and Image Analysis 56
2.4.1 Co-transfection experiments 56
2.4.2 Assessment of serum free and EV-depleted FBS media conditions 57
2.4.3 Transwell assay 58
2.4.4 Imaginal discs 58
2.5 Statistical analysis and visualisation 60
2.5.1 Analysing proportional data 60
2.6 Bioinformatic analysis 61
2.6.1 Biological sample preparation 61
2.6.2 Library preparation and RNA sequencing 61
2.6.3 RNA sequencing data analysis 62

3 Identification of novel EV-enriched motif combinations in *Drosophila melanogaster* 75
3.1 Introduction 76
3.2 Characterisation of cellular and EV RNA profiles 77
3.2.1 *Drosophila* cells and EVs showed distinct RNA profiles, with EV RNA profiles appearing more variable 79
3.2.2 Read coverage distribution for cellular and EV RNAs showed no distinct 5-prime or 3-prime bias 81
3.2.3 Cell and EV RNA reads predominantly mapped to coding regions 82
3.2.4 Comparing *Drosophila* cell and EV RNA profiles highlighted the presence of EV-specific RNA cargo 83
3.2.5 Cellular and EV RNAs were predominately protein-coding . 85

3.2.6 Specific transcripts are enriched in EVs compared to the cell 87
3.3 Novel 3'UTR motifs identified within EV-enriched mRNAs 90
3.3.1 Discovery of novel *Drosophila* EV-enriched motifs 90
3.3.2 The six novel *Drosophila* EV-enriched motifs were not analogous to one another 93
3.3.3 Novel *Drosophila* EV-enriched motifs showed no significant sequence similarity with known mammalian EV-enriched motifs 93
3.3.4 Each motif presented a unique distribution pattern along the 3'UTR 95
3.3.5 Motifs commonly occur in combinations within the 3'UTR . 96
3.3.6 3'UTR motifs were specifically arranged within EV mRNAs 98
3.3.7 RNA motifs were part of specific RNA secondary structures 100
3.3.8 RNA motifs potentially interact with specific RNA-binding proteins 101
3.3.9 Novel *Drosophila* motifs were more prevalent in longer transcripts 103
3.3.10 EV-enriched *Drosophila* motifs present in transcripts with specific gene product attribute predictions 105
3.3.11 Novel motif occurrences found within mRNA coding regions and the 5'UTR 107
3.4 Discussion 110
3.4.1 Characterisation of *Drosophila* S2R+ cell and EV RNAs . . 110
3.4.2 Mechanisms of selective EV-loading in *Drosophila* 113
3.5 Future Work 121
3.5.1 Determining RNA binding protein and miRNA interactions with EV-enriched motifs 121
3.5.2 Biological validation that the novel *Drosophila* EV-enriched motifs encourage the EV-loading of longer transcripts 122
3.5.3 Comparison of surface EV RNAs with those loaded within the EV lumen 122
3.6 Chapter Summary 124

4 Developing novel tools to study cell-to-cell communication 125
4.1 Introduction 126
4.2 Rational design of the Cre-LoxP based reporter system in *Drosophila* 126
4.3 Validation of the Cre-LoxP based reporter system in cell culture . . 128
4.3.1 Cre-LoxP reporter system functionality in *Drosophila* SR2+ cells 128
4.3.2 Validation of *cre* mRNA EV-loading in *Drosophila* cell culture 129

4.3.3 The *Drosophila* Cre-LoxP system is functional in cell culture 136
4.4 Validation of the Cre-LoxP reporter system *in vivo* 138
4.4.1 Transfer of Cre detected within complex tissues using the *in vivo Drosophila* Cre-LoxP reporter system 139
4.4.2 Cre transfer across the anterior/posterior boundary of the imaginal wing disc was limited to early larval development . 140
4.4.3 Cre transfer observed between different cell types in close proximity 142
4.4.4 The mammalian 25-nt EV motif did not alter reporter signalling *in vivo* 143
4.5 Discussion 146
4.5.1 A novel Cre-LoxP based reporter system to study EV mediated communication in *Drosophila* 146
4.5.2 Comparison of the *Drosophila* Cre-LoxP reporter system with other reporters of EV-mediated transfer 152
4.5.3 The mammalian 25-nt EV-loading motif did not influence *cre* mRNA loading in *Drosophila* 153
4.6 Future work 155
4.6.1 Further validation of *cre* mRNA EV-loading and transfer . . 155
4.7 Chapter Summary 158

5 Identifying novel factors involved in modulating long-range communication 159
5.1 Introduction 160
5.2 Adapting the Cre-LoxP system to validate novel *Drosophila* EV motifs 160
5.2.1 Determining the effect of the *Drosophila* EV-enriched motifs in a cell culture based Cre-LoxP system 162
5.2.2 Determining the effect of the *Drosophila* EV-enriched motifs in the *in vivo* based Cre-LoxP system 164
5.3 Investigating the impact of caspase activity on the Cre-LoxP reporter system *in vivo* 167
5.3.1 High levels of Cre expression increased caspase activity and enhanced reporter signalling *in vivo* 168
5.3.2 Enhanced reporter signalling *in vivo* is likely not the result of cell debris release 169
5.3.3 Inhibiting effector caspases modulated the activity of the Cre-LoxP reporter system *in vivo* 170
5.4 Discussion 173
5.4.1 Cell culture and *in vivo* analysis of the *Drosophila* EV-enriched motif combination 173

5.4.2 Using the *Drosophila* Cre-LoxP system to identify novel factors involved in EV-loading and secretion 175
5.5 Future work 178
5.5.1 Using the *Drosophila* Cre-LoxP system to identify novel motif interacting factors 178
5.5.2 Determining the effect of mRNA stability on EV-loading . . 178
5.5.3 Using the *Drosophila* Cre-LoxP system to identify novel proteins involved in EV uptake 179
5.5.4 Using the *Drosophila* Cre-LoxP system to study EV-mediated communication in a disease context 180
5.6 Chapter Summary 181

6 Discussion **182**
6.1 Thesis summary 183
6.2 Key findings 185
6.2.1 Bioinformatic analysis of *Drosophila* cellular and EV RNA profiles identified EV-enriched mRNAs 185
6.2.2 Combinations of novel mRNA features may interact to modulate EV-loading versus cell-retention 186
6.2.3 EV-enriched motifs are prevalent within longer transcripts . 186
6.2.4 EV-enriched motifs show significant sequence similarity with RBP binding sites 187
6.2.5 Atypical *Drosophila* motif identified which may be involved in alternative regulatory processes 188
6.2.6 Successful establishment of a *Drosophila* Cre-LoxP system to study cellular communication 189
6.2.7 Advantages of the *Drosophila* Cre-LoxP reporter system . . 190
6.2.8 Function of the novel *Drosophila* EV-loading motifs *in vivo* . 191
6.2.9 Modulating the caspase pathway altered the activity of the *Drosophila* Cre-LoxP reporter system 191
6.3 Concluding remarks 194

7 Supplementary **196**
7.1 Supplementary Information 197
7.1.1 Generation of *cre* fragments with the *paralytic* 3'UTR 197
7.1.2 Optimisation of conditions for *Drosophila* SR2+ cell culture, EV extraction and EV characterisation 200
7.2 Supplementary Figures 204
7.3 Supplementary Tables 213
7.4 Supplementary Listings 237

References **245**

List of Figures

1.1 Schematic diagram depicting different classes of cell communication. 3
1.2 EV release and uptake mechanisms 9
1.3 Biogenesis mechanisms for exosomes 11
1.4 Biogenesis mechanisms for microvesicles 14
1.5 Schematic diagram depicting the caspase cascade in *Caenorhabditis elegans*, vertebrates and *Drosophila melanogaster*. 17
1.6 Schematic diagram depicting the *Drosophila* life cycle. 32
1.7 Schematic diagram depicting the 19 *Drosophila* imaginal discs within the 3rd instar larvae life stage. 33
1.8 Schematic diagram depicting a top view and transverse cross section of an imaginal wing disc. 34

2.1 Transwell assay functionality 44
2.2 Schematic diagram describing the timeline experiment to control Cre expression during specific stages of development 55
2.3 Determining normalised percentage prevalence of motif combinations at the gene level 72
2.4 Determining all possible motif orders present within a 3'UTR . . . 73
2.5 Schematic depiction of the bioinformatic pipeline described in **Chapter 3**. 74

3.1 Bioanalyser traces for the cell and EV RNA samples isolated for bioinformatic analysis 78
3.2 Percentage of total RNA abundance made up of 18s, 28s and 5.8s rRNA species 79
3.3 Principal component analysis of cellular and EV RNAs 80
3.4 Gene body coverage for cell and EV samples 82
3.5 Read distribution within gene regions 83
3.6 zFPKM transformation of average Cell and EV FPKM values . . . 84
3.7 Percentage prevalence of transcripts within each transcript type category 86
3.8 Active genes encoding mRNAs within the 'Cell only', 'EV only' and 'Cell and EV' groups. 86
3.9 Differential expression analysis of genes within the 'Cell and EV' active gene group. 87
3.10 Significantly enriched cellular component gene ontology (GO) terms associated with genes encoding 'EV+' transcripts. 89
3.11 Sequence logos for six motifs identified via MEME and FIMO analysis as significantly enriched in 'EV+' transcripts when compared to 'Cell only' transcripts. 92

3.12 Percentage prevalence of motifs within the 'EV+' and 'Cell only' active gene lists. 92
3.13 TOMTOM alignment of the previously published Temoche-Diaz et al. (2019) and the 25-nucleotide mammalian motif 94
3.14 Distribution patterns of motif hits along the 3'UTRs of the 'EV+' genes. 96
3.15 Combinations of 3'UTR A, B, C and D motif occurrences 97
3.16 Number of A, B, C and D motif 3'UTR hits per shortlisted transcript 100
3.17 Percentage of bases paired per motif occurrence within the shortlisted transcript 3'UTRs. 101
3.18 Sequence similarity hits observed between the novel *Drosophila* motifs and a database of known RBP binding sites. 102
3.19 Sequence similarity hits observed between the novel *Drosophila* motif C and the known RNA-binding sites for Smooth and CNOT4. . . . 102
3.20 Length distribution of the median transcript length of the top 1000 most abundant genes (excluding *Arc1*) detected in EV and cell samples 103
3.21 Mean transcript 3'UTR length of genes containing any combination of the ABCD motifs. 104
3.22 Mean transcript length of genes containing any combination of the ABCD motifs. 104
3.23 Significantly enriched cellular component gene ontology (GO) terms associated with genes encoding transcripts carrying EV-motifs. . . . 106
3.24 Percentage prevalence of motif occurrences within untranslated and coding mRNA regions. 107
3.25 Sequence length of untranslated and coding mRNA regions. 109

4.1 Cre and reporter construct design 127
4.2 Schematic diagram to show EV transfer between a Cre donor cell and a recipient reporter cell 128
4.3 Cell states observed during co-transfection experiments within a single culture 129
4.4 Cell viability for Cre transfected cells during media conditioning period prior to EV extraction 130
4.5 Particle concentrations of EVs extracted from Media control, Ubi control, Ubi Cre-V5 and Ubi Cre-V5-25 conditioned media 130
4.6 Size distribution profile (15 to 375 nm) of EVs extracted from Ubi control, Ubi Cre-V5 and Ubi Cre-V5-25 conditioned media 131
4.7 Estimated particle size of EVs extracted from Ubi control, Ubi Cre-V5, Ubi Cre-V5-25 and media control conditions 132

4.8 qPCR data showing cellular and EV *cre* mRNA levels in Cre transfected cells . . . 134
4.9 Raw Ct values for qPCR data showing cellular and EV *cre* mRNA levels in Cre transfected cells . . . 135
4.10 RNase protection assay to detect luminal *cre* EV mRNA . . . 136
4.11 Quantified cell states during co-transfection experiments within a single culture . . . 137
4.12 Lineage tracing of the ptcGal4 driver . . . 139
4.13 *In vivo* Cre-LoxP system reporting Cre transfer across the IWD anterior/posterior boundary. . . . 140
4.14 Proportion of Reporter^{GFP+} signal during timeline experiment with ptcGal4 . . . 141
4.15 Eye/antennal disc Reporter^{GFP+} signal when Cre was expressed in hemocytes . . . 142
4.16 Proportion of IWD posteriors showing any Reporter^{GFP+} signal for the Cre-V5 and Cre-V5-25 conditions . . . 143
4.17 Percentage area of IWD posterior Reporter^{GFP+} signal for the Cre-V5 and Cre-V5-25 conditions . . . 144
4.18 Proportion of eye/antennal (E/A) discs showing Reporter^{GFP+} signal when Cre was expressed in hemocytes . . . 145
4.19 Schematic diagram depicting anterior/posterior (A/P) boundary during stages of *Drosophila* imaginal wing disc development. . . . 150
4.20 Schematic diagram depicting direction of growth during *Drosophila* imaginal wing disc development. . . . 151
4.21 Cre RNAi *in vivo* experimental design. . . . 156

5.1 Adapted Cre constructs with the wild-type and shuffled versions of the *paralytic* 3'UTR . . . 161
5.2 Predicted secondary structure of wild-type and shuffled versions of *paralytic* 3'UTR regions . . . 161
5.3 Fold differences observed in *cre* mRNA cellular and EV abundances for Cre-V5, Cre-V5-25, Cre-V5-paraWT and Cre-V5-paraACD conditions, determined via qPCR. . . . 163
5.4 Frequency and percentage area of posterior Reporter^{GFP+} signalling events under Cre-V5, Cre-V5-25, Cre-V5-paraWT and Cre-V5-paraACD conditions. . . . 165
5.5 Schematic diagram depicting the posterior regions becoming Reporter^{GFP+} under Cre-V5, Cre-V5-25 and Cre-V5-paraWT conditions. . . . 166
5.6 Caspase cascade and inhibition points . . . 167

5.7 Cre toxicity in the IWD using the ptcGal4 driver in the absence of TubGal80^{ts} . . . 168
5.8 Long-range Reporter^{GFP+} signalling was not observed when Cre was expressed within the IWD wing pouch using the spaltGal4 driver. . 170
5.9 Posterior Reporter^{GFP+} signalling observed when UAS Cre-V5-25 and UAS p35 co-expressed in the IWD . . . 171
5.10 Percentage of IWD posteriors showing Reporter^{GFP+} signal in the absence or presence of p35. . . . 171
5.11 No posterior Reporter^{GFP+} signal was observed when UAS Cre-V5-25 and UAS p35 were co-expressed during late larval development. . . 172

6.1 Possible mechanism of action of caspase activity for modulating RNA and RBP associations to influence RNA EV incorporation. 193

7.1 GENEWIZ construct containing *cre* along with the *paralytic* 3'UTR in wild-type (WT) form or with the A motif mutated. 197
7.2 GENEWIZ construct containing *cre* along with the *paralytic* 3'UTR with the ACD motifs mutated or the AC motifs mutated. 198
7.3 GENEWIZ construct containing *cre* along with the *paralytic* 3'UTR with the C motif mutated or the B motif mutated. 199
7.4 Cell culture conditioning trial . . . 201
7.5 Nanosight Nanoparticle Tracking Analysis and protein concentration analysis of SEC fractions 1 to 20 from *Drosophila* S2R+ cell conditioned media. . . . 202
7.6 TEM of *Drosophila* S2R+ EVs . . . 203
7.7 Western Blots of *Drosophila* EVs . . . 203
7.8 MultiQC data . . . 204
7.9 MultiQC data for duplicated and overrepresented sequences . . . 205
7.10 Pairwise Spearman correlation coefficients (ρ) between normalised read count values obtained for cell and EV sample replicates . . . 205
7.11 Mean read length of paired cell and EV RNA-sequencing reads. . . 206
7.12 Percentage prevalence of motifs within the 2-fold significantly enriched or depleted genes in EVs relative to cell. . . . 206
7.13 Motif start and end positions for all ABCD transcripts. . . . 207
7.14 Motif start and end positions for the shortlisted ABCD transcripts . 208
7.15 Size distribution profile (ranging from 15 to 1005 nm) for EVs extracted from Ubi control, Ubi Cre-V5 and Ubi Cre-V5-25 conditioned media . . . 209
7.16 Proportion of imaginal wing discs presenting Reporter^{GFP+} signal positive staining in the absence of a Gal4 driver . . . 209

7.17 Proportion of Reporter^{GFP+} signal in the IWD posterior respecting the 'gap region' 210
7.18 Percentage area of IWD posterior Reporter^{GFP+} signal when Cre-V5-25 is expressed throughout larval development, or during early development stages only. 210
7.19 Snapshot from the neighbour-joining phylogenetic tree generated from a ClustalW alignment of human miR-1289 and all *Drosophila* miRNAs 211
7.20 Fold differences observed in *cre* mRNA cellular and EV abundances, relative to matched Cre-V5 cell condition. 211
7.21 Quantification of the rate of Reporter^{GFP+} cells in the Transwell assay 212
7.22 Percentage area of IWD posterior Reporter^{GFP+} signal when Cre-V5-25 is expressed in combination with p35. 212

List of Tables

1.1 Description of common signalling related cellular projections. 5

2.1 PCR primers used to generate the Cre-V5 and Cre-V5-25 constructs 40
2.2 Amount of DNA used for co-transfection experiments with Reporter and Cre (or control) plasmids. 42
2.3 Antibody suppliers and their required dilutions. 42
2.4 Amount of DNA used for Transwell assay experiments with Reporter and Cre (or control) plasmids 44
2.5 Western blotting antibody suppliers and their required dilutions. . . 48
2.6 qPCR primers and literature sources. 52
2.7 Oxford Fly Food Ingredients 53
2.8 Cre and Reporter *Drosophila* fly lines generated for *in vivo* experiments. 53
2.9 *Drosophila* fly lines utilised in *in vivo* experiments. 54
2.10 Percentage of rRNA contamination 63

3.1 RNA concentrations and absorbance ratios obtained for cell and EV RNA samples. 77
3.2 Number of active genes in the three groups: 'Cell and EV', 'Cell only' and 'EV only'. 85
3.3 Number of active protein-coding genes in the 'Cell only' and 'EV only plus EV enriched' groups. 87
3.4 Number of mRNA transcripts with a 3'UTR sequence available and those 3'UTR sequences greater than 30 nucleotides long for the 'Cell only' and 'EV only plus EV enriched' groups. 91
3.5 Motif similarity matrix generated via a MAST analysis 93
3.6 Previously published EV-enriched motif sequences from miRNA and mRNA transcripts. 94
3.7 Order of motif occurrences present in the shortlisted ABCD 'EV+' transcripts. 99

7.1 Summary of cell junctions utilised for cell communication via the transfer of molecules or transduction of signals 213
7.2 Mechanisms of EV uptake or signal transduction in recipient cells. . 215
7.3 ESCRT components plus accessory factors and their involvement in ILV formation and cargo loading 217
7.4 Read alignment metrics for each cellular and EV RNA replicate, together with means and standard deviations. 217
7.5 Sidak's multiple comparisons test to compare the difference in relative abundance of genes aligning to particular gene regions in our cell and EV samples (**Figure 3.5**). 218

7.6 Percentage of transcripts within each transcript type category for 'Cell only', 'Cell and EV' and 'EV only' active gene groupings. . . . 218
7.7 Clustered biological process gene ontology terms significantly enriched in 'EV+' genes. . . . 220
7.8 Clustered cellular component gene ontology terms significantly enriched in 'EV+' genes. . . . 221
7.9 Clustered molecular function gene ontology terms significantly enriched in 'EV+' genes. . . . 221
7.10 Percentage prevalence of motifs within the 'EV only plus EV enriched' and 'Cell only' active gene lists. . . . 222
7.11 Dunn's multiple comparisons test to compare the distribution of normalised 3'UTR position hit for each motif (**Figure 3.14**). . . . 223
7.12 Skewness and standard error values determined for the distribution of normalised 3'UTR position hit for each motif (**Figure 3.14**). . . 223
7.13 Dunn's multiple comparisons test to compare the percentage of paired bases per motif occurrence (**Figure 3.17**). . . . 223
7.14 ExoCarta database literature hits from human or mouse models obtained when searching for human orthologs of *Drosophila* RNA binding proteins which have a binding motif significantly similar to the novel *Drosophila* EV-enriched motifs. . . . 225
7.15 Clustered biological process gene ontology terms significantly enriched in genes encoding transcripts containing 3'UTR ABCD motif combinations. . . . 229
7.16 Clustered cellular component gene ontology terms significantly enriched in genes encoding transcripts containing 3'UTR ABCD motif combinations. . . . 232
7.17 Clustered molecular function gene ontology terms significantly enriched in genes encoding transcripts containing 3'UTR ABCD motif combinations. . . . 235
7.18 Novel motif occurrences within 5' untranslated and coding mRNA regions. . . . 236
7.19 Number of genes and transcripts for which 5'UTR, CDS and 3'UTR sequences were available from Ensemble Biomart and used for FIMO motif analysis. . . . 236

List of Abbreviations

25-nt . . . . . Mammalian 25-nucleotide motif for EV-loading
3' 3-prime
5' 5-prime
Ago2 . . . . . Argonaute 2
ALIX . . . . . ALG-2 interacting protein X
A/P . . . . . Anterior/posterior boundary of the imaginal wing disc
ApoEV . . . . . Apoptotic EV
ARF . . . . . ADP-ribosylation factor
ARS . . . . . Aminoacyl-tRNA synthetase
Bcl-xL . . . . . B-cell lymphoma-extra large
Cas9 . . . . . Cas9 nuclease
CNOT4 . . . . . CCR4-NOT transcription complex subunit 4
Cre . . . . . Cre recombinase
CRISPR . . . . . Clustered regularly interspaced short palindromic repeats
DCP-1 . . . . . Death caspase-1, used as marker for caspase activity / cell death
DE . . . . . Differential expression
Dpp . . . . . Decapentaplegic
E/A ID . . . . . Eye-Antennal imaginal disc
ECM . . . . . Extracellular matrix
EM . . . . . Electron microscopy
ESCRT . . . . . Endosomal sorting complex required for transport
EV . . . . . Extracellular vesicle
EV+ . . . . . EV-only plus EV-enriched transcripts
exRNA . . . . . Extracellular RNA
FBS . . . . . Fetal bovine serum
FMR1 . . . . . Fragile X mental retardation 1
FPKM . . . . . Fragments Per Kilobase of transcript per Million mapped reads
GO . . . . . Gene ontology
Hh . . . . . Hedgehog
hml . . . . . Hemolectin
hnRNPA2B1 . . Heterogeneous nuclear ribonucleoprotein A2B1
hnRNPK . . . . . Heterogeneous nuclear ribonucleoprotein K
hnRNPL . . . . . Heterogeneous nuclear ribonucleoprotein L
hnRNPU . . . . . Heterogeneous nuclear ribonucleoprotein U
Hrs . . . . . Hepatocyte growth factor-regulated tyrosin kinase substrate
ILV . . . . . Intraluminal vesicle
IWD . . . . . Imaginal wing disc
KRAS . . . . . Kirsten rat sarcoma 2 viral oncogene homolog
KRS . . . . . Lysyl-tRNA synthetase
lncRNA . . . . . long non-coding RNA

miRNA . . . . . micro RNA
MLCK . . . . . Myosin light-chain kinase
mRNA . . . . . messenger RNA
MT . . . . . Microtubule
MVB . . . . . Multivesicular body
ncRNA . . . . . non-coding RNA
nm . . . . . nanometers
ng . . . . . nanograms
n.s. . . . . . Not significant
NSUN2 . . . . . NOP2/Sun RNA Methyltransferase 2
NTA . . . . . Nanoparticle Tracking Analysis
nts . . . . . Nucleotides
PC . . . . . Phosphatidylcholine
PCA . . . . . Principal Component Analysis
PE . . . . . Phosphatidylethanolamine
PH3 . . . . . Phospho-Histone H3, used as a marker for proliferation
piRNA . . . . . piwi-interacting RNA
PS . . . . . Phosphatidylserine
ptc . . . . . Patched
PTMs . . . . . Post-transcriptional modifications
qPCR . . . . . Quantitative real-time polymerase chain reaction
RBP . . . . . RNA-binding protein
RISC . . . . . RNA-induced silencing complex
rRNA . . . . . ribosomal RNA
RNAi . . . . . RNA interference
RNP . . . . . Ribonucleoprotein
SEC . . . . . Size-exclusive chromatography
siRNA . . . . . Small interfering RNA
SM . . . . . Smooth
SMase . . . . . Sphingomyelinase
SNARE . . . . . Soluble N-ethylmaleimide-sensitive factor attachment protein receptor
snRNA . . . . . small nuclear RNA
snoRNA . . . . . small nucleolar RNA
SPH . . . . . Sphingomyelin
SYNCRIP . . . . Synaptotagmin-binding cytoplasmic RNA interacting protein, also known as hnRNP-Q
TSG101 . . . . . Tumor Susceptibility 101
tRNA . . . . . transfer RNA
TNT . . . . . Tunnelling nanotube

UAS . . . . . Upstream activated regulatory sequence promoter
Ubi . . . . . Ubiquitin promoter
UTR . . . . . Untranslated region
µg . . . . . micrograms
µm . . . . . micrometers
µl . . . . . microliter
VAP-A . . . . . Vesicle-associated membrane protein-associated protein-A
Vps . . . . . Vacuolar protein sorting-associated protein
Wg . . . . . Wingless
WT . . . . . Wild-type
Y-BOX1 . . . . . Y box binding protein-1, also known as YB-1

1 Introduction

Contents

1.1 Intercellular communication is a fundamental aspect of multicellular life 2
1.1.1 Soluble factor transfer as a means of paracrine or endocrine signalling 3
1.1.2 Direct cell contacts mediate cell communication 4
1.1.3 Cellular projections allow direct cell communication over long distances 5
1.2 An introduction to extracellular vesicles 6
1.2.1 Subtypes of extracellular vesicles 7
1.2.2 Biogenesis pathways for extracellular vesicles 10
1.2.3 Caspases and their non-apoptotic role in EV-biogenesis, loading and uptake 15
1.2.4 Selective cargo loading into extracellular vesicles 17
1.2.5 EV targeting to recipient cells 26
1.2.6 EV uptake or signal transduction in recipient cells . . . 26
1.2.7 Functional effects of EV cargo in recipient cells 27
1.2.8 Tools to study EV mediated communication and functional cargo transfer 28
1.3 *Drosophila* as a model organism 31
1.3.1 Imaginal wing disc structure 34
1.3.2 Evolutionary conservation of EV-mediated cell communication in *Drosophila* 35
1.4 Aims and objectives 36

1.1 Intercellular communication is a fundamental aspect of multicellular life

Cells within multicellular organisms do not behave as individual entities, but form 'microsocieties' where precise signalling events between cells and tissues allow co-ordinated responses to be established [1]. Such co-ordinated responses govern vital cellular processes such as growth, development, metabolism, motility and immunity. Intercellular communication can be modulated by both physiological and pathological scenarios, with dysfunction of these pathways associated with a number of diseases, such as cancer and neurodegeneration [1, 2]. Due to the complexity of the diverse interacting pathways which regulate cell communication events, much work in this field has harnessed genetically compact model organisms, such as *Drosophila melanogaster* or *Caenorhabditis elegans*. These studies have accelerated this research field and have provided novel insights which have been shown to be relevant in mammalian systems.

Classically, intercellular communication events can be classified as either direct or indirect [2]. Direct communication can come in the form of self-communication, involving only the 'donor' cell itself (either intracrine or autocrine), or adjacent communication, where signals are transferred between cells via direct cell contacts (juxtacrine). Indirect communication involves the transfer of secreted signals, such as soluble factors, either locally (paracrine) or at a distance via the blood stream (endocrine). In recent years, the influence of alternative indirect cell communication mechanisms, such as the transfer of extracellular vesicles (EVs), has been more widely explored (**Figure 1.1**).

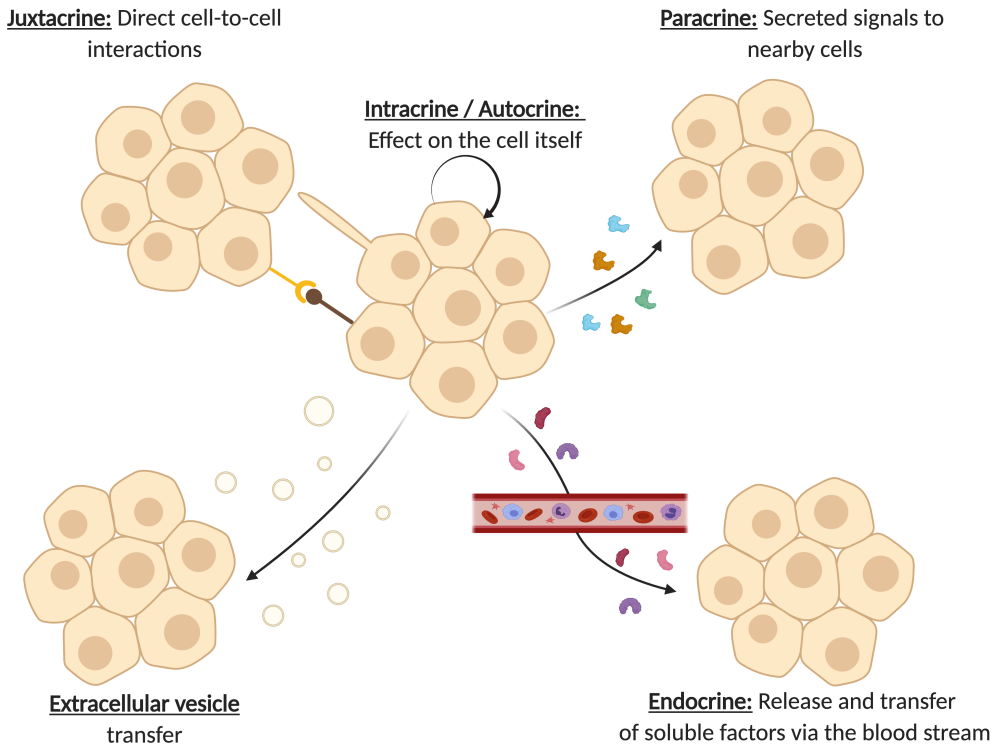


Figure 1.1: Schematic diagram depicting different classes of cell communication. Intracrine signalling is a direct form of self-communication where signals remain within the cell to influence the donor cell itself. Juxtacrine signalling is a direct form of adjacent communication where signals are sent between cells via direct cell contacts. Paracrine signalling is an indirect form of local communication where secreted signals are sent to a different cell nearby. Endocrine signalling is an indirect form of long-range communication where secreted signals are sent to a distant target via the blood stream. Extracellular vesicles are membrane bound carriers which facilitate the indirect signalling between cells which are not in contact. Information extracted from Brucher et al. (2014) [2].

1.1.1 Soluble factor transfer as a means of paracrine or endocrine signalling

Paracrine signalling was the first example of non-autonomous signalling to be described. Spemann and Geinitz (1927) showed that in amphibian embryos, inducers were passively diffused to determine patterning in the surrounding cells [3, 4]. Growth factors, morphogens, cytokines and chemokines are often secreted to form a concentration gradient as they disperse from their secretion site; the response of target cells is often concentration dependent [4, 5]. Hedgehog (Hh) and Wingless (Wg) are two key morphogens secreted to form concentration gradients. Hh was first discovered in *Drosophila*, where it was shown to be involved in segment polarity

[6]. Wg was also shown to be involved in segment polarity within the *Drosophila* model, where it was described as a homolog of the mouse mammary oncogene, int-1 [7]. Since then, these molecules have been shown to play important roles in tissue patterning, development, homeostasis and regeneration, across multiple species [8, 9].

Soluble factors can also function at a distance via the circulatory system. This mechanism is known as endocrine signalling, where soluble factors are referred to as hormones. For example, insulin and glucagon signalling regulates glycogen metabolism and glucose levels [10], with similar mechanisms identified in other species, such as insulin-like peptides (dILPs) and adipokinetic hormone (AKH) signalling to modulate growth and energy homeostasis in *Drosophila* [11]. Identifying the evolutionary conservation of key cell communication mechanisms highlighted their functional importance in cellular processes such as development and tissue homeostasis.

1.1.2 Direct cell contacts mediate cell communication

Cells can also transfer factors and transduce signals via direct cell contacts, such as gap junctions or ion channels, summarised in **Supplementary Table 7.1**. A common example of contact dependent signalling is the Notch pathway. This pathway, originally discovered and extensively explored in *Drosophila*, plays vital roles in cell-fate decision making and tissue patterning during development [12, 13]. Interestingly, both the Notch receptor and its ligands, Delta and Serrate, are membrane bound and therefore, close cell contact and precise membrane orientation is required to induce signal transduction in recipient cells [13, 14]. This spatiotemporal sensitivity ensures signalling events occur at defined locations and timings during development [13]. Typically, juxtacrine signalling occurs between adjacent cells; however, evidence has also shown that events can occur between distant cells in a tissue via the extension of cellular projections [14–16].

1.1.3 Cellular projections allow direct cell communication over long distances

Cellular projections are extensions of the cell membrane which facilitate long-range cellular signalling. Multiple forms of cellular extensions have been described, such as filopodia, cytonemes, microtubule (MT)-nanotubes and tunnelling nanotubes (TNTs) [17, 18]; each with their own characteristic features (**Table 1.1**). Filopodia, first described in sea urchins in 1995 [19], are thin, closed-ended, actin-based cellular projections which vary in diameter and in length [20]. Cytonemes, first discovered in *Drosophila*, are structurally similar to filopodia but are polarised in nature [21]. MT-nanotubes are closed-ended, microtubule-based extensions [22], with TNTs being actin-based, open-ended connections which act like bridges between cells [23]; however, TNTs also containing microtubules have been described [24].

Projection	Description	Approx. Diameter (nm)	Approx. Length (µm)	Ref.
Filopodia	closed-ended actin-based	100-400	1-200	[20]
Cytonemes	closed-ended actin-based polarised	100–200	2–150	[21, 25]
Microtubule (MT)-nanotubes	closed-ended microtubule-based	400	3	[22]
Tunnelling nanotubes (TNTs)	open-ended actin based (some microtubule content has been described)	50-200	6-120	[23, 24, 26]

Table 1.1: Description of common signalling related cellular projections. Approx. = Approximate. Ref. = References.

Cellular projections have been associated with processes such as stem cell signalling [22], pattern formation [15, 16], morphogenesis [27], establishment of morphogen gradients (e.g. Hh and Decapentaplegic (Dpp)) [21, 28–31] and morphogen signal reception (e.g. Wg) [32]. Cellular projections function by transferring signalling cargo or cytoplasmic contents between cells [18] or by facilitating distinct contact sites for ligand-receptor binding in a 'synapse-like' manner [29, 33].

In concordance with their 'synapse-like' function, cytonemes and filopodia can act as sites for extracellular vesicle (EV) budding and release. Similarly to the vesicular release of Wg from neurons at the neuromuscular junction (a synapse between a motor neuron and a muscle fibre) observed in *Drosophila* [34], morphogen release via cellular projections has also been described in other contexts [20, 35]. Vesicles have been shown to be released from cellular extensions to aid the restricted distribution of morphogens in *Drosophila* [36] and in the frog, *Xenopus laevis* [37]. In addition, filopodia are involved in vesicle uptake, providing cell projections for vesicles to 'surf' along prior to their internalisation [38]. These examples highlight how cells can simultaneously use multiple cell communication methods to achieve the precision and direction required to co-ordinate complex cellular processes, such as tissue patterning and development.

1.2 An introduction to extracellular vesicles

Extracellular vesicles (EVs) are a heterogeneous group of membrane bound carriers which are often classified based on their biogenesis pathway, size and origin. Based on these principles, EVs are commonly subdivided into three main categories: exosomes (40-100 nm), microvesicles (100-1000 nm) and apoptotic bodies (1000-5000 nm) [39]. EVs can be found in many biological fluids such as urine [40], breast milk [41], platelets [42] and blood plasma [43]. EV-mediated transfer of protein, lipid and RNA cargo is important for cell communication [44] in the context of neuronal communication [45], development [46, 47], immunity [48], cancer progression [49, 50] and inflammation [51].

EVs were potentially identified as early as the 1940s, when Chargaff and West described platelet-derived particles in plasma [52]. These particles were later described as 'platelet dust' by Wolf in 1976 [53]. Additional early sightings were potentially made by Anderson in 1969, when cellular vesicles were described as 'matrix vesicles' [54]. Matrix vesicles are extracellular, membrane-bound microparticles which anchor to the extracellular matrix to act as mineralization sites in tissues such as bone.

Whilst matrix vesicles and EVs may differ in function, their size, morphology and contents indicate that they may be homologous structures [55].

The physiological relevance of EVs, particularly *in vivo*, however, was not fully understood during these early investigations. Studies into their relevance intensified in the late 1970s and early 1980s. The term 'exosome' was coined during this period and their potential to mediate cell communication at a distance was first realised [56]. One early function observed for EVs was their involvement in transferrin receptor recycling and shedding during the differentiation of immature red blood cells in various species [57, 58], also later demonstrated using electron microscopy (EM) [59, 60]. In 1996, Raposo et al. showed that exosomes isolated from Epstein-Barr virus-transformed B lymphocytes played roles in antigen-presentation and induction of a T cell response [61].

For many years, EVs were also seen as a means for cells to remove redundant or 'junk' cellular content [62], maintaining donor cell homeostasis. However, a deeper exploration of EV cargo from the early 2000s onwards suggested their contents may, at least to some extent, be selectively loaded to carry out specific functions in recipient cells [63–70]. Additionally, rapid clearance of specific cargo via EVs may also be a way to activate particular cellular processes within the donor cell itself [71]. Many studies have focused on uncovering the mechanisms by which cargos become selectively loaded into EVs (see **Section 1.2.4**), and more specifically into particular EV subtypes [72–79]. An improved understanding of these processes would provide insights into novel functions for EVs in donor and recipient cells, which may also identify potential systems which can be harnessed for the manipulation of EV cargo loading for biological or therapeutic purposes.

1.2.1 Subtypes of extracellular vesicles

Three common subtypes of EVs, namely exosomes, microvesicles and apoptotic bodies, have defined cellular origins (**Figure 1.2**). Exosomes are derived from the intracellular endocytic pathway. Via the invagination of the plasma membrane, endosomes form within the cell. When subsequent invagination of the endosomal

membrane occurs, intraluminal vesicles (ILVs) are generated, after which the endosome is referred to as a multivesicular body (MVB). These MVBs can be sent to the lysosome for degradation or can be involved in protein recycling to other cellular locations [80, 81]. Alternatively, MVBs can fuse with the plasma membrane where the ILVs are released extracellularly as exosomes. Compared to other types of EVs, exosomes are often homogeneous in size [82], ranging from 40 to 100 nm in diameter [39]. Microvesicles, however, originate directly from the plasma membrane via outward budding and they are often more variable in size, ranging from 100 to 1000 nm in diameter [39].

Apoptotic cells generate the third category of EVs which are collectively referred to as Apoptotic EVs (ApoEVs) [83]. Apoptotic cells undergo a range of morphological changes, such as nuclear and cytoplasmic condensation, as well as the breaking up of the cell into membrane bound fragments [84]. Membrane blebbing results in the formation of protrusions resembling *'beads-on-a-string,'* with the 'beads' being released as apoptotic bodies [85]. Apoptotic bodies are often large EVs, ranging from 1000 to 5000 nm in diameter [39]. Apoptotic cells, however, can also release smaller EVs via membrane budding, such as ApoMVs, with a diameter of less than 1000 nm [83].

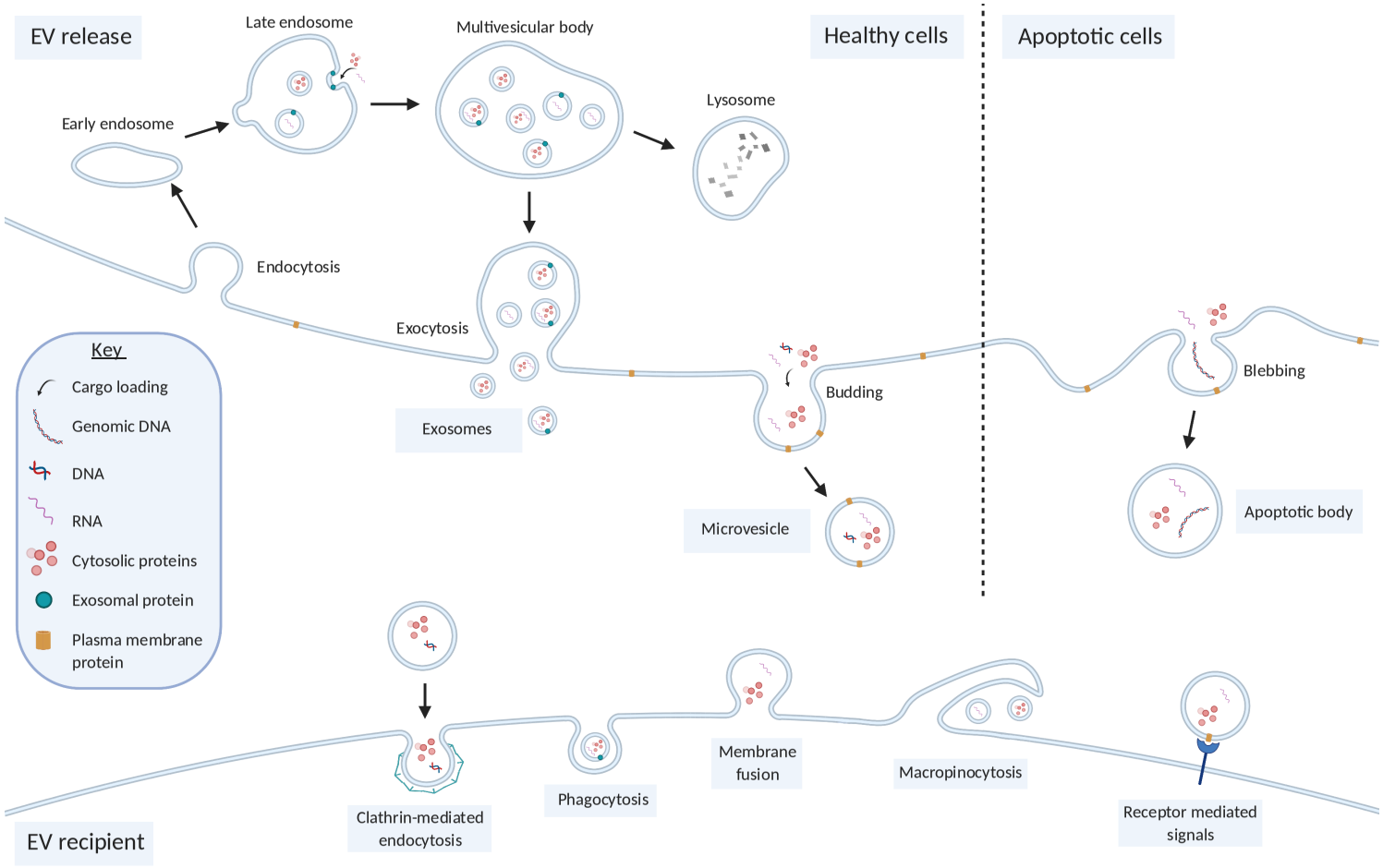


Figure 1.2: Schematic diagram depicting the release pathways for subtypes of EVs, together with recipient cell uptake and signalling mechanisms [86]. More detail on EV uptake and signalling mechanisms is included in **Supplementary Table 7.2**.

1.2.2 Biogenesis pathways for extracellular vesicles

Exosome biogenesis

Due to the different cellular origins of EV subtypes, EVs are generated via distinct biogenesis pathways. For exosome generation, one pathway depends on the endosomal sorting complex required for transport (ESCRT) machinery (ESCRT-dependent), whilst the other does not (ESCRT-independent) (**Figure 1.3**). The ESCRT machinery plays a role in a wide range of cellular processes, such as viral release, plasma membrane wound repair and neuronal pruning [87]. It is made up of four main complexes (ESCRT-0, -I, -II and -III) and associated proteins, such as vacuolar protein sorting-associated protein 4 (Vps4) and ALG-2 interacting protein X (ALIX) [82, 87]. In 2001, components of the ESCRT machinery were shown to be present within exosomes [88] and were later shown to be involved in the exosome biogenesis and protein-loading processes [63–67, 89].

ESCRT-0, composed of heterodimers of Hrs (hepatocyte growth factor-regulated tyrosin kinase substrate) and Stam subunits [90], is involved in protein recruitment to the endosomal membrane [91–93]. Although ESCRT-0 is not considered a core component of the membrane budding machinery [87], Hrs is indirectly involved in MVB formation due to its role in the recruitment of ESCRT-I subunits, such as TSG101 [94]. ESCRT-I is involved in the ubiquitin-dependent sorting of proteins into MVBs [95]. ESCRT-II and -III are key factors required for the recruitment of ubiquitinated cargo [96–98], with additional roles for ESCRT-III in membrane deformation [99]. Vps4, an AAA ATPase (ATPase Associated with diverse cellular Activities), is also required for cargo delivery [100], with additional roles in the disassembly of the ESCRT-III complex [99, 101, 102]. ALIX provides an alternative branch to the ILV formation pathway, bypassing the early ESCRT machinery but converging on the ESCRT-III component [66, 87, 103, 104]. Since these early studies, the roles for ESCRT components and associated factors in exosome formation have been more deeply explored and more details on these functions are summarised in **Supplementary Table 7.3**.

Exosomes can form in the absence of the ESCRT machinery [105] and in these scenarios, tetraspanins play a critical role [106]. Tetraspanins are a superfamily of small, four transmembrane domain proteins, frequently localised to endosomal membranes to regulate protein trafficking and commonly used as exosomal markers [106, 107]. Numerous studies have implicated various tetraspanins (such as CD9, CD82 and CD63) in the formation of exosomes and the loading of protein cargo in a range of cell lines [108–110].

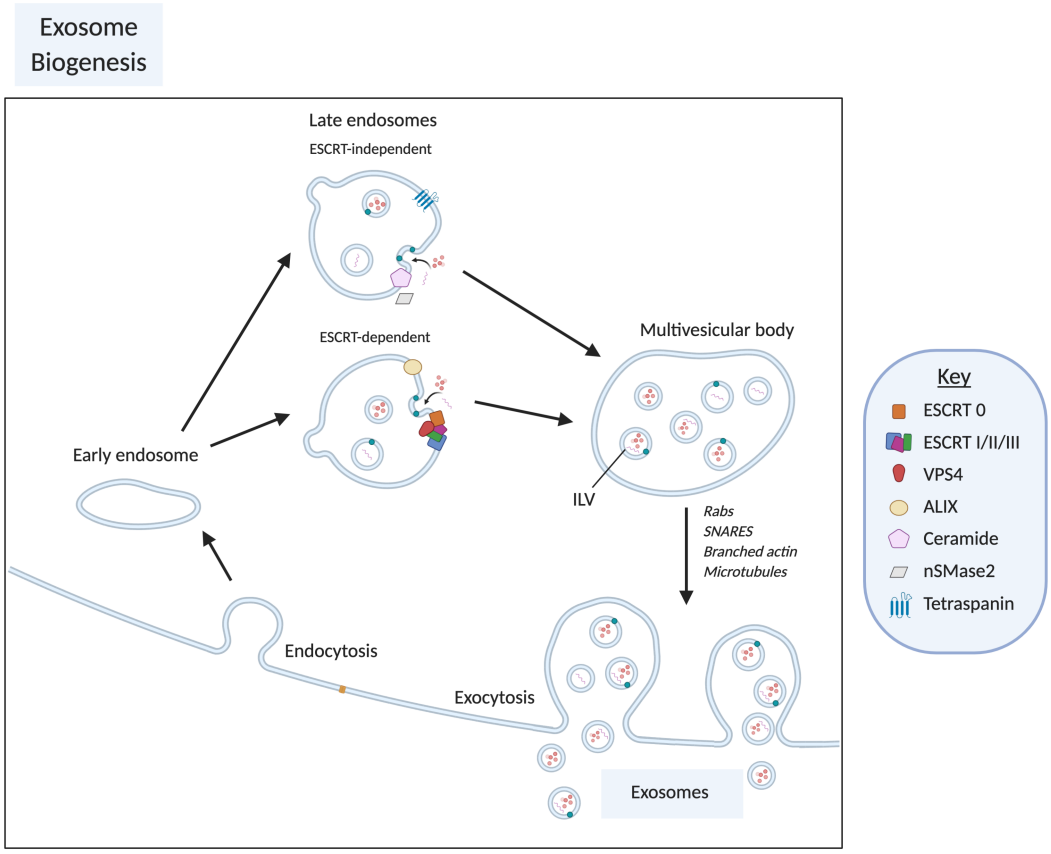


Figure 1.3: Schematic diagram depicting the biogenesis mechanisms for exosomes, highlighting key factors involved in these processes. ILV = Intraluminal vesicle. [86, 111–115].

The process of ILV formation for liposomal degradation is induced via lysobisphosphatidic acid (LBPA); however, this lipid is absent within exosomes [116, 117]. Therefore, cells may achieve ILV formation for exosome secretion via an alternative mechanism using the sphingolipid, ceramide. EVs are enriched in ceramide and the release of exosomes can be reduced after the inhibition of neutral sphingomyelinase

2 (nSMase2), an enzyme required to generate ceramide from sphingomyelin (SPH) [103, 118]. Lipid and ceramide rich regions appear within endosomal membranes, with ceramide self-associating to influence budding [117]. The cone shaped nature of ceramide may also influence negative curvature of endosomal membranes to influence ILV formation [118]. Not only do ceramide and lipid rich regions influence budding, they are also involved in cargo sorting [118], explored later in **Section 1.2.4**.

Once the ILVs have been formed, additional factors such as coat proteins, Rabs and SNAREs (soluble N-ethylmaleimide-sensitive factor attachment protein receptors) are involved in transporting the MVB to the plasma membrane [112]. This occurs via the interaction with actin or the microtubule cytoskeleton [115, 119], with the aid of molecular motors such as myosin, kinesin and dynein [120–122]. Rabs are small GTPases of the Ras superfamily which cycle between the cytoplasm and membranes [119]. Studies have implicated Rabs in the trafficking and secretion of vesicles in yeast, *Drosophila* and mammalian studies [46, 123–127]. Disrupting the activity of a single Rab protein, however, does not completely block exosome secretion, suggesting that redundancy may exist. Alternatively, each Rab may be responsible for the secretion of a particular exosome subpopulation [112], with future work needed to fully understand the complexity of these biological processes.

SNAREs are proteins which facilitate the tethering and fusion of two membranes. Vesicle SNAREs (vSNAREs) are present on the MVB which associate with target SNAREs (tSNAREs) on the plasma membrane to facilitate fusion and the release of ILVs as exosomes [119]. Although much of the machinery involved in this process remains elusive, SNAREs such as VAMP7, YKT6 and Syntaxin 1A have been highlighted as possible regulators of exosome release in mammalian or *Drosophila* systems [47, 125, 128]. The diverse range and influence of SNAREs throughout the endocytic and secretory pathways has, thus far, made it difficult to define specific functions in exosome secretion and more research is required in this area [112, 129].

Microvesicle biogenesis

Microvesicles (MVs) are an alternative subtype of EV which are generated via direct budding of the plasma membrane (**Figure 1.4**). A key factor regulating MV budding is the lipid composition of the inner and outer leaflets of the plasma membrane bilayer. Generally, the aminophospholipids phosphatidylserine (PS) and phosphatidylethanolamine (PE) are present within the inner leaflet, with phosphatidylcholine (PC) and sphingomyelin (SPH) enriched in the outer leaflet [130, 131]. Three components regulate lipid movement between the membrane leaflets: a flippase called aminophospholipid translocase (regulating inward lipid movement), a floppase (regulating outward lipid movement) and scramblase (regulating unspecific, bidirectional movement) [130, 132–134]. Upon stimulation, an increase of cytosolic Ca^{2+} levels causes flippases to be inhibited, whilst floppases and scramblases are activated, promoting outward lipid movement and the accumulation of PS on the outer leaflet [135]. Ca^{2+} levels also regulate the degradation of the cytoskeleton to promote MV release. In addition, membrane reorganisation also leads to lipid and protein accumulation into raft regions which promote specific cargo loading [130].

The ADP-ribosylation factor (ARF) family are part of the Ras superfamily of small GTPases and like other members of this superfamily, such as Rabs, ARFs play roles in membrane trafficking [136]. ARF6 in particular is important for EV release, with nucleotide cycling converting ARF6 from its GDP to its GTP form, ultimately leading to myosin light-chain kinase (MLCK) activation to induce membrane fission and microvesicle release [137]. Other small GTPases, such as RhoA, are also involved in MV biogenesis via cytoskeletal remodelling [138].

Some of the machinery required for MV biogenesis is common to the exosome biogenesis pathway, for example, the ESCRT machinery [111, 139]. ESCRT-I (TSG101) and ESCRT-III (Vps32) are recruited to the plasma membrane for MV production [140], with mammalian studies highlighting that ESCRT components and associated factors were required for the production of specific subtypes of MVs [141].

Sphingomyelinases (SMases) have also been shown to play roles in both exosome and MV generation [111]. In MV production in particular, P2X receptors activate acid sphingomyelinase (aSMase) to enhance MV shedding [142, 143]. aSMase catalyses the same reaction as nSMase [144], but these two SMase enzymes have different locations in the plasma membrane. aSMase is present within the outer leaflet, whilst nSMase is present within the inner leaflet [144]. They may carry out varying functions depending on their localisation. It is believed that deregulation of their activity may shift the secretion of cargo from exosomes to larger vesicles, or vice versa [144].

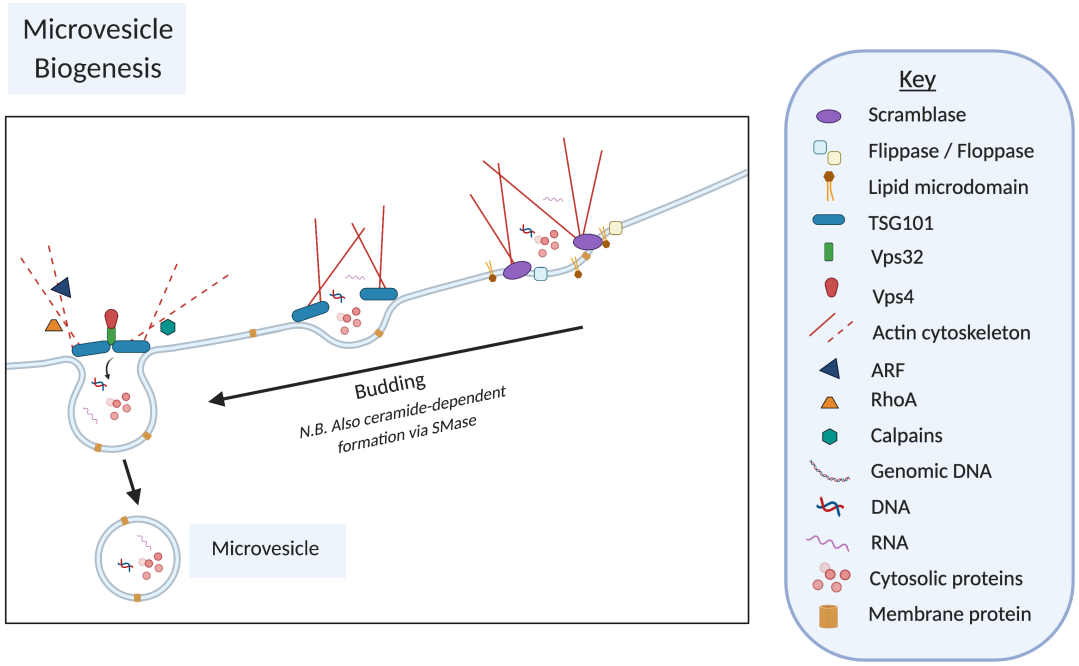


Figure 1.4: Schematic diagram depicting the biogenesis mechanisms for microvesicles, highlighting key factors involved in these processes [86, 111–115].

Apoptotic body biogenesis

Similar processes to those involved in MV release have been implicated in the generation and release of ApoEVs from apoptotic cells. A decrease in cell volume and phosphatidylserine (PS) externalisation are some of the earliest indicators of apoptosis [145, 146]. As highlighted previously, PS movement to the outer leaflet

of the plasma membrane is a key step in membrane budding, with cytoskeleton remodelling an additional step in vesicle release. In non-apoptotic cells, cytoskeleton remodelling occurs via calcium dependent pathways, for example via calpains, a family of Ca^{2+}-dependent cysteine proteases [147]. Under apoptotic conditions, however, cytoskeleton remodelling can be achieved via caspases (cysteinyl aspartate proteases). Caspases are a family of enzymes which play a central role in apoptosis [148, 149], with additional roles in ApoEV generation [150–152]. For example, caspase-3 mediates the cleavage of ROCK I (Rho-associated, coiled- coil containing protein kinase 1) during apoptosis, leading to MLCK phosphorylation, membrane contraction and EV release [153].

In the absence of caspase activity, alternative cell death enzymes called granzymes (a family of serine proteases) can facilitate cell death. ROCK II is activated by granzyme B to induce membrane blebbing in a caspase-3 independent pathway [154]. The co-ordinated cytoskeleton-driven processes induced by caspases, or alternative cell death enzymes, drives membrane blebbing, membrane protrusion formation, cell fragmentation and ApoEV release [155, 156].

1.2.3 Caspases and their non-apoptotic role in EV-biogenesis, loading and uptake

The caspase cascade is an evolutionarily conserved pathway (**Figure 1.5**). It is a key pathway in the induction of inflammation and cell death, with links to EV production in these contexts. In recent years, however, the non-apoptotic role of caspases in relation to EV-loading, release and uptake have been explored. Sapet at al. (2006) highlighted that caspases can activate ROCK II independently of cell death and this activation was required for thrombin-induced EV release [157]. Kim et al. (2017) later showed that caspase-8 controls the exosomal secretion of lysyl-tRNA synthetase (KRS) [158]. KRS is a type of aminoacyl-tRNA synthetase (ARS), essential for protein synthesis by charging amino acids to their cognate tRNAs [159]. They have also been implicated in diverse signalling pathways [160], for example, KRS is released from cancer cells to modulate the inflammatory

response [161]. KRS is cleaved by caspase-8, making its PDZ domain accessible for Syntenin binding (a process involved in EV cargo loading [162]), facilitating its inclusion into MVBs and its release via exosomes [158, 163].

Caspases have also been implicated in the process of EV-uptake. Vardaki et al. (2016) showed that caspase-3 cleavage of Bcl-xL (B-cell lymphoma-extra large, the most abundant Bcl-x protein which inhibits apoptosis [164]) occurs within EVs; a process necessary for EV uptake by myeloma and lymphoma cells [165]. Caspase-8, an enzyme which activates caspase-3, together with components of the death-receptor and toll-like receptor pathways were also found to be EV-loaded. This, together with the reported absence of cytochrome C or caspase-9, suggested that the extracellular stimuli pathway was likely responsible for the caspase-3 activation (**Figure 1.5**), with this stimulation perhaps occurring at the EV membrane surface [165]. Interestingly, reduction of this EV-uptake using a caspase-3 inhibitor resulted in decreased proliferation of target tumour cells, suggesting that the caspase-mediated regulation of EV-uptake may have functional consequences. These studies highlight the influence of caspase activity in EV-mediated cellular communication, even in the absence of cell death. Further work is needed to understand these mechanisms more clearly, potentially identifying novel ways to manipulate EV-release and uptake pathways.

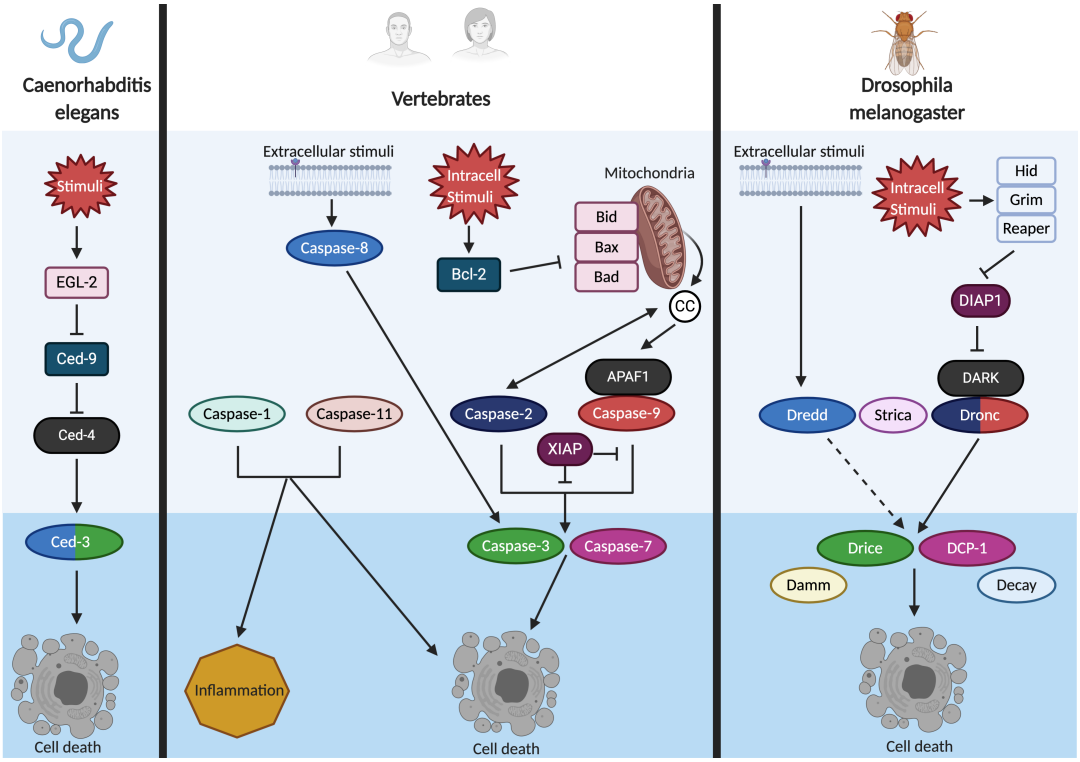


Figure 1.5: Schematic diagram depicting the caspase cascade in *Caenorhabditis elegans*, vertebrates and *Drosophila melanogaster*. Squares represent accessory proteins, with ovals indicating caspase enzymes. Similar proteins are depicted using the same colours. The upstream apical (initiator) caspase pathway is shown in the light blue region, the downstream executioner (effector) caspase pathway is shown via the dark blue region. CC = Cytochrome C. Diagram reproduced from Baena-Lopez et al. (2017) [166].

1.2.4 Selective cargo loading into extracellular vesicles

As **Section 1.2.2** has highlighted, subtypes of EVs are generated via distinct biogenesis pathways. Cargo within these EV subtypes is highly dynamic, varying depending on the cellular source, physiological state and environmental condition [167–170]. EV heterogeneity and dynamic content was realised as early as 1987 by Johnstone et al., who stated that it was unclear if *'each exosome contains a mixture of all externalized components or if a mixed population is externalized'* [171]. EV heterogeneity adds a layer of complexity when trying to understand how the EV-mediated transfer of cargo mediates specific biological responses. Today, multi-omic analyses have allowed complex EV cargo profiles to be better understood, with the identification of defined markers of EV subtypes aiding EV isolation and separation processes [73, 75]. This extensive profiling of EV contents has also

revealed the presence of EV-enriched cargos, suggesting that specific molecules may be selectively loaded into EVs to carry out particular functions in recipient cells.

Comprehensive studies of protein and lipid content of EVs have been carried out, with databases such as EVpedia, Exocarta and Vesiclepedia documenting the EV cargo released from different cell types and under varying cellular conditions [172–176]. In recent years there has been a strong focus on studying EV RNA cargo. Messenger RNA (mRNA) and micro RNA (miRNA) cargo is more commonly studied [68–70], with recent investigations exploring the significance of non-coding RNAs, such as small nucleolar RNAs (snoRNAs) or small nuclear RNA (snRNAs) [177, 178], highlighting their role in long-range gene regulation.

Protein loading into EVs

Extensive work has been carried out to characterise the protein content of EVs [88, 179–183]; however, variations in the cell type studied, as well as differences in the EV isolation methods and cell culture conditions used, has made it challenging to build a universal profile of protein cargo within EVs [168]. Proteins commonly found within EVs are those related to EV biogenesis pathways, for example, TSG101, ALIX, Rabs (RAB27A, RAB11B), ARF6 and tetraspanins (CD63, CD81, and CD9) [168]. Annexin A1 has recently been identified as a specific marker for microvesicles shed directly from the plasma membrane [76].

A number of different mechanisms have been proposed to control EV protein loading. Fang et al. (2007) showed that higher-order oligomerization (the oligomerization of oligomers) and membrane binding were sufficient to target proteins to sites of membrane budding [184]. For example, common EV tetraspanin markers such as CD63, CD81 and CD9 are membrane proteins which form large heterogeneous complexes referred to as the ‘tetraspanin web’ [185–187].

Protein ubiquitylation is a key process involved in the ESCRT-dependent sorting of proteins into ILVs for lysosomal degradation [89]. The role of protein ubiquitylation and ESCRT machinery in EV protein loading, however, is less well understood [184, 187, 188]. The involvement of ESCRT machinery in EV biogenesis

and cargo loading was demonstrated by Colombo et al. (2013), who showed that silencing the expression of ESCRT components and associated proteins (such as ALIX, TSG101, Hrs and Vps4) resulted in altered EV protein composition and in some instances, changes in the EV size profile [63]. Protein ubiquitylation has also been shown to play a role in EV protein loading; however, conflicting reports have shown that protein ubiquitylation can have either inhibitory or facilitatory roles depending on the context [189, 190], highlighting the molecular complexity of these processes.

In addition to ubiquitylation, protein sorting into EVs can be achieved via alternative post-translational modifications such as SUMOylation (addition of small ubiquitin related modifier (SUMO)) [191], phosphorylation, glycosylation, oxidation, WW domains or coiled coil domains also play roles in selective protein loading [192].

Flotillins (Flotillin-1/reggie-2 and flotillin-2/reggie-1) are additional proteins which may play roles in EV cargo sorting. For example, Phuyal et al. (2014) highlighted that the small interfering RNA-mediated down-regulation of flotillin-1 and flotillin-2 did not disrupt exosome secretion but did alter exosome composition [193]. Therefore, it has been suggested that flotillins play a distinct role in cargo loading into specific populations of ILVs; however, the mechanisms of this loading are not fully understood [194].

Although protein cargo and the mechanisms by which it becomes EV-loaded have been widely studied, controversy remains regarding the universal nature of classical EV protein cargo. Jeppesen et al. (2019) determined that 10 of the top 25 proteins commonly associated with EVs via the ExoCarta database (including GAPDH, Clathrin and Hsp90 (Heat shock protein 90)) [175] were actually present in non-vesicular fractions [76]. In addition, the widely used exosome marker CD63, thought to be present in approximately 5 copies per EV [195, 196], has also shown variable EV incorporation, dependent on the experimental parameters [197]. Ji et al. (2019) have harnessed the power of a microchip platform for multiplexed profiling of single-cell EV secretion and have shown that a decrease in certain EV phenotypes, for example CD63 positive EVs, was associated with

the invasive potential of oral squamous cell carcinoma cell lines and primary cells (biopsy material) [198]. These authors have highlighted previously unobserved EV heterogeneity and have shown how this heterogeneity can have significant effects on EV function, particularly in pathological scenarios.

Lipid loading into EVs

Similarly to proteins, the lipid content of EVs has also been widely profiled [199–202]. Although the lipid content of EVs largely reflects that of the donor cell, some lipids are particularly associated with specific EV subtypes [168, 203]. Lipids key for EV biogenesis, such sphingomyelin (SPH), cholesterol, phosphatidylserine (PS), phosphatidylethanolamine (PE) and ceramide, are often enriched within EVs [202, 204], with phosphatidylcholine (PC) and diacylglycerol levels often decreased compared to the lipid membrane composition of the donor cell [204, 205]. The ratio of these lipids in donor cells versus EVs can change depending on the cell type under investigation [204], highlighting the need for a more large-scale comparison of various cellular contexts to understand the intricacies of selective lipid EV-loading. Nevertheless, studies have shown that the lipid content of EVs is a vital factor determining their uptake by recipient cells [206]. For example, EV uptake was significantly reduced in dendritic cells when EV donor cells were pre-treated with compounds which decreased plasma membrane levels of glycosphingolipid [207]. EV lipids may play important roles in EV 'docking' and endocytosis, possibly via cholesterol-rich microdomains in recipient cells [206]. Studies have also shown that protein and RNA loading into EVs can be achieved via their association with the lipid-raft regions of EV membranes [208, 209].

RNA loading into EVs

Studies in the mid 2000s highlighted that EVs contain RNA which can be transferred to recipient cells [68–70]. Since then, EVs have been shown to carry many subtypes of RNA, such as microRNAs (miRNAs), piwi-interacting RNA (piRNA), small nuclear RNA (snRNA), small nucleolar RNAs (snoRNAs), Y RNA, transfer RNAs

(tRNAs), messenger RNA (mRNAs), ribosomal RNAs (rRNAs) and long non-coding RNAs (lncRNAs), or fragments thereof [177, 210–224].

Studies have suggested that the RNA content of EVs largely reflects the RNA profile of donor cells [225]; however, non-random loading of particular RNAs has also been observed [216, 226–228]. Valadi et al. (2007) highlighted that no clear correlation between RNA levels in EVs and donor cells was observed [68]. Tosar et al. (2015) state that whilst miRNA secretion in MCF-7 cells was predominantly non-specific, with intracellular levels of miRNAs correlating well with their extracellular abundances, a subset of miRNAs were consistently present at higher extracellular levels than would be expected based on their intracellular abundance [78]. These authors suggested that both non-selective and selective mechanisms for RNA EV-loading may be present [78], with the potential for RNAs loaded via either mechanism to be functionally relevant [229]. Particular EV-enrichment of mRNA, ncRNA (including miRNA), rRNA and tRNA cargo has been noted [77, 216, 217, 222, 224, 230–234]. Interestingly, some studies have highlighted that it was the RNAs present at low cellular abundancies which were prone to become EV-loaded [222, 235, 236], suggesting EV RNA incorporation may, at least to some extent, be an active and regulated process, not reliant on cellular abundance levels.

Janas et al. (2015) proposed a number of mechanisms for RNA EV-loading [237]. As previously highlighted, lipid-raft regions are involved in the recruitment of nSMase2, an enzyme involved in the generation of the cone-shaped ceramide which self-associates to encourage inward budding of the MVB to form ILVs. The first step in RNA EV-loading into ILVs is likely the association of RNAs with lipid-raft regions within the outer layer of the MVB membrane [208]. Interestingly, Kosaka et al. (2010) showed that RNA sorting in mammalian cells was ceramide dependent [238]. In addition to the *in situ* generation of ceramide for EV production [118], recent studies have revealed that ceramide transfer at membrane contact sites can occur to promote EV biogenesis and cargo loading. Ceramide transport protein (CERT) mediates ceramide transport at endoplasmic reticulum (ER)-endosome membrane contact sites to regulate hepatocyte exosome biogenesis [239]. In a

recent study, vesicle-associated membrane protein-associated protein-A (VAP-A) was shown to bind CERT and ORP1L (a cholesterol transporter), to control the lipid, protein and RNA content of EVs [240]. Barman et al. (2020) proposed a model whereby lipid transfer at ER membrane contact sites drives the biogenesis of selective EV subpopulations, selectively recruiting RNA cargo [240].

Associations between RNAs and membranes can be determined via hydrophobic modifications and via specific nucleotide sequences within the RNA [241–244]. Particular RNA sequence motifs have been shown to be enriched in the RNAs released via EVs [222, 234, 245–249]. These motif sequences often displayed specific RNA secondary structures, such as stem-loops. Interestingly, lipid-rafted membranes can effectively distinguish between different RNA secondary structures, with randomly structured RNAs showing a 20-fold lower affinity to lipid-raft regions [209]. Intriguingly, miRNA exosome sorting motifs (EXOmotifs) identified by Villarroya-Beltri et al. (2013) were shown to be present in the lipid-raft binding sequences of two RNAs identified by Janas et al. (2006) [209, 234]. These studies emphasise the influence of RNA-membrane association as a possible mechanism for modulating RNA EV-loading.

RNA sequence motifs may influence EV-loading via interactions with RNA binding proteins (RBPs), modulating their cellular localisation. Villarroya-Beltri et al. (2013) showed that heterogeneous nuclear ribonucleoprotein A2B1 (hnRNPA2B1) binds miRNAs carrying specific EXOmotifs to promote their exosomal release [234], with hnRNPA2B1 even co-localising with ceramide within MVB membranes [234]. Barman et al. (2020) showed that hnRNPA2B1 is also one of the RBPs dependent on VAP-A for its incorporation into EVs [240]. SYNCRIP (synaptotagmin-binding cytoplasmic RNA interacting protein, also known as hnRNP-Q) is an additional RNA binding-protein whose EV packaging is modulated by VAP-A [240], with this protein shown to bind exosome-enriched miRNAs which carry an alternative exosome motif (hEXO) [249]. HuR (ELAV-like protein 1) [250–252], Y box binding protein-1 (YBOX-1 / YB-1) [217, 227, 247, 253], NSUN2 [247], Fragile X mental retardation 1 (FMR1) [254], heterogeneous nuclear ribonucleoprotein K (HNRNPK)

[177] and Lupus La protein (La protein) [74] have also been implicated in the loading of specific RNAs into EVs. Intriguingly, a number of these RNA binding proteins (YBOX-1, SYNCRIP, hnRNPA2B1 and HNRNPK) have been identified as caspase-3 substrates, with mass spectrometry analysis highlighting that non-apoptotic caspase-3 activity cleaves these EV-associated proteins [255]. Together, these studies suggest that nucleotide motifs, secondary structure and RBP associations may collectively facilitate RNA EV incorporation, with non-apoptotic caspase activity potentially modulating these processes.

Beyond this RBP dependent (chaperone-mediated) pathway of RNA EV-loading, an Argonaute 2 (Ago2)-associated pathway, which utilises this key component in the RNA-induced silencing complex (RISC), has also been identified [217]. Ago2 dependent loading has been shown in a colorectal cancer cell model, where KRAS was implicated in the sorting of miRNAs into exosomes [216, 256]. Whilst a general absence of Ago2 in exosomes has been reported [257], other studies have highlighted the presence of Ago2 and other RISC components within EVs [220, 240, 256, 258–260]. This suggests possible overlap may exist between the mechanisms involved in RNA targeting for gene silencing purposes and RNA targeting for EV incorporation. However, as not all RNAs are sensitive to these processes, it is likely that multiple mechanisms exist to incorporate distinct RNAs into EVs.

Some studies suggest that the RBP and Ago2 loading pathways may not intersect, for example, miR-223, whose loading is dependent on the RBP YBX1, was not affected by KRAS status [216, 217, 256]. Other evidence suggests that these two loading pathways may not always be mutually exclusive. Studies have shown that specific EV-enriched RNAs do have binding capabilities for both RBPs and Ago2 [261]. In addition, Yin et al. (2020) described how the RBP hnRNPA2B1 denatures the stem structure of a specific mRNA to reveal a miRNA binding site to facilitate the recruitment of the RISC complex for gene silencing [262]. Although this study does not specifically relate to EV RNA loading, it does provide a basis by which RBPs and the RISC complex may work together to regulate gene expression. In the future, it would be interesting to explore this phenomenon in

an EV context, particularly since hnRNPA2B1 is localised to MVB membranes and is a key regulator of EV-loading [226, 234, 240, 249].

Many studies to date have focused on the loading of RNA into small EV subpopulations such as exosomes; however, a recent study has shown that caveolin-1 forms a complex with hnRNPA2B1 to specifically traffic RNAs into microvesicles [226]. Additionally, Zietzer et al. (2020) have implicated heterogeneous nuclear ribonucleoprotein U (hnRNPU) in microvesicle RNA loading via a sequence specific interaction. In this instance the RBP negatively regulates RNA EV-loading by stabilising RNA within the nucleus [263].

Recent work has focused on delineating the biogenesis and RNA loading pathways for different EV subpopulations, which have been shown to carry unique RNA cargos [264, 265]. This remains a significant challenge, however, due to extensive shared cellular machinery and technical limitations in isolating and classifying EV subpopulations [195, 266, 267]. Moreover, much research into the mechanisms of RNA loading into EVs has focused on the miRNA loading pathway [74, 216, 217, 233, 234, 249, 250, 254], whilst mechanisms of loading for other RNA species, such as mRNAs, has been less well explored. Future work is needed to address these fundamental knowledge-gaps.

It is also important to note that although RNAs are present within EVs, they have been shown to occur in low copy numbers, with less than one copy per EV observed, on average, for most RNA species [215, 268]. For example, one miRNA copy was shown to be present per single EV and only mRNA copy (or fragments thereof) was shown to be present in every ten EVs [215]. Low RNA copy numbers suggest that EVs may need to be delivered in large quantities in order to elicit functional effects [215]. Albanese et al. (2020) utilised a novel fusion assay with improved sensitivity and specificity to show that miRNAs are a minor constituent of EV RNA cargo, with no functional effects detected in recipient cells [269]. Interestingly, these authors do describe functional mRNA transfer via EVs using an alternative version of their tool [269]. Harnessing a tool which facilitates the sub-fractionation of EV populations [73], Barman et al. (2020) showed that whilst the dense fraction contained only

around 10% of small EVs, it was highly enriched in RNA (approximately 9-fold/EV) compared to the light fraction [240]. This uneven distribution of EV RNA cargo may explain the low RNA copy numbers previously determined, especially given the fact that studies have shown functional delivery of RNA EV-cargo [270–274].

Recently, the importance of non-vesicular extracellular RNA carriers as confounding factors when studying EV RNA has been highlighted [195, 235]. Carriers such as lipoproteins (protein and lipid particles, now shown to carry miRNA) [275], RNA-binding proteins (such as Ago2) [276], and exomeres (small, non-membranous nanoparticles) [75, 277] can also carry extracellular RNA, referred to as exRNA. Karimi et al. (2018), however, used advanced methods to separate EVs from lipoprotiens and showed that most RNA was packaged inside blood derived EVs [278]. In addition, Wei et al. (2017) showed that ribonucleoproteins (RNPs) presented a distinct RNA signature, distinguishable from that of EVs, with RNPs predominantly associated with ncRNAs such as tRNA and Y RNA [215], with similar observations made by Tosar et al. (2015) [78]. The biological importance of non-vesicular exRNA should not be underestimated, particularly as selective RNA secretion via these alternative mechanisms has been described [78, 235].

Tosar et al. (2020) introduce the additional consideration of survivorship bias when studying RNA enrichment, whereby specific RNAs become enriched, not because of selective release mechanisms but due to the degradation of other RNAs [229]. This concept was introduced in the context of exRNAs and due to the presence of the EV membrane to protect luminal RNAs from RNase degradation, it is less likely to be a confounding factor in the study of EV RNA. As highlighted by Chakrabortty et al. (2015), however, it is possible that RNA fragmentation occurs within EVs [279], which could bias enrichment investigations. These caveats do not invalidate current investigations of EV RNA or exRNA, but they do draw attention to the presence of '*dark matter*' which may remain undetected [229]. Advanced RNA extraction and sequencing methods are required to explore this uncharted territory to provide complete RNA profiles, potentially identifying novel functions, mechanisms and cell communication processes.

1.2.5 EV targeting to recipient cells

Once EVs have been successfully loaded and released from donor cells, EVs can be targeted to particular cells or tissues to elicit functional effects. This may be achieved via variations in EV surface molecules; however, the processes of directed EV uptake remain unclear [267]. Understanding these processes more fully may provide knowledge to harness EVs for the targeted delivery of therapeutic cargo.

Studies have shown that EV transfer can occur between various donor and recipient cell types [280, 281], with Horibe et al. (2018) showing that EV uptake was largely dependent on the recipient cell type [282]. Cell culture studies have even shown the transfer of EVs between species [68]. Alternative studies, however, have shown the directed targeting of EVs to defined cell types [283, 284], even *in vivo* [285–287], suggesting the specific targeting of EVs may regulate their biological effect [288].

It is likely that the targeting of EVs to specific cells and organs is achieved via their enriched protein surface receptors and adhesion molecules [206, 288–294]. For example, treatment of EVs or cells with proteinase K significantly impeded EV uptake, suggesting that cell and/or EV surface proteins are required for EV uptake [295]. Negative-selection mechanisms may also be in place to aid EV targeting. For example, the EV surface integrin-associated protein, CD47, has been shown to prevent macrophage and monocyte uptake of EVs, increasing the lifetime of blood circulating EVs, facilitating more efficient targeting of EVs to pancreatic cells [296].

1.2.6 EV uptake or signal transduction in recipient cells

Whether directed or random, once EVs are delivered to a recipient cell, various mechanisms are in place to facilitate the induction of cell signalling processes within that cell; either via surface molecule interactions or direct cargo uptake (**Figure 1.2** and **Supplementary Table 7.2**). A recent tool developed by Sung et al. (2020) allows the dynamic monitoring of exosome release and uptake, highlighting that exosome internalisation happens very rapidly, occurring in just 1 to 20 minutes after contact [297]. Various studies have implicated the endocytic pathways in EV uptake, with these processes strongly inhibited at 4°C, suggesting their energy

dependence [290, 295, 298]. Further evidence that EV uptake is an active process is the inability of fixed cells to uptake EVs [283, 299]. Cytochalasin D treatment was also shown to significantly reduce EV uptake in a dose dependent manner [290, 295, 298], with Cytochalasin D known to inhibit endocytic pathways by depolymerising the actin filament network [300, 301].

Diverse mechanisms are utilised by cells to facilitate EV uptake or surface interactions. Molecular similarities exist between the multiple forms of endocytosis [302], making it hard to distinguish which EV uptake routes are utilised in particular cellular scenarios. The specific pathway harnessed appears to depend on a number of factors, such as the donor cell type, recipient cell type or cell culture conditions [303]. Future work will focus on uncovering the complexity of EV-targeting and uptake mechanisms, with improved tools required to fully investigate this, particularly *in vivo*. In addition, once EV uptake is achieved, the consequences of cargo delivery on recipient cells must be more fully explored, with improved tools required to confirm the functional significance of this process.

1.2.7 Functional effects of EV cargo in recipient cells

Once EVs have been successfully transferred to recipient cells, their functional effect depends on the ability of the EV cargo to successfully induce a response. A recent review by Masaharu Somiya (2020) highlighted a knowledge-gap in the 'EV cargo transfer hypothesis,' with uncertainties remaining about the ability of EV cargo to escape endosomal and lysosomal degradation in recipient cells [303, 304]. Bonsergent et al. (2021) determined that cytosolic cargo release only occurred for approximately 30% of uptaken EVs [305]. Indeed, despite harnessing methods to significantly enhance EV cargo loading and recipient cell uptake, a reduction in the expression levels of the target genes or the expression of particular proteins is not always observed in recipient cells [306, 307].

Studies have reported, however, the functional delivery of protein, mRNA, siRNA and miRNA via EVs [269, 308–320]. Mizrak et al. (2013) presented the first report of therapeutic mRNA and protein transfer via microvesicles, highlighting

the functional delivery of molecules which resulted in cancer regression [321]. One caveat of this study, however, was that endosomal escape of this mRNA or the resultant protein was not confirmed and thus, functional effects could have occurred even without endosomal escape [304]. Recent advances have allowed endosomal escape of RNAs after nanoparticle-mediated uptake to be studied, but the relevance of these tools to study EV uptake has not yet been explored [322]. Improved tools and new models are needed to understand the intricacies of functional cargo delivery via EVs more fully, allowing biological effects in recipient cells to be attributed to particular EV cargos [195].

1.2.8 Tools to study EV mediated communication and functional cargo transfer

EV uptake is commonly studied using fluorescent or bioluminescent signals from cells treated with labelled purified EVs [206, 269, 270, 323, 324]. These methods have a number of drawbacks: they rely on EV isolation methods which remain highly variable, all EVs are labelled making it difficult to differentiate between subpopulations, long term studies are difficult due to the short lifetime of membrane dyes and other non-EV lipid membranes may be labelled, leading to false positive results [325, 326]. Furthermore, it is difficult to discriminate between EV uptake versus EV-cell surface interactions alone. Cell surfaces are therefore stripped with acid [327] or trypsin [328] to confirm EV internalisation; however, these treatments can alter EV biology [206].

Alternative methods have been developed to overcome a number of these drawbacks, for example, the covalent fluorescent labelling of purified EVs which has a decreased detrimental effect on EV function [329]. Other studies have utilised fluorescently tagged protein [330, 331] or taken advantage of the Palmitoylation signal to target proteins to the cell membrane to allow live cell imagining of bi-directional EV transfer [326]. CD63-pHluorin EV tracking models have been utilised in the zebrafish model to study endogenous exosome transfer *in vivo* [332]. Methods have also been developed to quantify fluorescently labelled mRNA EV

transfer between cells [326]. More recently, a pair of chimeric reporter proteins which associate when within the same compartment have also been harnessed to quantify EV uptake [333].

Many of these methods, however, do not extensively prove functional delivery of EV cargo. Studies to associate biological functions to EV uptake are commonly carried out via the treatment of recipient cells with EV preparations collected from donor cells in culture. Phenotypic changes, such as gene expression levels, can be measured in recipient cells to determine if the EVs had any functional effects [303]. Beyond this, -omic analysis of EV cargo is carried out to infer what processes the EV cargo may manipulate in recipient cells. These methods do present limitations: EV heterogeneity means functional effects cannot be attributed to a specific subtype, cargo heterogeneity means effects cannot be attributed to a single EV cargo and cell type specificities, sample variation or differences in experimental methods mean functional effects may not be readily reproducible [303, 334]. In addition, serum-derived cell culture contaminants [335, 336] or extracellular RNA in non-EV carriers [337] can also have effects on recipient cells. Fujita et al. (2019) even demonstrated that empty liposomes were capable of altering lipid metabolism in recipient cells [338].

In recent years, tools have been developed to gain a deeper understanding of the functional non-coding RNA delivery via EVs. de Jong et al. (2020) describe a CRISPR-Cas9 based reporter system to allow the study of EV-mediated transfer of small non-coding RNAs at single-cell resolution [318]. Albanese et al. (2020) describe a tool to study the functional transfer of miRNAs or mRNAs in cell culture [269]. Alternative models have been designed to specifically study EV protein cargo, for example, the functional transfer of Cre recombinase protein between donor and recipient cells [313, 320].

Tools to study functional transfer of EV mRNAs

Although mRNA transfer via EVs has been described, the capacity for EV mRNAs to be functional in recipient cells has been less thoroughly explored. Whilst Valadi et al. (2007) successfully showed that EV mRNAs could be successfully translated into full-length proteins via an *in vitro* translation method, actual protein translation within recipient cells was not confirmed [68, 303]. A significant advance in the study of functional mRNA transfer via EVs came with the establishment of an *in vivo* Cre-LoxP reporter [309–312]. These studies showed that *cre* mRNA could be successfully transferred via EVs to recipient cells, where it was subsequently translated into Cre protein to facilitate the excision of a loxP flanked stop-cassette, resulting in the induction of a fluorescent change.

Ridder et al. (2014) reported that the injection of purified EV preparations from the bone marrow or peripheral blood samples of Vav1-iCre mice (Cre driven in the hematopoietic cell lineage) into reporter mice resulted in positive Purkinje neurons and glial/microglial-like cells. Recombination events in the brain were rare in healthy animals but increased during injury or inflammation. Upon EV-uptake, changes in the miRNA profile were observed, allowing additional functional effects to be studied [310].

Ridder et al. (2015) used similar techniques, observing reporter positive leukocytes infiltrating transplanted Cre-producing tumour cells [309]. Once again, altered miRNA profiles were observed in recipient cells, with immunosuppression phenotypes observed after EV uptake. Zomer et al. (2015) also studied EV-mediated transfer from tumour cells to their microenvironment [311], confirming that EVs released by tumour cells were taken up locally and at a distance, with the uptake of tumour derived EVs enhancing cell migration and metastatic behaviour in recipient cells [312].

These studies confirmed the functional transfer of EV mRNAs and provided a means to allow recipient cells to be differentiated and additional physiological effects to be studied. These systems, however, do present a number of drawbacks. The presence of *cre* mRNA within EVs was confirmed in the mammalian cell culture

Cre-LoxP systems via qPCR; however, Cre protein was not detected within EVs via Western Blotting or ELISA analysis [309–312]. Whilst this could suggest that EV *cre* mRNA transfer alone was responsible for the recombination events, further analysis is required to completely eliminate the possibility that Cre protein transfer occurred to facilitate the recombination events, but at a level low enough to not be detected via Western Blotting or ELISA. In addition, *in vivo* recombination efficiencies were low (approximately 0.1 to 5.8% depending on the cellular scenario [310, 312]), highlighting concerns regarding the physiological significance of EV-mediated cargo transfer or alternatively, the sensitivity of these tools to detect it. Interestingly, these studies showed an increase in recombination efficiencies upon the induction of cancer, injury or inflammation, suggesting the relevance of EV-mediated communication may become more important under cellular stress conditions. Although these studies utilised *in vivo* mouse models, they still required the injection of purified EVs or the implantation of transgenic cells. These models, therefore, were still affected by technical variations, reducing the reproducibility or physiological relevance of the result. Scope exists to further adapt this Cre-LoxP *in vivo* tool to study the functional effects of EV cargo in alternative model organisms to overcome a number of these limitations.

1.3 *Drosophila* as a model organism

Within the EV field, much research has utilised cell culture techniques, whether that be 2D, or more recently 3D, cellular scenarios [339, 340]. Whilst these techniques have successfully built up a fundamental knowledge of EV biogenesis, loading, release and uptake, they do not fully represent the complex microenvironment of cells within complex tissues. Furthermore, specific cell lines may behave differently due to their gene expression profiles, with even the same cell line showing altered EV production under varying cell culture conditions, such as passage number and cell density [341]. Van Den Berg et al. (2019) have recently described an 'Organs-on-chips' model which has the potential to be applied to the EV field to mimic the true cellular microenvironment more closely [342]. Other EV studies have utilised

in vivo mouse [270, 271, 326, 343], zebrafish [332], monkey [344] and pig [345] models to gain a better understanding of EV biology within and between tissues, complementing cell culture systems. The EV-field would benefit from additional tools to aid the investigation of EV-biology in complex cellular scenarios.

Drosophila melanogaster is an attractive model organism for studying cell communication due to their short generation time (**Figure 1.6**), large number of progeny, affordability, extensive genetic tractability and accessible tissues [346]. *Drosophila* imaginal discs are one such tissue; sac-like epithelial tissues found inside the third instar larval stage of *Drosophila* undergoing metamorphosis [347]. *Drosophila* have a total of 19 imaginal discs which give rise to adult appendages such as wings, eyes, legs and genitalia (**Figure 1.7**). The term 'imaginal' stems from the Latin word 'imago' meaning 'image,' the word used to refer to the adult insect life stage [347].

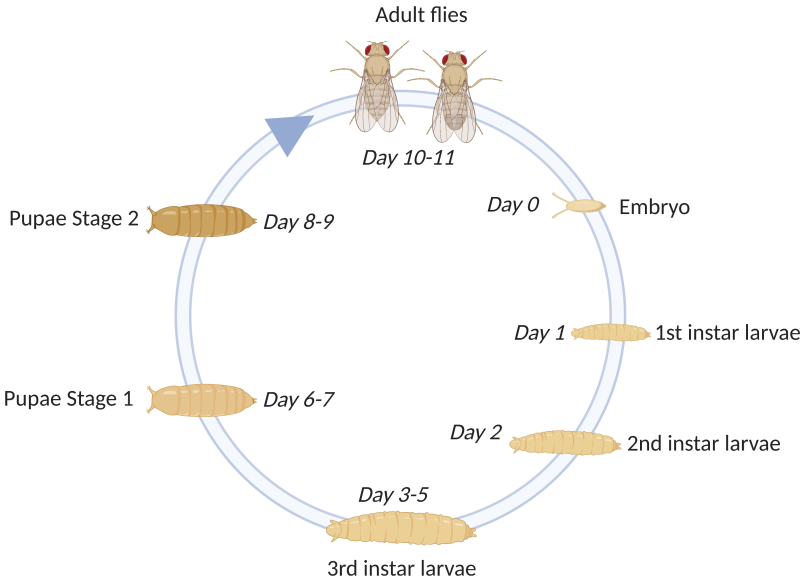


Figure 1.6: Schematic diagram depicting the *Drosophila* life cycle at 25°C. Life stage is detailed on the outside of the circle, with days after egg laying detailed on the inside. Information extracted from Flatt, 2020 [348].

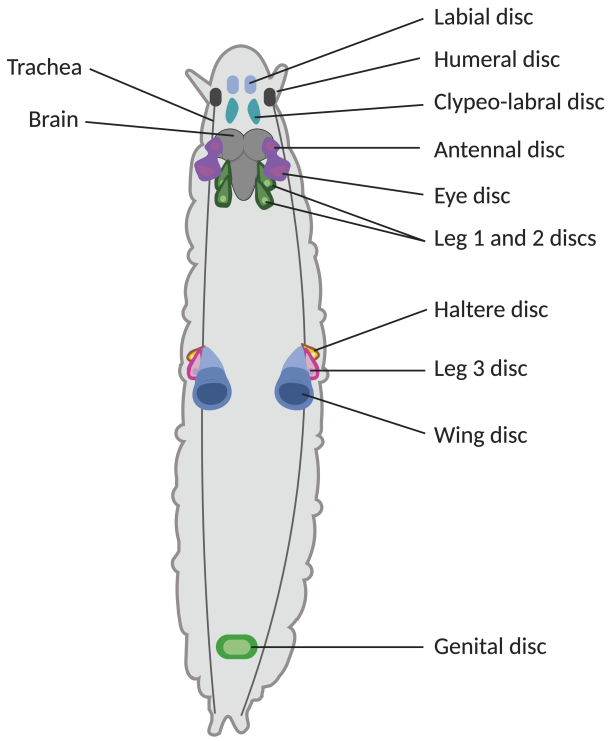


Figure 1.7: Schematic diagram depicting the 19 *Drosophila* imaginal discs. Information extracted from Phalle et al. (2004) [349].

Imaginal discs are double layered structures connected by intermediate cells; the disc proper comprises of a single layer of tall and thin columnar epithelial cells, with the peripodial membrane made up of a single layer of flat and wide squamous cells (**Figure 1.8**). The disc proper is the primary focus for many studies as this layer will eventually make up the majority of the adult appendage; however, the peripodial membrane is also important for growth, development and tissue patterning [346].

Drosophila imaginal discs have been extensively studied in the field of regeneration, development, morphogenesis, wound-healing, cell competition and tumorigenesis [350–356]. Due to the powerful genetic tools available in *Drosophila*, complex genetic manipulations can be made in specific cellular compartments [346], allowing gene or protein functions in various cellular scenarios to be studied. Furthermore, because the imaginal disc epithelial tissues share similarities with epithelial tissues found in mammals, knowledge gained using these tissues is relevant for research and medical purposes [350, 357].

1.3.1 Imaginal wing disc structure

Imaginal wing discs (IWDs) are often utilised for cell communication studies. IWDs contain the wing, hinge and notum (back of the fly) regions (**Figure 1.8**), which correspond to the relevant region in the adult fly [347, 351]. IWDs have a high proliferation rate, forming as a group of 30 cells and dividing up to every 10 hours during the larval stage to form a tissue made up of approximately 50,000 cells [347, 358, 359]. IWDs have been utilised to study cytoneme mediated signalling [21], hedgehog morphogen gradients [360] and compensatory proliferation [361]. IWDs have been harnessed to study EV biology, for example, the EV-mediated transfer of the morphogen Wg [47]. A recent study has also implicated microvilli-derived EVs in the transfer of the morphogen Hh within the IWD [36], building on previous studies highlighting the importance of cytonemes and exosomes for Hh signalling in *Drosophila* [28, 29, 33, 362]. The IWD, therefore, would provide an excellent model tissue for the development of a novel *in vivo Drosophila* tool to study EV-loading and transfer.

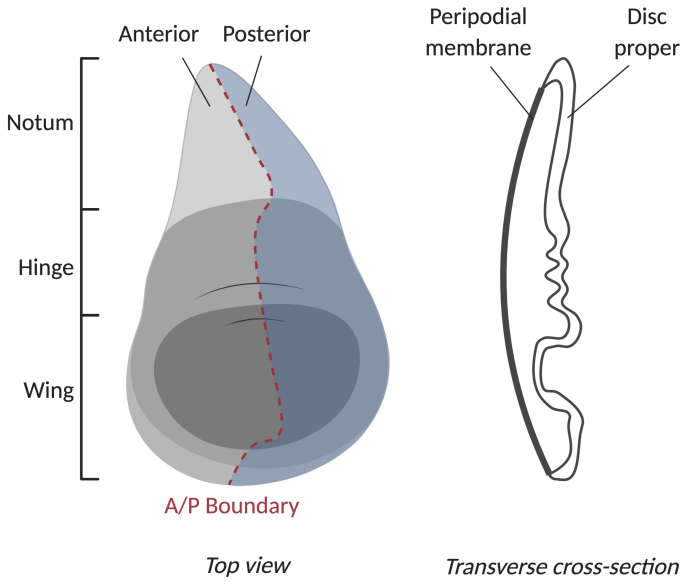


Figure 1.8: Schematic diagram depicting a top view and transverse cross section of an imaginal wing disc. The disc proper comprises of a single layer of tall and thin columnar epithelial cells, with the peripodial membrane made up of a single layer of flat and wide squamous cells. The IWD comprises of approximately 50,000 cells [347, 358, 359] and is approximately 250 µm wide at the widest point. Information extracted from Butler et al. (2003) [363].

1.3.2 Evolutionary conservation of EV-mediated cell communication in *Drosophila*

Establishing a *Drosophila* tool to study EV-loading and transfer would provide both a means to investigate new biological mechanisms involved in these processes and investigate their potential evolutionary conservation. *Drosophila* have previously been harnessed to demonstrate a novel role for GAPDH in EV assembly, with the potential for this target to be exploited to increase EV-mediated drug delivery to the brain [364]. In addition, this model organism was used to show that mutations in the enzyme glucocerebrosidase, often associated with Parkinson's disease, resulted in increased EV production which may lead to protein aggregation and neurodegeneration [365]. These studies highlight how research in *Drosophila* has relevance for the study of human pathologies and therapeutics.

Beyond studying the mechanisms by which EVs are generated or released, *Drosophila* have also been utilised to study EV cargo loading. Lefebvre at al. (2016) carried out a comparative analysis of human and *Drosophila* cellular and EV RNAs [366]. This study highlighted the presence of RNAs which were full length and enriched for specific functions such as translation, in both human and *Drosophila* derived EVs. Considerable conservation, particularly in the abundance of ribosomal RNA and retrotransposons, was observed between these two species. EV-exclusive RNAs were identified in both the human and *Drosophila* conditions and whilst gene-specific mRNA enrichment was not strongly correlated between these two species, the exclusivity of such EV mRNAs suggested the presence of targeting signals within their sequence that may promote their EV incorporation [366]. Such targeting signals, such as EV-enriched RNA motifs, have been identified in mammalian systems, as outlined in **Section 1.2.4**; however, their presence in *Drosophila* has not been explored, providing the opportunity to determine the universal nature of RNA EV-loading mechanisms across species.

1.4 Aims and objectives

Whilst the EV-loading mechanisms for miRNA and other ncRNA have been more widely explored, knowledge regarding the EV incorporation of mRNA is more limited, with a lack of *in vivo* tools available to investigate these processes. Research described within this thesis sought to explore the presence of EV-enriched features within *Drosophila* mRNAs which may facilitate their EV-loading and develop a *Drosophila* based tool to assess the biological function of these features to modulate long-range mRNA transfer *in vivo*. The three main aims of this research project are summarised below:

- Carry out an RNA-sequencing analysis of *Drosophila* cellular and EV RNAs to identify EV-enriched mRNAs and specific features they contain, such as sequence or structural patterns, which may promote their EV incorporation.
- Adapt a Cre-LoxP based reporter system originally developed in mice [309–312] for use in *Drosophila melanogaster* and harness this system to biologically validate the EV-enriched mRNA features *in vivo*, to determine their influence on long-range mRNA transfer.
- Harness the *Drosophila* Cre-LoxP system to identify novel molecular factors involved in long-range cell communication, dedicating special attention to caspases given the recent insights into their involvement in EV cargo loading and uptake processes in non-apoptotic contexts [158, 165, 255].

2
Methods

Contents

2.1 Generation of the Cre-LoxP reporter system 38
2.1.1 Key molecular biology techniques 38
2.1.2 Actin-loxP-myrGFP Reporter plasmid generation 39
2.1.3 Cre-V5 and Cre-V5-25 plasmid generation 39
2.1.4 Cre-V5-para construct generation to test the novel EV-enriched motifs 40
2.2 Cell culture techniques 41
2.2.1 Transfection 41
2.2.2 Fixation and immunostaining 42
2.2.3 Reporter and Cre co-transfection experiments 43
2.2.4 Reporter and Cre Transwell assay 43
2.2.5 Optimisation of cell culture conditions for EV extraction 44
2.2.6 Preliminary EV extraction and characterisation 45
2.2.7 EV extraction from Cre transfected *Drosophila* S2R+ cells 48
2.2.8 Cellular and EV RNA extraction 50
2.2.9 Quantitative Real-Time PCR 51
2.2.10 RNase protection assay 52
2.3 *Drosophila melanogaster* husbandry 53
2.3.1 Fly handling 53
2.3.2 Fly stocks 53
2.3.3 Experimental fly handling 54
2.3.4 Dissection, fixation and immunostaining 55
2.4 Microscopy and Image Analysis 56
2.4.1 Co-transfection experiments 56
2.4.2 Assessment of serum free and EV-depleted FBS media conditions 57
2.4.3 Transwell assay 58
2.4.4 Imaginal discs 58
2.5 Statistical analysis and visualisation 60
2.5.1 Analysing proportional data 60
2.6 Bioinformatic analysis 61
2.6.1 Biological sample preparation 61
2.6.2 Library preparation and RNA sequencing 61
2.6.3 RNA sequencing data analysis 62

2.1 Generation of the Cre-LoxP reporter system

2.1.1 Key molecular biology techniques

Bacterial transformation

Competent cells (*Mix & Go E. coli* Transformation Kit, Zymo Research, T3001) were transformed with 1 to 5 µl of DNA (less than 5% of the total volume, approximately 500 ng - 2 µg DNA) and shaken gently. Cells were incubated on ice for ten minutes and then spread on nutrient agar plates with ampicillin (100 µg/ml) and incubated at 37°C, overnight. Single colonies were spiked into LB broth with ampicillin (100 µg/ml) and incubated in a shaking incubator at 37°C, overnight.

Plasmid purification

DNA extraction was carried out using the QIAprep Spin Miniprep Kit (Qiagen, 27106) or, if transfection grade DNA was required, the Plasmid Plus Midi Kit (Qiagen, 12941). DNA extractions were carried out using approximately 30 µl of bacterial culture.

Restriction enzyme digestion

All restriction enzyme digestions were carried out using enzymes from New England Biolabs (NEB) according to the manufacturer instructions.

Gel electrophoresis and gel extraction

Agarose gel electrophoresis was carried out using 1% agarose (Sigma, A9539), Tris-Acetate–EDTA buffer, 2-Log DNA Ladder (0.1 to 10 kb, NEB, NO550), SYBR® Safe (Invitrogen, S7567) and gel loading dye (NEB, B7025S). Gel extraction was carried out using QIAquick Spin Columns (Qiagen, 28104).

Polymerase chain reactions

Polymerase chain reactions (PCRs) were carried out using Taq PCR Master Mix Kit (Qiagen, 201443) unless otherwise stated. For colony PCRs, 1 µl forward primer (50 µM stock), 1 µl reverse primer (50 µM stock), 10 µl Master Mix and 7 µl RNase-free water were used. For PCRs from purified DNA, DNA was used at a concentration less than 1 µg per reaction, with the volume of RNase-free water adjusted accordingly to result in a 20 µl total volume per reaction.

2.1.2 Actin-loxP-myrGFP Reporter plasmid generation

pJFRC172-10XUAS-loxP>-dSTOP-loxP>-myr::GFP (a gift from Gerald Rubin, Addgene plasmid # 32145, http:n2t.netaddgene:32145; RRID:Addgene_32145 [367]) was digested with Zhra and ApaI. The Actin 5C attb polylinker plasmid [368] was digested with NheI and religated to remove a loxP site before digesting with ApaI and PmeI. The loxP>-dSTOP-loxP>-myr::GFP Zhra-ApaI fragment and the Actin 5C ApaI-PmeI fragment were ligated using T4 DNA ligase (NEB, M0202).

2.1.3 Cre-V5 and Cre-V5-25 plasmid generation

The UAS plasmid (pUASTattB [369]) was digested with NotI and Xbal. The Ubiquitin (Ubi) plasmid [370] was digested using NotI and SpeI. The Cre-V5 and Cre-V5-25 fragments were generated via PCR using Q5 High-Fidelity DNA Polymerase (NEB, M0491). Primers used are shown in **Table 2.1**. The pcDNA3.1-CMV-CFP;UBC-Cre25nt template was a gift from Jacco van Rheenen (Addgene plasmid # 65727, http://n2t.net/addgene:65727; RRID:Addgene_65727) [312]. The UAS or Ubiquitin backbones were ligated to the individual Cre fragments using T4 DNA ligase (NEB, M0202).

PCR Fragment	Forward Primer (FP)	FP sequence	Reverse Primer (RP)	RP sequence
Cre-V5	FP Cre	ATTATTGCGGCCG CATGTCCAATTTA CTGACCGTACACC	RP Cre V5	ATAAATGCTAGCC TACGTAGAATCGA GACCGAGGAGAGG GTTAGGGATAGGC TTACCATCGCCAT CTTCCAGCAG
Cre-V5-25	FP Cre	ATTATTGCGGCCG CATGTCCAATTTA CTGACCGTACACC	RP Cre 25V5	ATAAAGCTAGCAC AGGCGGAGTCCAG GCGGCAGGGTCTA CGTAGAATCGAGA CCGAGGAGAGGGT TAGGGATAGGCTT AC

Table 2.1: PCR primers used to generate the Cre-V5 and Cre-V5-25 fragments.

2.1.4 Cre-V5-para construct generation to test the novel EV-enriched motifs

Three GENEWIZ constructs were designed to contain Cre-V5 plus two versions of the *paralytic* (*para*) 3-prime untranslated region (3' UTR), separated by enzyme digestion sites. These constructs were manipulated to generate the five individual Cre-V5 *para* constructs (*para* wild-type (WT), *para* A shuffled, *para* C shuffled, *para* AC shuffled and *para* ACD shuffled) in a cost effective manner. Sequences for the three constructs generated are described in **Supplementary Section 7.1.1**. GENEWIZ constructs containing two 3'UTRs were digested with NotI and Xbal to liberate a fragment containing Cre-V5 and the first 3'UTR. These were then ligated to either the UAS or Ubiquitin backbone vectors previously digested with NotI and Xbal. This gave the UAS and Ubi constructs containing Cre-V5 with the para WT, C shuffled or ACD shuffled 3'UTR.

The GENEWIZ constructs were separately digested with ZraI to remove the first 3'UTR and leave only the second. The constructs were then religated before digesting with NotI and XbaI. Fragments containing Cre-V5 and the second 3'UTR were retained. These fragments were then ligated to either the UAS or Ubiquitin backbones digested with NotI and XbaI. This successfully resulted in the generation of UAS and Ubi constructs containing Cre-V5 with the *para* A shuffled or AC shuffled 3'UTR.

2.2 Cell culture techniques

Data described in **Figures 4.3 - 4.11**, as well as **Supplementary Figures 7.4 - 7.5**, was generated from cells maintained in the following culture conditions. *Drosophila* S2R+ cells were grown in Shields and Sang M3 Insect Media (Sigma Aldrich, S8398) supplemented with yeast extract (1 g/l, Sigma, Y1000), bactopeptone (2.5 g/l, ThermoFisher, 211677), potassium bicarbonate (0.5 g/l, Sigma, 237205). Media was additionally supplemented with 10% FBS and pen/strep (100 U/ml penicillin, 100 µg/ml streptomycin) unless otherwise stated. Cells were cultured in CELLSTAR flasks (Greiner Bio-One, 690160 or 658170) at 25°C. All other data was generated as described here, except for the use of Schneider's Insect Media (Gibco, 21720-024) instead of the supplemented M3. *Drosophila* S2R+ cells are semi-adherent and present hemocyte-like characteristics (the *Drosophila* immune cells). They are an original isolate from the S2 cell line, sourced from the embryo [371].

2.2.1 Transfection

Cells were plated at a concentration of approximately 1 - 1.5 million cells per ml. 1 ml of cell suspension was added to each well of a CELLSTAR 12 well plate (Greiner Bio-One, 665180) and left to settle for at least 8 hours. Cells were transfected using Effectene Transfection Reagent (Qiagen, 301425). Unless otherwise stated, reagent volumes were used as described within this section, with DNA prepared as detailed in **Table 2.2**, diluted in no more than 5 µl of water. In tissue culture, 75 µl of EC buffer was added to the DNA, followed by 8 µl of enhancer. Samples were vortexed vigorously before incubating for 5 minutes at room temperature. 8.3 µl of Effectene was added, samples were vortexed vigorously and incubated for 10 minutes at room temperature. 500 µl of media (without FBS or pen/strep) was added to each well, bringing the total volume up to 1.5 ml per well. 96 µl of transfection complexes was then pipetted over the cells, drop by drop. Plates were shaken gently before being returned to the incubator.

Condition	Cre or control plasmid amount (ng)	Reporter plasmid amount (ng)
Reporter + Ubi Cre-V5	300	700
Reporter + Ubi Cre-V5-25	300	700
Reporter + Ubi Control	300	700
Reporter alone	N/A	700

Table 2.2: Amount of DNA used for co-transfection experiments with Reporter and Cre (or control) plasmids.

2.2.2 Fixation and immunostaining

Cells were gently washed in sterile PBS (3 x 20 minutes) before being fixed (4% formaldehyde in PBS) for 20 to 30 minutes. Cells were then washed in 0.1% PBS-Tween (PBSTw) (3 x 20 minute washes). Samples were incubated with primary antibodies (2 hours at room temperature (RT) or overnight at 4°C). Samples were washed in 0.1% PBSTw (3 x 20 minute washes) and incubated with secondary antibodies (2 hours at RT). Samples were washed in 0.1% PBSTw (3 x 20 minute washes) and mounted on SuperFrost Microscope Slides (Thermo Scientific, 12372098) using Vectashield Mounting Medium (Vector Laboratories, H-1000 or H-1200 (with DAPI if not added as a secondary)). Antibodies used and their required dilutions (unless otherwise stated) are described in **Table 2.3**.

Antibody	Supplier	Product code	Dilution
Anti-V5	Thermo Fisher Scientific	R960-25	1:200
Anti-GFP	Abcam	Ab6673	1:200
Anti-Phospho-Histone H3	Abcam	Ab32388	1:100
Anti-Death caspase-1	Cell Signalling	9578	1:200
Anti-Beta-galactosidase	Abcam	Ab9361	1:500
Anti-goat 488 secondary	Life Technologies	A1105	1:200
Anti-rabbit 488 secondary	Life Technologies	A21206	1:200
Anti-mouse 555 secondary	Life Technologies	A3157	1:200
Anti-chicken 647 secondary	Life Technologies	A21449	1:200
Anti-rabbit 647 secondary	Life Technologies	A31573	1:200
DAPI	Thermo Scientific	62248	1:1000

Table 2.3: Antibody suppliers and their required dilutions.

2.2.3 Reporter and Cre co-transfection experiments

Cells were plated and transfected as described in **Section 2.2.1**. Cells were maintained in transfection complexes for 120 hours before being fixed and immunostained according to **Section 2.2.2**. Anti-V5 primary and anti-mouse 555 secondary antibodies were used for immunofluorescence. Cells were mounted in Vectashield with DAPI.

2.2.4 Reporter and Cre Transwell assay

Cre and reporter cells were separately plated and transfected. A cell suspension of 1 million cells per ml was used, with 1 ml of this suspension plated in each well of a 12 well plate, and 300 µl added to the ThinCert Cell Culture 12 well Inserts (1.0 µm pore diameter, Greiner Bio-One, 665610). An additional 700 µl of media was added to the lower portion of the well containing the Transwell insert to prevent drying out. Cells were left to settle for at least 8 hours before transfections were carried out according to **Section 2.2.1**, using DNA preparations described in **Table 2.4**. 48 hours post transfection, Cre transfected cells were thoroughly washed before they were suspended approximately 1 mm above reporter transfected cells (**Figure 2.1**). Cells were co-cultured for a further 48 hours. Following this, cells were fixed and immunostained as described in **Section 2.2.2**. Anti-V5 and anti-GFP primary antibodies, followed by anti-goat 488 (1:600) and anti-mouse 555 (1:600) secondary antibodies, were used for immunofluorescence. The filter was removed from the insert using a scalpel and mounted face up under a coverslip for imaging purposes. Cells and filters were mounted in Vectashield with DAPI.[1]

[1] in collaboration with Elizabeth Dellar

Condition	Cre or control plasmid amount per balcony (ng)	Reporter plasmid amount per well (ng)
Reporter + Ubi Cre-V5	300	700
Reporter + Ubi Cre-V5-25	300	700
Reporter + Ubi Control	300	700
Reporter alone	N/A	700

Table 2.4: Amount of DNA used for Transwell assay experiments with Reporter and Cre (or control) plasmids

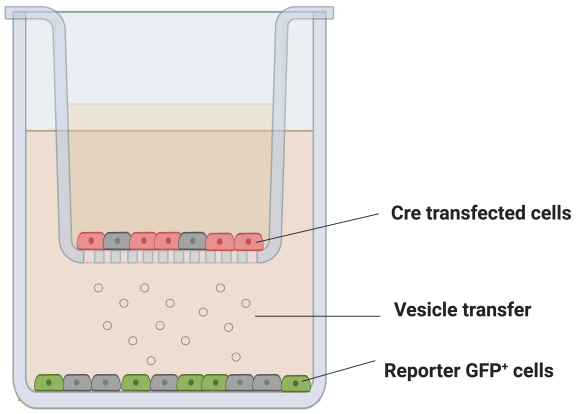


Figure 2.1: Schematic diagram to represent the experimental set up of the Transwell assay to test Cre transfer in the absence of cell contact.

2.2.5 Optimisation of cell culture conditions for EV extraction

During the experiments carried out to optimise cell culture conditions for EV extraction described in **Supplementary Section 7.1.2**, time spent in the EV-depleted or serum free media was minimised to mitigate their potentially negative effects on cell growth and survival. Cells were first plated and maintained in normal media conditions (media supplemented with 10% FBS and pen/strep). After 24 hours, the media was either refreshed, replaced with serum free media or replaced with media supplemented instead with 10% EV-depleted FBS. Cells were cultured for a further 72 hours, before being fixed and assessed for cell death (via anti-Death caspase-1 immunostaining, 1:100) or cell proliferation (via anti-Phospho-Histone H3

immunostaining). Secondary antibody staining was carried out using anti-rabbit 488.

EV-depleted FBS was generated via ultracentrifugation (120,000g for 16 h at 4°C) in open-topped tubes (Beckman Coulter, 344058) using a SW32Ti swing –bucket rotor (Beckman Coulter $\varkappa$-factor 204). Supernatant was removed and filter sterilised with a 0.22 micron filter (Merck, SLGP033R).

2.2.6 Preliminary EV extraction and characterisation

Cells were plated at a concentration of 1 million cells per ml. 40 ml of cell suspension was plated into four 10 ml plates (10 ml per plate) and left to settle overnight. Cells were washed (4 x 4 ml media without FBS or pen/strep) before media was replaced with media supplemented with 10% EV-depleted FBS and pen/strep. Cells were then incubated for a further 48 hours. Conditioned media from the four plates was pooled and centrifuged for 5 minutes at 300 g (4°C). The supernatant was retained and centrifuged for a further 20 minutes at 16000 g (4°C); once again, the supernatant was retained. A 0.22 micron filter (Merck, SLGP033R) was blocked using 1 - 2 ml of 0.01% bovine serum albumin. 1 - 2 ml PBS was passed through the blocked filter before the media was filtered through. 2 ml PBS was run through a Vivaspin 100 KDa concentrator (Fisher, 10774797). The filtered media was then centrifuged using the concentrator at 3000g in a swing bucket rotor. This was repeated as necessary to concentrate the media samples down to 1.2 ml.[2]

Size exclusion chromatography (SEC)

Size exclusion chromatography (SEC) columns (Biorad, 7321010) were prepared at least 4 hours before EV extraction. 14 ml of Sepharose CL-2B (Fisher, 10217754) was added to a column, resulting in a 10 ml column. After placing the column bed support above the resin, columns were washed two times in PBS with MgCl2 and CaCl2 (Sigma, D86626) and once using 0.03% PBSTw. The 1.2 ml media sample was split between four SEC columns, with approximately 290 µl of sample being added per column. When the top of the sample touched the column bed, the

[2] in collaboration with Elizabeth Dellar

columns were topped up with PBSTw. 500 µl fractions were collected, from fraction 1 to fraction 20.[3]

BCA Protein Assay

Protein concentrations in each fraction were determined via a Pierce BCA Protein Assay Kit (Thermofisher, 23225). Standards were made up in PBSTw (similarly to the fractions being analysed). 10 ml of working reagent was made up (9800 µl of Reagent A, 200 µl of Reagent B). 5 µl sample was added to 45 µl of working reagent. Samples were incubated at 37°C for 30 minutes with a gentle shake. Samples were left to cool for 5 minutes at room temperature before being analysed on a Molecular Devices Platereader, Spectramax i3x (Softmax Pro 7 software).[3]

Nanoparticle Tracking Analysis (NTA)

Particle concentrations for each fraction were determined using NanoSight NS500 Nanoparticle Tracking Analysis (NTA 3.2 Dev Build 3.2.16). Samples were diluted 1:10 or 1:4 in PBS before being analysed.[3] The settings used were as follows:

- Camera type = sCMOS
- Laser Type = Blue488
- Camera level = 13
- Camera Histogram Lower Limit = 0
- Camera Histogram Upper Limit = 8235
- Detection = 5
- Frame rate (frames per second) = 24.9825
- Syringe Pump Speed (arbitrary units) = 15
- Slider Shutter = 1232
- Slider Gain = 219
- Shutter (millisecond) = 30.8
- Temperature (°C) = 23-26
- Total frames analysed = 749
- Videos per samples = 3-5

[3] in collaboration with Elizabeth Dellar

Transmission Electron Microscopy (TEM)

10 µl of EV samples (approximately 2 x 10^{10} - 3 x 10^{11} particles, measured via NTA) were pipetted onto 300 mesh copper grids (TAAB, C267, previously glow discharged for 20 seconds at 10 mA). Grids were incubated for 2 minutes before being dabbed with filter paper and placed onto a 20 µl drop of 2% uranyl acetate (for 10 seconds). The grids were left to air dry. Grids were visualised using Jeol JEM-1400 Flash Transmission Electron Microscope with Gatan OneView 16 Megapixel camera at 100kV.[4]

Western Blotting

Western Blotting was carried out using the Invitrogen NuPage system and MES running buffer (Thermofisher, NP0002). 6 - 40 ng total protein was prepared in LDS buffer (Thermofisher, NP0007) and boiled at 95°C (5 minutes). For anti-Beta-actin, this was carried out with 0.4 mM DDT (Sigma, 10708984001). 20 µl of sample and 5 µl of ladder (Biorad, 1610374) was run on 4 - 12% polyacrylamid Bis-Tris pre-cast gels (Thermofisher, NP0322). Gels were run at 125 V (90 minutes). Wet-transfer onto nitrocellulose membranes was carried out in NuPage transfer buffer (Thermofisher, NP0006) with 15% methanol. Membranes were blocked for 1 hour using 5% milk in 0.1% PBSTw. Primary antibodies were incubated overnight (4°C). Membranes were washed (4 x 20 mins) in 0.1% PBSTw. Secondary antibodies were incubated for 1 hour (RT). Membranes were washed (4 x 20 mins) in 0.1% PBSTw and then incubated with ECL (Millipore Sigma, GERPN2232) for 5 minutes. Membranes were imaged using ChemiDoc MP (Bio-Rad). Antibodies are shown in **Table 2.5**.[4]

[4]by Elizabeth Dellar

Antibody	Supplier	Product code	Dilution
Anti-Latebloomer	DSHB	10C9	1:500
Anti-Beta-actin	Proteintech	66009-1-lg	1:5000
Anti-TSG101	Abcam	Ab30871	1:500
Anti-Hsp70	Abcam	Ab5439	1:1000
Anti-Golgin-84	DSHB	Golgin84 12-1	1:500
Anti-Histone H3	Cell Signalling	4499	1:1000
Anti-mouse	Promega	W4021	1:20000
Anti-rabbit	Promega	W4011	1:5000

Table 2.5: Western blotting antibody suppliers and their required dilutions.

2.2.7 EV extraction from Cre transfected *Drosophila* S2R+ cells

Cell transfections using the Ubi Cre-V5, Ubi Cre-V5-25 and Ubi control (containing the Ubi backbone in the absence of Cre) plasmids were carried out according to **Section 2.2.1**, with the following changes. Cells were plated in 10 cm dishes (Greiner Bio-One, 664960), with 5 dishes used per condition (per replicate). 10 ml of a 1 million cells per ml cell suspension was added per plate and left to settle for at least 8 hours. Reagents were used in the following amounts per 10 cm dish: 1900 ng DNA, 150 µl EC buffer, 15.2 µl enhancer and 19 µl Effectene. 199 µl of complexes was added per dish. Cells were then washed thoroughly in media (without FBS or pen/strep) before 10 ml of serum free media was added for the conditioning period. At this point, a media control sample was set up (5 x 10 cm dishes of serum free media conditioned in the absence of any cells). Media was conditioned for approximately 48 hours. After this time, conditioned media for each sample was pooled from the 5 dishes and retained. Cells were resuspended and a cell viability assay was carried out using the Trypan Blue Exclusion Test of Cell Viability (Biorad, 1450021) [372]. This cell suspension was centrifuged for 5 minutes at 300 g (4°C) and the cell pellet was retained at -80°C for downstream RNA extraction.

The conditioned media then underwent EV extraction via SEC as described in **Section 2.2.6**, except for the following changes. Media samples were concentrated down to a volume of 500 µl. Instead of the PBSTw wash, PBS alone was run through the column. Samples were eluted in PBS instead of PBSTw. The change to remove

Tween during SEC EV extractions was made as it was observed[5] that using PBS produced a cleaner SEC fraction profile, aiding EV isolation. 500 µl fractions were collected, with fractions 6 to 10 being retained and pooled together as the EV sample.

Particle concentrations for pooled fraction 6 to 10 were analysed via ParticleMetrix Nanoparticle Tracking Analysis (NTA ZetaView 8.04.02 SP). Pooled samples were diluted 1:500 in PBS before being analysed. The settings used were as follows:

- Camera = FpSec 30
- Cycles = 2
- Frame rate = 30
- Gain = 28.8
- Laser Wavelength = 488
- Minimum brightness = 25
- Maximum Brightness = 255
- Minimum area = 5
- Maximum area = 1000
- Minimum Trace length = 15
- Sensitivity = 80
- Shutter = 100
- Temperature (°C) = 21-30

Measurements were taken at 11 positions, some of which were automatically flagged as anomalies or manually excluded for calculation purposes. Mean particle sizes per replicate for each sample were estimated using the histogram size distributions obtained from NTA analysis, using the follow equation:

$$Estimated\,mean = \frac{\sum (Size\,midpoint \times Number\,of\,particles\,within\,size\,range)}{Total\,number\,of\,particles}$$

Overall means and standard deviations for the Ubi Cre-V5, Ubi Cre-V5-25, Ubi Control and media control samples (using three experimental replicates) were then determined.

[5]by Elizabeth Dellar

2.2.8 Cellular and EV RNA extraction

Fractions 6 to 10 obtained from the above EV extraction protocol were pooled together, vortexed and concentrated down to 200 µl using a Vivaspin 6 concentrator (VWR International Ltd, 28-9323-19). 700 µl of QIAzol Lysis Reagent (Qiagen, 79306) was added. Pelleted cells were recovered from the -80°C freezer and loosened by flicking the tube. 0.75 ml QIAzol Lysis Reagent was added per 5-10 $\times$ 10^7 cells. Cell and EV samples in QIAzol were homogenized by strongly pipetting up and down and vortexing for 1 minute before being incubated at room temperature for 5 minutes. 70 µl of 1-Bromo-3-chloropropane was added to samples and tubes were shaken vigorously for 15 seconds before being incubated at room temperature for 2 to 3 minutes. Samples were centrifuge for 15 minutes at 12500 g (4°C). The upper aqueous phase was transferred to a new collection tube. 1.5 volumes (usually 525 µl) of 100% ethanol was added and mixed thoroughly by pipetting up and down. Up to 700 µl of sample, including any precipitate that may have formed, was added to a RNeasy Mini spin column (Qiagen, 74104) and centrifuged for 15 seconds at 10000 g (room temperature). The flow-through was discarded and this process was repeated until all the sample was used up. 700 µl of Buffer RWT (prepared with Isopropanol) was added to the column and centrifuged for 15 seconds at 10000 g to wash. The flow-through was discarded. 500 µl of Buffer RPE was added to the column and centrifuged for 15 seconds at 8000 g to wash the column. The flow-through was discarded and another 500 µl of Buffer RPE was added to the column and centrifuged for 2 minutes at 8000 g to dry the column membrane. The columns were placed into a new 2 ml collection tube and centrifuged in a microcentrifuge at full speed for 5 minutes. The columns were added to new 1.5 ml collection tube. 30 µl of RNase-free water was added directly onto the column membrane and centrifuged for 1 minute at 100 g and then for 1 minute at 8000 g to elute the RNA.[6]

[6]Protocol optimised by Elizabeth Dellar

2.2.9 Quantitative Real-Time PCR

Qiagen Rotor-Gene Q

The following protocol was used to generate the qPCR data shown in **Figure 4.8** and **Figure 4.9**. To generate cDNA from cellular and EV extracted RNA, the Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Scientific, K1641) was used according to the manufacturer's instructions (incubated at 25°C for 10 minutes, 50°C for 15 minutes and 85°C for 5 minutes). 55 ng of media control RNA, 110 ng of cell RNA and 110 ng of EV RNA were used for the reverse transcription reactions. qPCR reactions were carried out using the QuantiNova SYBR Green PCR Kit (Qiagen, 208054) and Rotor-Gene Q Real-time PCR 536 cycler (Qiagen). Log_2FC values were determined via 2^{-DDCt} calculations, using *Arc1* as the housekeeping gene. The following cycling conditions were used on the Qiagen Rotor-Gene Q:

- Hold: 95°C (10 minutes)
- Cycle (90 times): 95°C (10 seconds), 60°C (15 seconds), 72°C (20 seconds)
- Melt: Ramp from 72 - 95°C (rising 1 degree each step). Wait for 90 seconds of pre-melt conditioning on first step. Wait for 5 seconds for each step afterwards.

Bio-Rad CFX96 Touch Thermal Cycler

For qPCR data shown in **Figure 5.3**,[7] DNase treatment was carried out (Sigma: AMPD1-1KT) according to manufacturers instructions (RNA in 16 µL water, 2 µL of 10x Reaction Buffer, 2 µL of DNase I (Amplification Grade, 1 unit/mL)). Samples were mixed gently and incubated for 15 minutes at room temperature. 2 µL of stop solution was added to bind calcium and magnesium ions and to inactivate the DNase I. Samples were heated at 70°C for 10 minutes to denature both the DNase I and the RNA.

cDNA synthesis was carried out using 15 - 40 ng of RNA and the Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Scientific, K1641). Samples were placed in a thermal cycler and incubated at 25°C (10 minutes), then 65°C (30 minutes) and finally 85°C (5 minutes). qPCR data was generated using SYBR Green (Bio-Rad, 1725121). *rpl32* was used as the housekeeping gene.

[7] by Elizabeth Dellar

The following cycling conditions were used in the CFX96 Touch Thermal Cycler (Bio-Rad):

- Hold: 95°C (30 seconds)
- Cycle (40 times): 95°C (5 seconds), 60°C (30 seconds)
- Melt: Ramp from 65 - 95°C (rising 1 degree each step). Wait for 90 seconds of pre-melt conditioning on first step. Wait for 5 seconds for each step afterwards.

qPCR primers

Primers used for the qPCR reactions are shown in **Table 2.6**.

Primer	Reference	Sequence
FP arc1	[366]	TAGAAGGTATCAGCGACGAG
RP arc1	[366]	GCCATACCGTAGAACAGCA
FP cre	[373]	CGGTCTGGCAGTAAAAACTAT
RP cre	[373]	CAGGGTGTTATAAGCAATCCC
FP rpl32	[374]	AAGCGGCGACGCACTCTGTT
RP rpl32	[374]	GCCCAGCATACAGGCCCAAG

Table 2.6: qPCR primers and literature sources.

2.2.10 RNase protection assay

For the RNase protection assay,[8] EV samples were concentrated to approximately 55 µl each, immediately after extraction. The following day, EV preparations were split into three 15 µl samples. A stock solution of 2% PBS Triton-X100 (PBSTx) was prepared. RNase A (Fisher Scientific, 10546881) was prepared at a concentration of 5 mg/ml (1:1 dilution of stock in PBS). 15 µl of 2% PBSTx was added to the relevant tubes, or 15 µl of PBS as a control. Samples were vortexed briefly and incubated on ice (30 minutes), yielding a 1% final Triton concentration. 2 µl of RNase A was added to the relevant tubes, vortexing briefly and incubating at 37°C (for 20 minutes), yielding a 0.31 µg/µl final RNase concentration. 300 µl of Qiazol was immediately added to inactivate the RNase. RNA extraction was completed as described in **Section 2.2.8**. DNase treatment was carried out as described in **Section 2.2.9**. cDNA synthesis was carried out using 15 ng of RNA. For the qPCR data shown in **Figure 4.10**, data was generated using the Bio-Rad CFX96 Touch Thermal Cycler protocol, as described in **Section 2.2.9**.

[8]By Elizabeth Dellar

2.3 *Drosophila melanogaster* husbandry

2.3.1 Fly handling

Drosophila melanogaster stocks were maintained at 18°C or 25°C unless otherwise stated. All flies were maintained on Oxford Food (ingredients shown in **Table 2.7**), with yeast added to each tube prior to usage.

Oxford Fly Food Ingredients	Concentration
Agar	0.292590028 w/v
Malt	6.232686981 w/v
Molasses	1.817867036 w/v
Maize	6.232686981 w/v
Yeast	1.267313019 w/v
Soya	0.747922438 w/v
Methyl 4-hydroxy-b	0.256232687 w/v
Ethanol	2.434210526 v/v
5% phosphoric acid in propionic acid	0.484764543 v/v

Table 2.7: Oxford Fly Food Ingredients

2.3.2 Fly stocks

Fly stocks generated for the *in vivo* Cre-LoxP reporter system are shown in **Table 2.8**.

Fly line	Chromosome (Injection Site)	Genotype
UAS Cre-V5	X (9753)	$UASCre-V5$
UAS Cre-V5-25	X (9753)	$UASCre-V5-25$
UAS Cre-V5-para WT	X (9753)	$UASCre-V5-paraWT$
UAS Cre-V5-para A shuffled	X (9753)	$UASCre-V5-paraA$
UAS Cre-V5-para C shuffled	X (9753)	$UASCre-V5-paraC$
UAS Cre-V5-para AC shuffled	X (9753)	$UASCre-V5-paraAC$
UAS Cre-V5-para ACD shuffled	X (9753)	$UASCre-V5-paraACD$
Actin-loxP-myrGFP reporter	II (9740)	w; $\frac{Actin-LoxP-Stop-LoxP-myrGFP}{Cyo}$; $\frac{MKRS}{TB6B}$

Table 2.8: Cre and Reporter *Drosophila* fly lines generated for *in vivo* experiments.

Fly stocks utilised during experiments are shown in **Table 2.9**.

Fly line	Sources	Genotype
ptcGal4	BL2017, BL7017, BL56555	w; $\frac{ptcGal4}{Cyo}$; $\frac{TubGal80^{ts}, UASHistoneRFP}{TM6B}$
hmlGal4	BL30140, [375], BL7017, BL56555	w; $\frac{hmlGal4,UASGFP}{Cyo}$; $\frac{droncKO, TubGal80^{ts}, UASHistoneRFP}{TM6B}$
spaltGal4	BL80573	w; $\frac{spaltGal4}{Cyo}$; $\frac{TM2}{TM6B}$
lacZ	BL4539, BL6355	w; $\frac{UASFlp}{Cyo}$; $\frac{Actin-FRT-STOP-FRT-LacZ}{TM6B}$
Actin-loxP-myrGFP reporter, p35	BL5073	w; $\frac{Actin-loxP-STOP-loxP-myrGFP}{Cyo}$; $\frac{UASp35}{TM6B}$

Table 2.9: *Drosophila* fly lines utilised in *in vivo* experiments. BL = Bloomington Drosophila Stock Center.

2.3.3 Experimental fly handling

Experimental fly crosses were maintained at 25°C unless otherwise stated.

Timeline experiments

For timeline experiments, the schematic shown in **Figure 2.2** describes how Cre expression was controlled during specific larval stages of development using the temperature sensitive (ts) Gal4 suppressor, $Gal80^{ts}$, under the control of the tubulin promoter ($TubGal80^{ts}$). Cre expression was either switched on during early or late larval development only. Flies were maintained at their initial temperature for 3 days (25°C condition) or 6 days (18°C), as developmental time roughly doubles at 18°C [376]. Flies were then shifted to the alternative temperature and maintained there until ready to dissect at the third instar larval (L3) stage.

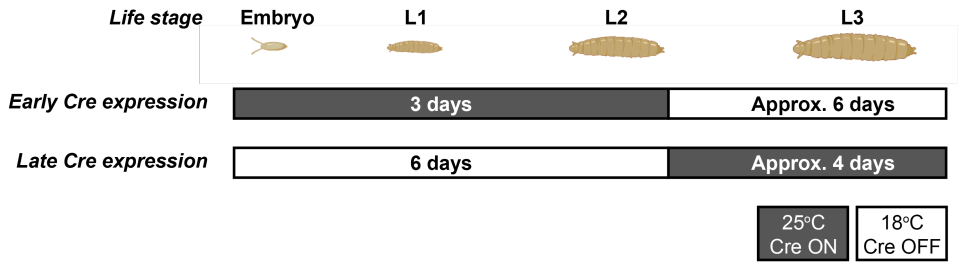


Figure 2.2: Schematic diagram describing the timeline experiment to control Cre expression during specific stages of larval development. In the presence of ptcGal4 TubGal80^{ts}, UAS Cre expression was turned off at 18°C (White) and switched on at 25°C (Grey). L1 = First instar larvae. L2 = Second instar larvae. L3 = Third instar larvae.

2.3.4 Dissection, fixation and immunostaining

L3 larvae dissections were carried out in PBS. Wing disc and eye disc dissections were carried out according to the following method [377]. Larvae were held halfway down with two forceps. Forceps were pulled apart to divide the larvae in two, retaining the head region. The tip of one side of the forceps was placed at the larvae mouth and using the other pair of forceps, the larvae was gently pushed over the single end of the forceps, turning the larvae inside out. The fat body and guts were gently detached, leaving the trachea and brain attached. Larvae were placed in fixative on ice until all dissections were complete.

Samples were fixed (4% formaldehyde in PBS) for 20 - 30 minutes at room temperature and washed in 0.3% PBS-TritonX (PBSTx) (3 x 20 minutes). Samples were incubated with primary antibodies (2 hours at room temperature (RT) or overnight at 4°C). Samples were washed in 0.3% PBSTx (3 x 20mins) and incubated with secondary antibodies (2 hours at RT). Samples were washed in 0.3% PBSTx (3 x 20 minutes) and mounted on SuperFrost Microscope Slides (Thermo Scientific, 12372098) using Vectashield Mounting Medium (Vector Laboratories, H-1000). Antibody suppliers and required dilutions are shown in **Table 2.3**.

2.4 Microscopy and Image Analysis

Confocal images of *Drosophila* larvae dissections were captured using a Live Cell Olympus confocal microscope. Fluorescent images of cell culture experiments were captured using the Olympus IX-70 or the EM7000. Images were analysed using ImageJ [378, 379] or CellProfiler™ [380].

2.4.1 Co-transfection experiments

Single focal plane fluorescent images were captured on the Olympus IX-70 using a 20x lens. DAPI counts were determined using the following CellProfiler parameters:

- Color to gray:
  - Conversion method - Split
  - Image type - RGB
- Identify primary objects in DAPI channel:
  - Typical diameter of objects - 5 to 60 pixels
  - Discard objects outside this diameter range - Yes
  - Discards objects touching the border of the image - Yes
  - Threshold strategy - Global
  - Thresholding method - Otsu
  - Three classes of thresholding
  - Assign pixels in the middle intensity class to the background.
  - Threshold smoothing scale - 1.3488
  - Threshold correction factor - 1.0
  - Lower and upper bounds on threshold - 0, 1
  - Method to distinguish clumped objects - Intensity
  - Method of draw dividing lines between clumped objects - Intensity
  - Automatically calculate size of smoothing filter for declumping - Yes
  - Automatically calculate minimum allowed distance between local maxima - Yes
  - Speed up by using lower-resolution image to find local maxima - Yes
  - Fill holes in identified objects - After both thresholding and declumping
  - Handling of objects is excessive number of objects identified - Continue
- Convert objects to image:
  - Color format - Binary (black and white)
- Save split images:
  - Saved file format - Tiff
  - 8-bit integer
  - Save every cycle
- Save binary images:
  - Saved file format - Tiff
  - 8-bit integer
  - Save every cycle
- Export to spreadsheet:
  - Select column delimiter - Comma
  - Select data to export - Nuclei
  - Calculate the per-image mean values for object measurements - Yes

Images in each cell status (Cre^{+} alone, Cre^{+} and Reporter^{GFP+}, or Reporter^{GFP+} alone), were manually quantified via visual analysis of the RBG split and binary images output in the protocol above.

2.4.2 Assessment of serum free and EV-depleted FBS media conditions

Single focal plane fluorescent images were captured on the Olympus IX-70 using a 10x lens. Using the ImageJ macro shown in **Listing 2.1**, quantification of DCP1 or PH3 positive cells was carried out. The nuclei channel was also extracted using this macro for downstream nuclei quantifications using CellProfiler.

```
 // pick directory
                macro "Split and save" {
                dir1 = getDirectory("Choose Source Directory") ;
                dir2 = getDirectory("Choose Destination Directory") ;
     list = getFileList(dir1);
L = 0
i = 0

// Nuclei
for (i=0; i<list.length; i++) {
showProgress(i+1, list.length);
Temp = File.getName(dir1+list[i]);
if (endsWith(Temp, ".tif")) {
open(dir1+list[i]); N = getTitle();
selectWindow(N);
run("Split Channels");
selectWindow(N+" (red)");
display summarize add");
saveAs("Tiff", dir2+list[i]+"nuclei");
while (nImages>0) {
          selectImage(nImages);
          close(); }}}

// Antibody staining
for (i=0; i<list.length; i++) {
showProgress(i+1, list.length);
Temp = File.getName(dir1+list[i]);
if (endsWith(Temp, ".tif")) {
open(dir1+list[i]); N = getTitle();
selectWindow(N);
run("Split Channels");
selectWindow(N+" (green)");
run("Despeckle");
run("Auto Threshold", "method=Intermodes white");
run("Convert to Mask");
run("Watershed");
saveAs("Tiff", dir2+list[i]+"green");
      while (nImages>0) {
          selectImage(nImages);
          close(); }}}
 print("done");}

```

Listing 2.1: Macro used for image processing and quantification of media trial images via ImageJ.

CellProfiler was then used with the following parameters to determine nuclei counts:

- Identify primary objects in DAPI channel:
  - Typical diameter of objects - 1 to 13 pixels
  - Discard objects outside this diameter range - Yes
  - Discards objects touching the border of the image - Yes
  - Threshold strategy - Global
  - Thresholding method - Otsu

  - Three classes of thresholding
  - Assign pixels in the middle intensity class to the background.
  - Threshold smoothing scale - 1.3488
  - Threshold correction factor - 1.0
  - Lower and upper bounds on threshold - 0, 1
  - Method to distinguish clumped objects - Intensity
  - Method of draw dividing lines between clumped objects - Intensity
  - Automatically calculate size of smoothing filter for declumping - Yes
  - Automatically calculate minimum allowed distance between local maxima - Yes
  - Speed up by using lower-resolution image to find local maxima - Yes
  - Fill holes in identified objects - After both thresholding and declumping
  - Handling of objects is excessive number of objects identified - Continue
- Export to spreadsheet:
  - Select column delimiter - Comma
  - Select data to export - Nuclei
  - Calculate the per-image mean values for object measurements - Yes

2.4.3 Transwell assay

Using the EVOS EM7000, cells were imaged using the 40x lens. 42 stitched fields of view were analysed for each condition, with a z-stack of 5 focal planes taken at each position. Serpentine horizontal, spiral out clockwise and default stitch parameters were used.

Images were analysed as follows[9]: Stitched and z-projected images were converted to 8 bit images using ImageJ. ROI segmentation carried out in the DAPI channel using Otsu thresholding method and watershed function. Measure Intensity measurements were taken using the green and red channels in defined ROIs with diameter 200 - 5000 pixels. Reporter^{GFP+} cells were defined by a mode intensity greater than 30, Cre^{+} cells were defined if mode intensity was greater than 80.

2.4.4 Imaginal discs

Z-stack confocal images (approximately 25 focal planes) were taken per imaginal disc using the Live Cell Olympus confocal 20x lens. Projections of the z-stack images for the epithelial membrane (excluding peripodial membrane) were made for the IWD. If stitching was required, this was carried out using the *Plugins>Stitching>Pairwise stitching* tool (using Linear Blending and default parameters). For visualisation purposes, individual channels underwent *Process>Noise>Despeckle* processing to

[9]by Elizabeth Dellar

remove background signal via ImageJ, as necessary. For quantification purposes, the ImageJ macro shown in **Listing 2.2** was used for image processing. For quantifications of the posterior or Reporter^{GFP+} area of IWDs, manual outlines of these areas were made using the ImageJ *polygon selection* tool. Area measurements were made of these selections via the ImageJ *Analyze>Measure* tool.

```
T = getTitle();
selectWindow(T);
path = getDirectory("image");
run("Split Channels");

selectWindow("C1-"+T);
Blue = getTitle();
selectWindow("C2-"+T);
Green = getTitle();
selectWindow("C3-"+T);
Red = getTitle();
selectWindow("C4-"+T);
FR = getTitle();

selectWindow(Red);
run("Despeckle");
selectWindow(FR);
run("Despeckle");

selectWindow("C1-"+T);
Blue = getTitle();
selectWindow("C2-"+T);
Green = getTitle();
selectWindow("C3-"+T);
Red = getTitle();
selectWindow("C4-"+T);
FR = getTitle();

run("Merge Channels...", "c1=["+Blue+"] c2=["+Green+"] c3=["+Red+"] c4=["+FR+"] create");

Stack.setChannel(1); run("Blue");
Stack.setChannel(2); run("Green");
Stack.setChannel(3); run("Red");
Stack.setChannel(4); run("Grays");

saveAs("tiff",path+"new"+T);
```

Listing 2.2: Macro used for image processing of IWD projections via ImageJ.

The percentage area of DCP1 positive staining within Cre producing cells of the IWD was determined by drawing an ImageJ Region of Interest (ROI) around the ptcGal4 domain within the wing pouch region of IWD z-projection. The ImageJ macro shown in **Listing 2.3** was then used.

```
T = getTitle();
selectWindow(T);
path = getDirectory("image");
run("Split Channels");
selectWindow("C4-"+T);
run("Despeckle");
run("Auto Threshold", "method=MaxEntropy ignore_black white");
roiManager("Select", 1);
run("Analyze Particles...", "  show=Outlines display clear summarize");
```

Listing 2.3: Macro used for quantifying DCP1 positive staining within the IWD.

2.5 Statistical analysis and visualisation

Statistical analysis was carried out via GraphPad Prism unless otherwise stated. Visualisations were carried out using GraphPad Prism, Matplotlib (v. 3.2.1) or Seaborn (v. 0.10.0) [381, 382]. Figures were generated using `https://biorender.com/`.

2.5.1 Analysing proportional data

When analysing the proportional data (for example, posterior clone event rates in the IWD, when driving Cre expression with ptcGal4), the true population proportion of these events was not known and therefore, standard deviation (SD) values could not be calculated. Instead, the standard error of the proportion ($SE_{\hat{p}}$) was determined, as an approximation of the SD. This was carried out using the following equation [383]:

$$SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $\hat{p}$ is the proportion of the sample that experiences the event of interest and n is the sample size.

Z-scores were determined when comparing the differences between proportions using the following equation [384]:

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_1(1-p_1)}{n_1}}}$$

$$\therefore Z = \frac{p_1 - p_2}{\sqrt{(SE_1)^2 + (SE_2)^2}}$$

where p_1 is the proportion of sample 1 that experiences the event of interest and p_2 is the proportion of sample 2 that experiences the event of interest. n_1 and n_2 are the sizes of sample 1 and sample 2, respectively. SE_1 and SE_2 are the standard errors calculated for sample 1 and sample 2, respectively. Z-scores were then used to determine p-values via the online GraphPad resource, QuickCalc [385].

2.6 Bioinformatic analysis

This work was carried out in collaboration with Elizabeth Dellar and Jana Terenova. All packages were run using command line unless otherwise stated. Galaxy Tools were run via `https://usegalaxy.org/` [386]. A schematic summary of the bioinformatic pipeline is shown at the end of this chapter (**Figure 2.5**).

2.6.1 Biological sample preparation

Biological triplicates of dual-indexed, strand-specific sequencing libraries were prepared from DNase-treated RNA extracted from EVs and their parental cells using a NEB Next Ultra Directional RNA kit (NEB, E7420L).[10] Each sample was generated using four T75 flasks of *Drosophila* S2R+ cells (seeded at approximately 1 million cells per ml) cultured in Schneider's media with 10% EV-depleted FBS and conditioned for 48 hours. Each RNA extraction was eluted in 30 ul of water, resulting in an approximate total volume 25 ul after elution. The optimisation of conditions for *Drosophila* SR2+ cell culture and EV extraction is described in **Section 7.1.2**.

Each sample underwent BioAnalyser capillary electrophoresis (**Figure 3.1**).[11] Cellular RNA was ribo-depleted, however, EV RNA was not.

2.6.2 Library preparation and RNA sequencing

Library preparation and sequencing was carried out by Liverpool sequencing facility. 150 base pair paired end (reverse stranded) sequencing was carried out on an Illumina HiSeq 4000 machine. Raw Fastq files were trimmed for the presence of Illumina adapter sequences using Cutadapt version 1.2.1 [387], using the option -O 3 to remove the 3' end of any reads which match the adapter sequence for 3 base pairs or more. Reads were further trimmed using Sickle version 1.200 with a minimum window quality score of 20 to remove reads shorter than 20 base pairs after trimming.

[10] by Elizabeth Dellar
[11] in collaboration with Elizabeth Dellar

2.6.3 RNA sequencing data analysis

Quality control

Quality control of sequencing data was done using FastQC Galaxy Tool (v.0.72) [388]. FastQC results were summarised using the MultiQC Galaxy Tool (v.1.6) (**Supplementary Figures 7.8 - 7.9**) [389].

Removal of rRNA contamination

Reads aligning to the rRNA FASTA sequence for *Drosophila melanogaster* were removed.[12] This was particularly important because EV RNA was not ribo-depleted during sample preparation. The rRNA FASTA sequence file was generated via Ensemble Biomart [390] (Ensemble Genes 95; Fruitfly genes (BDGP6)). The following Biomart filters used were:

· **Gene type** rRNA
· **Attributes** Sequences, Unspliced (Gene)

BBtools (v.38.42) [391] was then used to remove rRNA reads (**Table 2.10**) using the bbduk.sh script with the following parameters:[12]

· **in1** R1_001.fastq.gz *#input read file 1*
· **in2** R2_001.fastq.gz *#input read file 2*
· **out1** _1R.fq *#output read file 1, with rRNA removed*
· **out2** _2R.fq *#output read file 2, with rRNA removed*
· **ref** rRNA.fasta *#rRNA reference file from Biomart*

Ribosomal RNA species characterisation

Reads were filtered against a FASTA file of all rRNA sequences obtained from Flybase using Bowtie2 on sensitive end-to-end setting. Reads were aligned to the *Drosophila* genome with STAR aligner, with outFilterMultimapNmax set to 1000. Feature counts used a filtered *Drosophila* GTF file containing only rRNA annotations, with the exclusion of pre-rRNA. Manual calculation of FPKM allowed the percentage abundance of each rRNA species to be determined.[13]

[12] by Jana Terenova
[13] by Elizabeth Dellar

Sample	rRNA contamination (%)
Cell Rep 1 R1	7.93
Cell Rep 1 R2	7.87
Cell Rep 2 R1	13.72
Cell Rep 2 R2	13.53
Cell Rep 3 R1	4.82
Cell Rep 3 R2	4.77
EV Rep 1 R1	87.19
EV Rep 1 R2	87.25
EV Rep 2 R1	97.37
EV Rep 2 R2	97.40
EV Rep 3 R1	85.02
EV Rep 3 R2	84.84

Table 2.10: Percentage of rRNA contamination removed by BBtools in either Cell or EV samples. Each condition has three replicates (Rep), with R1 and R2 representing the separate paired end read files for each replicate.

Read alignment

Reads which remained after rRNA contamination removal were aligned to the *Drosophila melanogaster* reference genome FASTA file and GTF file downloaded from Flybase (v. dmel_r6.26_FB2019 01). STAR Aligner (v.2.5.3a) [392] was used in a two step process.[14] Firstly, it was used to generate the index:

· **runMode** genomeGenerate *#index the genome file*

· **genomeDir** STAR_index *#path to file with genome file*

· **genomeFastaFiles** dmel-all-chromosome-r6.26.fasta *#FASTA file for genomic sequences*

· **sjdbGTFfile** dmel-all-r6.26.gtf *#path to GTF file with annotation*

· **sjdbOverhang** 139 *#length of the genomic sequence around the annotated junction to be used in constructing the splice junction database (read length -1)*

· **runThreadN** 8 *#central processing units (CPUs) to use*

Secondly, STAR Aligner was used to align reads:

· **genomeDir** STAR_index

· **readFilesIn** _1R.fq _2R.fq *#path to file with the paired end read files*

· **out SAMtype** BAM SortedByCoordinate *#output BAM file sorted by co-ordinate*

· **sjdbGTFfile** dmel-all-r6.26.gtf

· **outFilterMultimapNmax** 1 *#maximum number of loci a read is allowed to map to. 1 allows only uniquely mapped reads*

· **outFilterScoreMinOverLread** 0.3 *#alignment only output if it's score (normalised to read length) is higher than this value*

· **outFilterMatchNminOverLread** 0.3 *#alignment only output if the number of matched bases (normalised to read length) is higher than this value*

· **outReadsUnmapped** Fastx *#output unmapped and partially mapped reads in a separate FASTA file*

[14] in collaboration with Jana Terenova

· **runThreadN** 8
· **outFileNamePrefix** *#set prefix to add to output files*

Determining read alignment metrics

Read alignment metrics were determined using the Collect Alignment Summary Metrics Galaxy Tool (v.2.18.2.1) with default settings and the dmel-all-transcript-r6.26.fasta reference genome fasta file obtained from Flybase.

Determining read counts

HTSeq-count (v.0.9.1) [393] was used to generate read counts, a count of how many reads map to a specific gene in the genome.[15] The following parameters were used:

· **f** bam *#format of the input data is a BAM file*
· **r** position *#Data sorted via position using STAR is the input*
· **s** reverse *#data is from a reverse strand-specific assay*

For the cell samples, a gene was only kept for downstream analysis if it had a minimum of 5 reads aligned to it in one out of three replicates. For the EV samples, a gene was only kept for downstream analysis if it had a minimum of 1 read aligned to it in one out of three replicates.

Read count Principal component analysis

Differential gene expression analysis on read counts was performed using DESEQ2 (Galaxy Tool v.2.11.40.2) [394]. Normalised read counts for each condition, obtained via DESEQ2 analysis, were used to carry out a Principal Component Analysis (PCA) using the Scikit-learn Python package (v.0.23.1) [395, 396]. PCA is affected by scale and therefore, StandardScaler was used to standardise the data so that all datasets had a mean of 0 and a variance of 1. For visualisation purposes the PCA was reduced to three components. Two components accounted for 57.7% of the variance and three components accounted for 75.7% of the variance. Code for analysis and visualisation is shown in **Supplementary Listing 7.1**.

[15] in collaboration with Jana Terenova

Read count correlation analysis

Normalised read counts for each gene were compared between each biological replicate using the Pandas .corr() function (using the 'Spearman' method) in Python [397–400].

Determining read coverage

An adapted *Drosophila* GTF (r26.6) file was generated by removing all regions related to miRNAs and FBgn0002781 (as this is a trans-spliced gene). This GTF file was converted to a BED12 file using the Galaxy tool Convert GTF to BED12 (Galaxy Version 357) using default parameters. Bigwig files were generated from each Bam file (sorted by co-ordinate) using the Galaxy tool bamCoverage (Galaxy Version 3.3.2.0.0) using default parameters except for the following:

- **Bin size in bases** 50
- **Scaling/Normalization method** Normalise to reads per kilobase per mission (RPKM) *#allows for better comparison between cell and EV samples where read coverage is expected to be highly different*
- **Coverage file format** bigwig

Read coverage was determined using the Galaxy tool Gene Body Coverage (Bigwig) (Galaxy Version 2.6.4.1) using the *Drosophila* BED12 and bigwig files generated above. Read distribution skewness values were manually determined for the histogram read distribution data using the following equations [401, 402]:

$$g_1 = \frac{((x - \overline{\mathrm{x}})^3 f)}{((x - \overline{\mathrm{x}})^2 f))^{1.5}}$$

$$G_1 = \frac{\sqrt{n(n-1)}}{(n-2)} \times g_1$$

Standard error of skewness (SES) values were determined using the following equation [401, 403]:

$$SES = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}$$

Skewness test statistic, Z_{g1}, values were determined using the following equation [401, 404]:

$$Z_{\mathrm{g1}} = \frac{G_1}{SES}$$

Assigning reads to specific gene regions

Reads were aligned to specific gene regions using the bam flies generated from the STAR alignment and the Galaxy tool, RSeQC read_distribution.py 2.6.4.1. A GTF file with FBgn0002781 (trans-spliced gene) and miRNAs removed (as they prevented the program from running) was used. The bed12 file was generated using the USC tools gtftoGenePred and GenePredtoBed via the command line. Relative abundance was determined by dividing tags assigned to that region by the total number of tags.[16]

Determining abundance estimates

To estimate the expression of genes in the Cell and EV samples, FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values were generated for transcripts using Cufflinks (v.2.2.1) [405–408] using the following parameters:[17]

· **library-type** fr-first strand *#sequencing platform and protocol information (dUTP)*
· **G** dmel-all-r6.26.gtf *#GTF reference annotation file*
· **input** Aligned.sortedByCoord.out.bam *#Sorted data from STAR in the input*

Determining active genes

FPKM values were used to carry out a zFPKM normalization process to more accurately separate biologically relevant genes in the cell and EV samples, from the ultralow-expression of noisy genes. An average Cell and EV FPKM value was determined for each gene across the three biological replicates. Following this, the zFPKM transformation package (v.1.4.1) [409] together with R studio (v.1.1.463) [410, 411] was used to generate active gene lists for cell and EV samples, with a zFPKM cut-off of $\geq$ -3 utilised. Using these active gene lists, genes were categorised into one of three groups:

· Cell only
· EV only
· EV and cell

[16] by Elizabeth Dellar
[17] in collaboration with Jana Terenova

Determining gene type

The type of gene was determined for each of the active genes (e.g. messenger RNA (mRNA), non-coding RNA (ncRNA), pseudogene) using 'Transcript Type' identifiers for each gene downloaded from Biomart Ensemble 96. For any genes which did not have a type identifier in this database, their type was identified using Flybase directly. For downstream analysis, only protein-coding (mRNA) active genes were analysed.

Differential gene expression analysis

Differential gene expression (DE) analysis on read counts was performed using DESEQ2 as described in **Section 2.6.3**. The standard p-adjusted value cut-off of < 0.05 was used to determine significance. Genes were identified as significantly differentially expressed in EVs if their Log_2FC value was > 2. This was a stricter cut-off than the commonly used Log_2FC cut-off of > 1, to identify more relevant genes highly upregulated in EVs. This DE analysis identified 83 protein-coding genes within the 'Cell and EV' grouping which were significantly upregulated in EVs. These 83 genes, together with the 231 'EV only' protein-coding genes, formed the 'EV only plus EV enriched' active gene grouping, referred to as the 'EV+' group. Carrying out the DE classification method, in conjunction with the zFPKM classification, allowed us to be more stringent about which mRNAs were characterised as being present in EVs versus cells.

Gene ontology analysis and visualisation

Gene ontology (GO) analysis was carried out using the ViSEAGO package (v. 1.4.0) in R Studio (v. 1.2.5033) and R (v. 4.0.4) [410–412]. Example code for this analysis is shown in **Listing 2.4**, altering for biological process (BP), molecular function (MF) or cellular component (CC) as necessary. The default cut-off of $p < 0.01$ was used to define significantly enriched GO terms. This analysis was carried out to identify enriched GO terms for the protein-coding genes encoding 'EV+' transcripts or protein-coding genes encoding transcripts carrying any combination of the A, B, C or D EV motifs (ABCD). These groups were used within the 'genSel'

option. The 'background' was all *Drosophila melanogaster* protein-coding genes. For the ABCD cellular component analysis, minClusterSize was set to 1 instead of 2 to allow the analysis to proceed due to the presence of small cluster sizes. For the ABCD biological process analysis, due to a large number of enriched terms, deepSplit was set to 1 instead of 2 to allow the formation of a smaller number of clusters, to aid downstream analysis.

```
1
2  EntrezGene<-ViSEAGO::EntrezGene2GO()
3
4  # load GO annotations from EntrezGene for Drosophila melanogaster
5  myGENE2GO<-ViSEAGO::annotate(
6      "7227",
7      EntrezGene
8  )
9
10 # create topGOdata for BP
11 BP_EV<-ViSEAGO::create_topGOdata(
12     geneSel=EV,
13     allGenes=background,
14     gene2GO=myGENE2GO,
15     ont="BP",
16     nodeSize=5
17 )
18
19 # TopGO test using classic algorithm
20 classic_BP_EV<-topGO::runTest(
21     BP_EV,
22     algorithm ="classic",
23     statistic = "fisher"
24 )
25
26 # merge results from topGO
27 BP_sResults_EV<-ViSEAGO::merge_enrich_terms(
28     Input=list(
29         condition=c("BP_EV","classic_BP_EV")
30     )
31 )
32
33 # initialise
34 myGOs_BP_EV<-ViSEAGO::build_GO_SS(
35     gene2GO=myGENE2GO,
36     enrich_GO_terms=BP_sResults_EV
37 )
38
39 # compute all available Semantic Similarity (SS) measures
40 myGOs_BP_EV<-ViSEAGO::compute_SS_distances(
41     myGOs_BP_EV,
42     distance="Wang"
43 )
44
45 # GOterms heatmap with the default parameters
46 Wang_clusters_wardD2_BP_EB<-ViSEAGO::GOterms_heatmap(
47     myGOs_BP_EB,
48     showIC=TRUE,
49     showGOlabels=TRUE,
50     GO.tree=list(
51         tree=list(
52             distance="Wang",
53             aggreg.method="ward.D2"
54         ),
55         cut=list(
56             dynamic=list(
57                 pamStage=TRUE,
58                 pamRespectsDendro=TRUE,
59                 deepSplit=2,
60                 minClusterSize =2
61             )
62         )
63     ),
64     samples.tree=NULL
65 )
66
67
68 # calculate semantic similarities between clusters of GO terms
69 Wang_clusters_wardD2_BP_EV<-ViSEAGO::compute_SS_distances(
70     Wang_clusters_wardD2_BP_EV,
71     distance=c("max", "avg","rcmax", "BMA")
72 )
73 )
74
75 # GOclusters heatmap
76 Wang_clusters_wardD2_BP_EV<-ViSEAGO::GOclusters_heatmap(
```

```
77      Wang_clusters_wardD2_BP_EB,
78      tree=list(
79          distance="BMA",
80          aggreg.method="ward.D2"
81      )
82  )
83
84  # display the GOClusters heatmap
85  ViSEAGO::show_heatmap(
86      Wang_clusters_wardD2_BP_EV,
87      "GOclusters"
88  )
89
```

Listing 2.4: Example code for ViSEAGO GO term enrichment analysis [413].

Motif discovery

The 3'UTRs of the transcripts encoded for by each active gene in the 'Cell only' and 'EV only plus EV enriched' groups were downloaded from Biomart Ensemble 97. Motif analysis was carried out using MEME Suite (v.5.0.4) [414, 415]. Multiple Em for Motif Elicitation (MEME, MEME Suite) was used to identify deferentially enriched motifs in the EV loaded RNAs. Only 3'UTRs $\geq$ 30 nucleotides long were utilised for motif detection via MEME to ensure sequences analysed were longer than the maximum width of motifs detected (which was defined as 25-nt).

MEME was used with the following parameters:

· **dna** *#input files are DNA alphabet files*

· **objfun** de *#differential enrichment analysis, comparing motif occurrences in test (e.g. EV only plus EV enriched') versus control (e.g. Cell only)*

· **mod** zoops *#each input sequence may contain zero or one occurrence of the motif*

· **V** *#print extensive status messages to terminal*

· **nmotifs** 30 *#MEME will stop searching for motifs once 30 have been found*

· **maxw** 25 *#search for motifs with width $\leq$ 25 nucleotides long*

· **neg** Cell_only_active_above30bp_mRNA.fasta *#name of the file containing the control sequences in FASTA format*

Motif scanning

Find Individual Motif Occurrences (FIMO, MEME Suite) [416] was then used to search for the thirty motifs identified by MEME in the following groups:

· EV only plus EV enriched (referred to as 'EV+')

· Cell only

· All transcripts regardless of cell or EV classification

FIMO was run on all available 3'UTR / 5'UTR / CDS sequences within each group, regardless of length. FIMO was used with default parameters. FIMO hits with a q-value of < 0.05 were considered significant. Only positive strand hits were utilised for downstream analysis. To account for any differences in nucleotide frequencies, an estimated Markov model was first generated using the fasta-get-markov MEME program. This model was then used as the background file (*bfile*). This allowed FIMO to better score and assign significance to the motif hits.

The following parameters were used for the fasta-get-markov program (MEME Suite) [414]:

· **dna** *#input files are DNA alphabet files*

· **bfile** *#input appropriate Markov background model of nucleotide base probabilities*

As an additional control, FIMO was used to search for motif occurrences in shuffled version of the sequences within the 'Cell only' and 'EV only plus EV enriched' groups. Shuffling was carried out using shuffleseq (Galaxy Tool v.5.0.0) [417, 418]. Taken together, this pipeline identified six EV-enriched motifs. Four of these motifs were selected for downstream analysis due to their high prevalence in 'EV+' transcripts, low prevalence in 'Cell only' transcripts and their high degree of sequence conservation, identified via their bit score.

Motif comparisons

MAST was one tool used to compare motifs [419]. MAST determines the pairwise correlations between each motif pair. The correlation between two motifs is the maximum sum of Pearson's correlation coefficients for aligned columns divided by the width of the shorter motif. The maximum is found by trying all alignments of the two motifs. Motif with a correlation of greater than 0.6 are considered similar.

The following parameters were used:

· **norc** *#do not score the reverse complements of target motifs.*

TOMTOM [420, 421] was an alternative tool utilised to compare the novel motifs with motifs of know *Drosophila* proteins binding sites or previously published EV motifs (**Table 3.6**. The following parameter was used:

· **norc** *#do not score the reverse complements of target motifs.*

Motif 3'UTR positions

The FIMO output provided motif hit locations along the 3'UTR. Overlapping motif hits for the same motif were observed, disrupting downstream motif quantification and visualisation. Overlapping motif hits were removed using the code described in **Supplementary Listing 7.2**.[18] The remaining nucleotide start and end positions for each motif hit were then identified. The start position values were normalised to the transcripts 3'UTR length so that comparisons between transcripts could be made. Skewness values were determined using GraphPad Prism. Standard error of skewness (SES) and skewness test statistic (Z_{g1}) values were determined as previously described in **Section 2.6.3**.

Motif combinations

Using the motif hit information contained within the *fimo.tsv* output of FIMO, transcripts containing specific combinations of the four chosen motifs were identified. Percentage prevalence of each motif combination was determined at the gene level. To account for the different number of transcripts encoded for by 'Cell only' versus 'EV only plus EV enriched' transcripts, normalised percentage prevalence values were determined, as described in **Figure 2.3**.

Visualisation of transcript combinations was carried out as follows. Transcripts which contained at least one occurrence of each of the four motifs were extracted and motif start and stop positions (normalised to the full 3'UTR length) were plotted using a custom Python script (**Supplementary Listing 7.3**). Multiple transcripts for the same gene which contained the same motif pattern were identified via visual analysis and removed to generate a 'shortlist' of transcripts containing all four motifs. This shortlist of transcripts is what was used for the motif order and structural analysis.

[18] Code developed by Jana Terenova

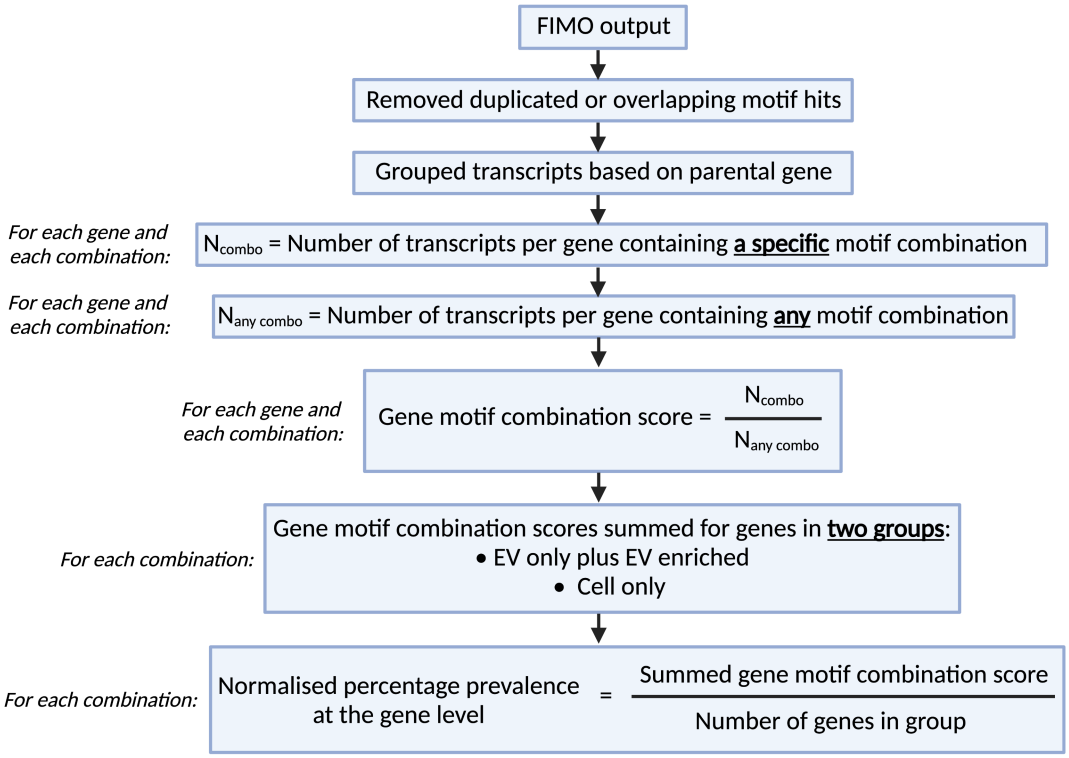


Figure 2.3: Flow diagram depicting the process of determining normalised percentage prevalence at the gene level for each motif combination.

Motif order

Using the 'shortlist' of transcripts which contain motifs A, B, C and D, the order in which these motifs appear in a 3'UTR sequences was analysed. This was carried out using a custom Python script (**Supplementary Listing 7.4**).[19] Because 3'UTRs contained multiple occurrences of the same motif along the 3'UTR, the code was developed to determine all possible motif combinations within that 3'UTR sequence, represented in **Figure 2.4**.

[19]Code developed by Jana Terenova

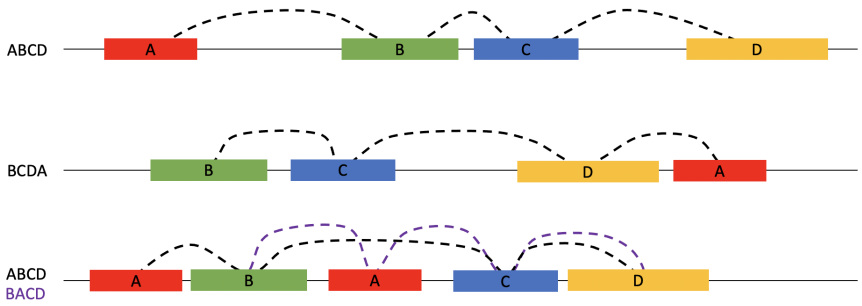


Figure 2.4: Schematic diagram to depict how all possible motif orders present within a 3'UTR were determined. This figure highlights the motif orders determined for three independent 3'UTR sequences.

RNAfold secondary structure analysis

To determine the percentage of bases paired in motif occurrences within the 3'UTRs of shortlisted transcripts carrying the ABCD motif combination, RNAfold (v. 2.4.14) [422] analysis was carried out using the following parameters:

- **p** *#Calculate the partition function and base pairing probability matrix*
- **d** 2 *#How to treat "dangling end" energies for bases adjacent to helices in free ends and multi-loops (default='2')*
- **T** 25 *#Rescale energy parameters to given temperature (°C)*
- **noLP** *#Produce structures without lonely pairs (helices of length 1)*

Code shown in **Supplementary Listing 7.5** was used to combine the FIMO output for the shortlisted transcripts (with overlapping hit removed as described in **Section 2.6.3**) with the RNAfold output to calculate the number of bases in each motif which were paired versus unpaired.

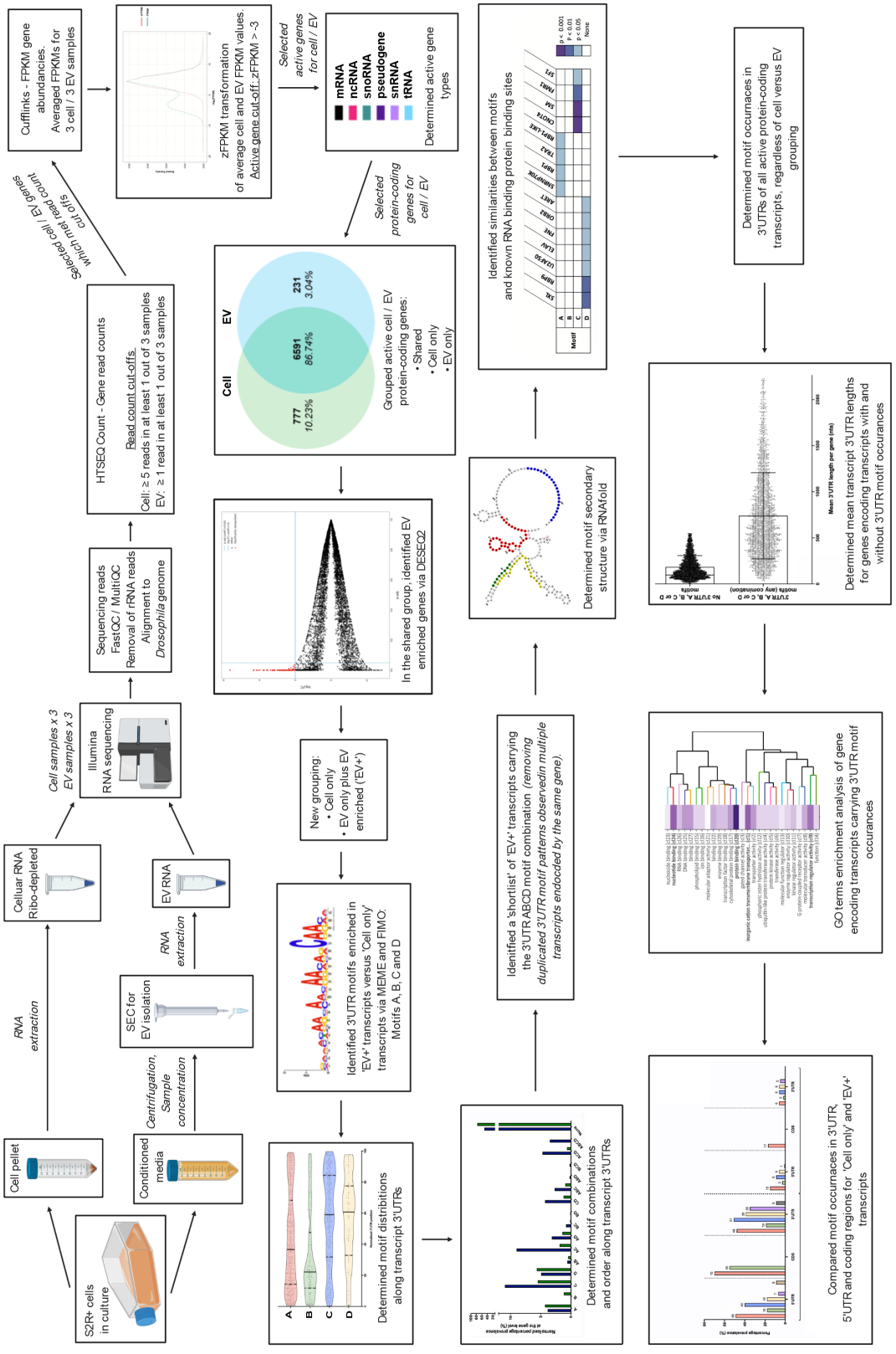


Figure 2.5: Schematic depiction of the bioinformatic pipeline described in **Chapter 3**.

3

Identification of novel EV-enriched motif combinations in *Drosophila melanogaster*

Contents

3.1 Introduction 76
3.2 Characterisation of cellular and EV RNA profiles . . . 77
3.2.1 *Drosophila* cells and EVs showed distinct RNA profiles, with EV RNA profiles appearing more variable 79
3.2.2 Read coverage distribution for cellular and EV RNAs showed no distinct 5-prime or 3-prime bias 81
3.2.3 Cell and EV RNA reads predominantly mapped to coding regions 82
3.2.4 Comparing *Drosophila* cell and EV RNA profiles highlighted the presence of EV-specific RNA cargo 83
3.2.5 Cellular and EV RNAs were predominately protein-coding 85
3.2.6 Specific transcripts are enriched in EVs compared to the cell 87
3.3 Novel 3'UTR motifs identified within EV-enriched mRNAs 90
3.3.1 Discovery of novel *Drosophila* EV-enriched motifs 90
3.3.2 The six novel *Drosophila* EV-enriched motifs were not analogous to one another 93
3.3.3 Novel *Drosophila* EV-enriched motifs showed no significant sequence similarity with known mammalian EV-enriched motifs 93
3.3.4 Each motif presented a unique distribution pattern along the 3'UTR 95
3.3.5 Motifs commonly occur in combinations within the 3'UTR 96
3.3.6 3'UTR motifs were specifically arranged within EV mRNAs 98
3.3.7 RNA motifs were part of specific RNA secondary structures 100
3.3.8 RNA motifs potentially interact with specific RNA-binding proteins 101
3.3.9 Novel *Drosophila* motifs were more prevalent in longer transcripts 103
3.3.10 EV-enriched *Drosophila* motifs present in transcripts with specific gene product attribute predictions 105

3.3.11 Novel motif occurrences found within mRNA coding regions and the 5'UTR 107
3.4 Discussion 110
3.4.1 Characterisation of *Drosophila* S2R+ cell and EV RNAs 110
3.4.2 Mechanisms of selective EV-loading in *Drosophila* . . . 113
3.5 Future Work 121
3.5.1 Determining RNA binding protein and miRNA interactions with EV-enriched motifs 121
3.5.2 Biological validation that the novel *Drosophila* EV-enriched motifs encourage the EV-loading of longer transcripts . 122
3.5.3 Comparison of surface EV RNAs with those loaded within the EV lumen 122
3.6 Chapter Summary 124

3.1 Introduction

The EV-mediated transfer of RNA cargo provides a means for cells to remotely manipulate gene expression and cell function; however, the processes involved RNA incorporation into EVs, and the extent to which this incorporation is selective, is not fully understood. Whilst the mechanisms involved in miRNA EV-loading have been more extensively studied [74, 216, 217, 233, 234, 249, 250, 254], mRNA EV-loading has been less widely explored. In addition, the level of evolutionary conservation of these basic mechanisms is not fully appreciated, providing an opportunity to expand our fundamental knowledge of EV biology.

Accordingly, an RNA-sequencing and bioinformatic analysis of *Drosophila* S2R+ cell and EV RNA content was carried out. The subsequent bioinformatic analysis was designed to identify unique EV-enriched mRNA features which may enhance their EV incorporation and long-range transfer. This chapter presents results associated with the methods described in **Section 2.6**. A summary of this pipeline is depicted in **Figure 2.5**.

3.2 Characterisation of cellular and EV RNA profiles

Optimisation of EV isolation from *Drosophila* S2R+ cells and the characterisation of EVs is described in **Supplementary Section 7.1.2**. Total RNA was extracted from biological triplicates of *Drosophila* S2R+ cells and the EVs they released.[20] RNA concentration and purity values are shown in **Table 3.1**. To characterise the RNA extracted from cell and EV samples, each sample underwent BioAnalyser capillary electrophoresis (**Figure 3.1**).[21] Small RNA species were highly abundant in all three EV RNA replicates, ranging from approximately 20 to 500 nucleotides (nts). Cell RNA samples also showed a small RNA peak from approximately 30 to 250 nts. Secondary peaks were observed in cell RNA samples at the 18S and 28S ribosomal RNA size ranges (approximately 1500 nts and 3000 nts, respectively). These ribosomal peaks were largely absent in the EV RNA samples. Cellular RNA, therefore, was ribo-depleted prior to sequencing. Studies have shown that even with ribo-depletion treatments on both cell and EV samples, rRNAs still appeared enriched in EVs, likely due to differences in depletion efficiencies between cells and EVs [223]. For this reason, as well as the lack of EV rRNA Bioanalyser peaks and low EV RNA concentrations, EV samples were not ribo-depleted.

		Concentration Nanodrop (ng/µl)	**A260/280**	**A260/230**	**Concentration Qubit (ng/µl)**
Cell	Rep 1	3929.6	2.17	2.35	-
	Rep 2	6313.9	2.17	2.18	-
	Rep 3	2422.5	2.2	2.19	-
EV	Rep 1	80	1.88	1.54	45
	Rep 2	14.1	1.97	0.23	8.15
	Rep 3	-	2	2.15	36.7

Table 3.1: RNA concentrations and absorbance ratios obtained for cell and EV RNA samples.[22]

[20] by Elizabeth Dellar
[21] in collaboration with Elizabeth Dellar
[22] Data generated by Elizabeth Dellar

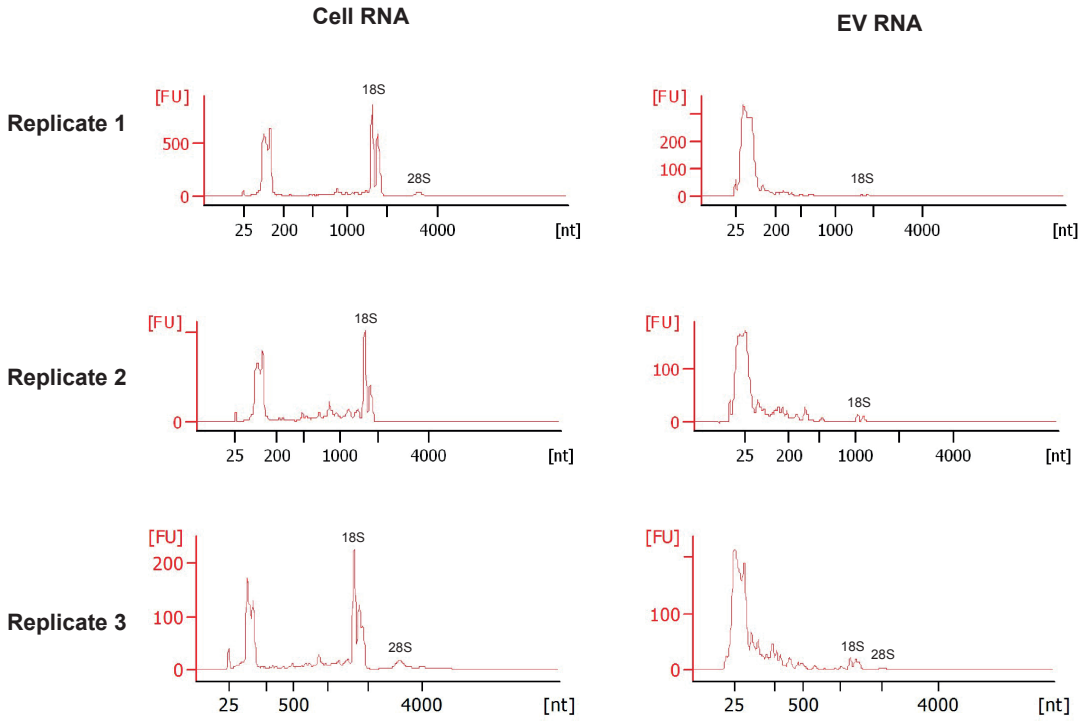


Figure 3.1: Bioanalyser traces for the cell and EV RNA samples. Small RNA species were highly abundant in cell and EV samples. 18S and 28S ribosomal RNA peaks were observed in cell samples but were largely absent in EV samples.[23]

Cellular and EV RNA samples underwent Illumina RNA-sequencing analysis and sequencing reads were processed as described in **Section 2.6**. Interestingly, although 28s and 18s rRNA peaks were largely absent in EV RNA Bioanalyser profiles, a large proportion of EV reads aligned to rRNA sequences (ranging from 84.84% to 97.37% per sample, **Table 2.10**). Cellular RNA-sequencing reads contained a much smaller proportion of reads aligning to ribosomal RNA (ranging from 4.77% to 13.72% per sample), likely due to the cellular RNA ribo-depletion treatment. These results suggested that ribosomal RNAs made up a large component of the EV loaded RNA. Further analysis of rRNA species confirmed that 18s and 28s were the most abundant rRNAs species present within EVs (56.9 ± 5.57% and 40.2 ± 3.12% of total abundance, respectively), with 5.8s rRNA much less prevalent (2.6 ± 2.73%) (**Figure 3.2**).[24]

[23]Data generated in collaboration with Elizabeth Dellar
[24]by Elizabeth Dellar

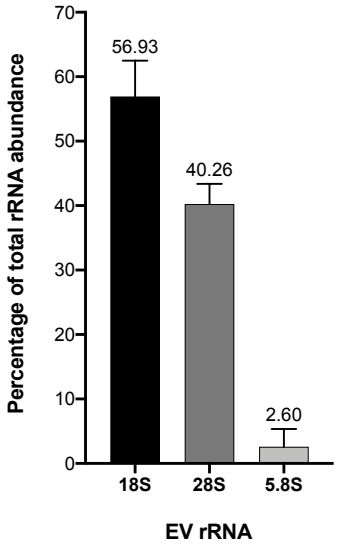


Figure 3.2: Percentage of total RNA abundance made up of 18s, 28s and 5.8s rRNA species. Shown are the values for three biological replicates and standard deviations. Highlighted above each bar are the mean values for each rRNA species.[25]

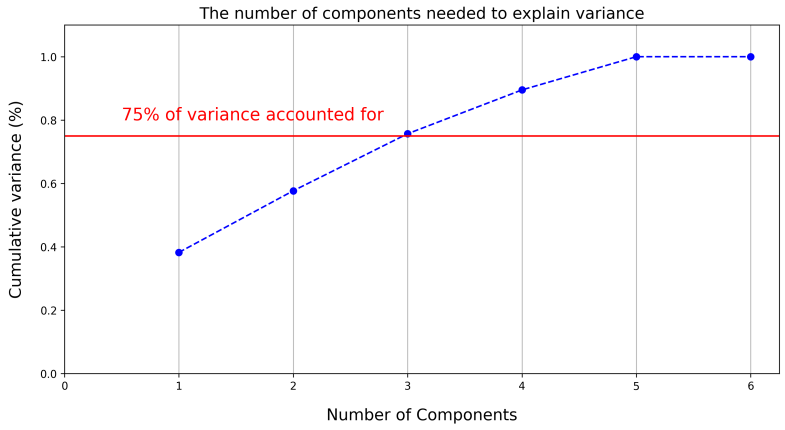
The disparity between the Bioanalyser profiles and RNA-sequencing analysis suggested that EV rRNAs may be fragmented and thus, contributing to the smaller RNA BioAnalyser peaks. In the future it will be interesting to study the fragmentation and EV-loading of ribosomal RNAs; however, as this was not the focus of this bioinformatic analysis, cell and EV sequencing reads aligning to *Drosophila* rRNA sequences were removed prior to downstream bioinformatic analysis.[26]

3.2.1 *Drosophila* cells and EVs showed distinct RNA profiles, with EV RNA appearing more variable

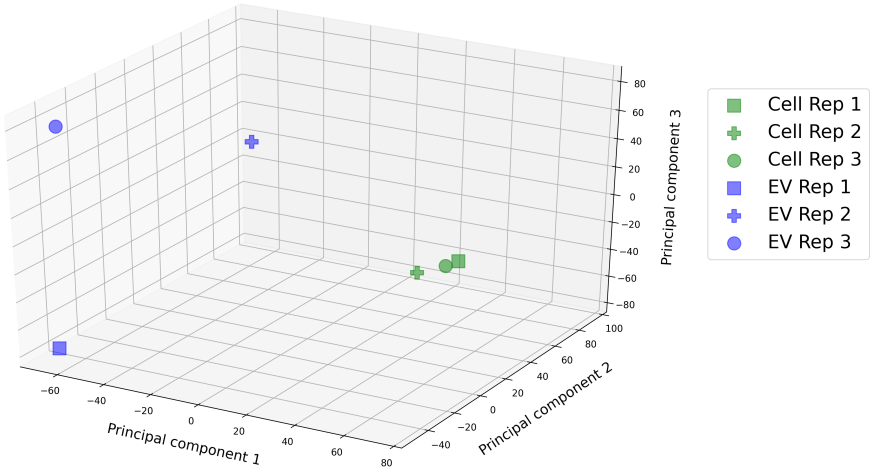
An initial comparison of cellular and EV RNA profiles was carried out using a Principal Component Analysis (PCA). Normalised read counts per gene (determined via DESEQ2 analysis) were compared for each replicate of the cell and EV samples. A three-component PCA accounted for 75.7% of the variance (**Figure 3.3i**). In the 3D PCA, cell and EV RNA samples were plotted at a distance from one another, highlighting the presence of distinct cellular and EV RNA profiles (**Figure 3.3ii**). Cellular RNA replicates were closely clustered, suggesting consistency in cellular

[25]Data generated by Elizabeth Dellar
[26]by Jana Terenova

RNA content. All three EV replicates, however, were distinct from one another in the y and z plane, suggesting a higher degree of variability in EV RNA compared to cellular RNA (**Figure 3.3ii**). Direct correlation of normalised read counts between replicates also reflected these patterns. Cellular RNA replicates were strongly correlated ($\rho = 0.95$), with EV RNA replicates less strongly correlated ($\rho = 0.87$ - 0.88) (**Supplementary Figure 7.10**), again suggesting consistency in cellular RNA content and variability in EV RNA cargo.



(i) The percentage variance accounted for with an increasing number of components included in the analysis.



(ii) Three-component analysis.

Figure 3.3: Principal component analysis (PCA) of cellular and EV RNA normalised read counts for each biological replicate (Rep.)

3.2.2 Read coverage distribution for cellular and EV RNAs showed no distinct 5-prime or 3-prime bias

After the initial comparison of normalised read counts which revealed distinct cellular and EV RNA profiles, it was important to compare read coverage and distribution between these samples. The depth and breadth of read coverage along *Drosophila* transcripts was determined via the analysis described in **Section 2.6.3**. A reads per kilobase per million mapped reads (RPKM) normalisation method was used to allow for better comparison between cell and EV samples. This was because, on average, the total number of mapped reads in proper pairs was much higher in cell samples (12.2 ± 2.6 million) compared to EV samples (0.2 ± 0.06 million) (**Supplementary Table 7.4**), likely due to the higher cellular RNA concentrations obtained. The read coverage distributions for cell and EV RNA samples are shown in **Figure 3.4**. Normalised 'total read count' wigsum values were used to account for any variation in read length between samples where, for example, a wigsum of 100 million equals the coverage achieved by 1 million reads with a length of 100 nucleotides or 2 million reads with a length of 50 nucleotides. This was a precautionary measure, however, as results shown in **Supplementary Figure 7.11** confirmed that the mean size of paired reads was not significantly different between cell and EV samples (Cell average = 143.4 ± 1.3 nts, EV average = 142.6 ± 2.5 nts). On average, the read distribution for cell samples peaked at position 34 with 5.89 ± 0.47 million wigsum, with EV profiles peaking at position 29 with 5.4 ± 0.42 million wigsum (**Figure 3.4**). Standard deviations were small, suggesting consistent read coverage profiles between all three biological replicates.

A small 5' skew was observed for both cellular and EV read distributions; however, for the reverse-stranded RNA-sequencing protocol used in this analysis, a small 5' skew is to be expected [423]. The 5' skew for cells samples was 4.94 ± 0.0113 x 10^{-6} (Skew ± Standard Error of Skew) and 6.14 ± 0.0119 x 10^{-6} in EV samples. The test statistic of these skews (Z_{g1}) were 0.044 and 0.052, respectively, not reaching the Z_{g1} critical value of 2 and therefore, not deemed significant. In addition, no strong 3' bias was observed, suggesting these samples were not undergoing RNA

degradation by 3' exonucleases [424–426]. These results confirmed that uneven read distribution from the 5' to 3' end of transcripts would not disturb downstream quantifications and analysis [427].

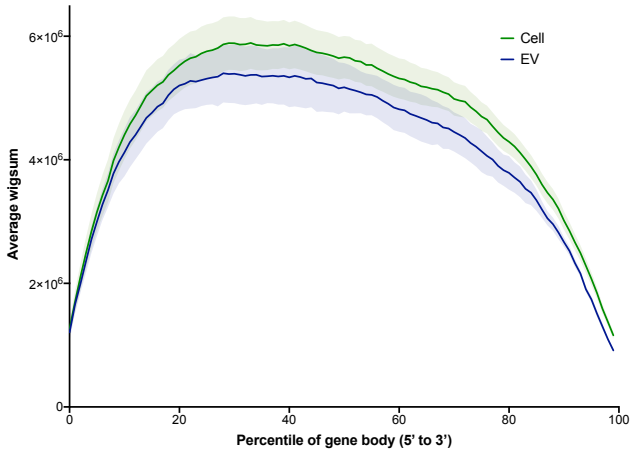


Figure 3.4: Gene body coverage for all non trans-spliced or miRNA genes in cellular and EV RNA-sequencing samples. Green and blue lines represent the mean values of the three biological replicates for cells and EVs, respectively. Shaded areas represent standard deviations. No significant 5' or 3' skews were observed.

3.2.3 Cell and EV RNA reads predominantly mapped to coding regions

Further analysis of how mapped reads were distributed over genome features (coding, 5'UTR, 3'UTR or intron regions) in cell and EV RNA fractions was carried out.[27] As shown in **Figure 3.5**, most reads were mapped to coding regions in both cell and EV samples (74.97 ± 1.57% and 69.73 ± 2.86%, respectively), followed by the 3'UTR (11.10 ± 0.27% and 12.40 ± 0.46%, respectively). The lowest percentage abundance was observed for the 5'UTR and intron regions (Cell 5'UTR = 7.11 ± 0.47, EV 5'UTR = 8.27 ± 0.49%, Cell intron = 6.33 ± 1.80%, EV intron = 8.63 ± 1.86%). Interestingly, the proportion of reads aligning to the coding region was significantly lower in EV RNA samples compared to cellular RNA samples ($p = 0.0006$). The proportion of reads aligning to 3'UTR, 5'UTR and intron regions was not statistically different in cell versus EV samples (**Supplementary Table 7.5**).

[27] by Elizabeth Dellar

These results highlighted that whilst both cellular and EV RNA reads predominantly aligned to coding regions, this proportion decreased in EV RNA samples.

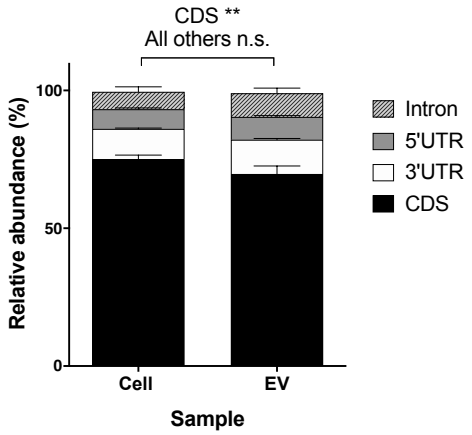
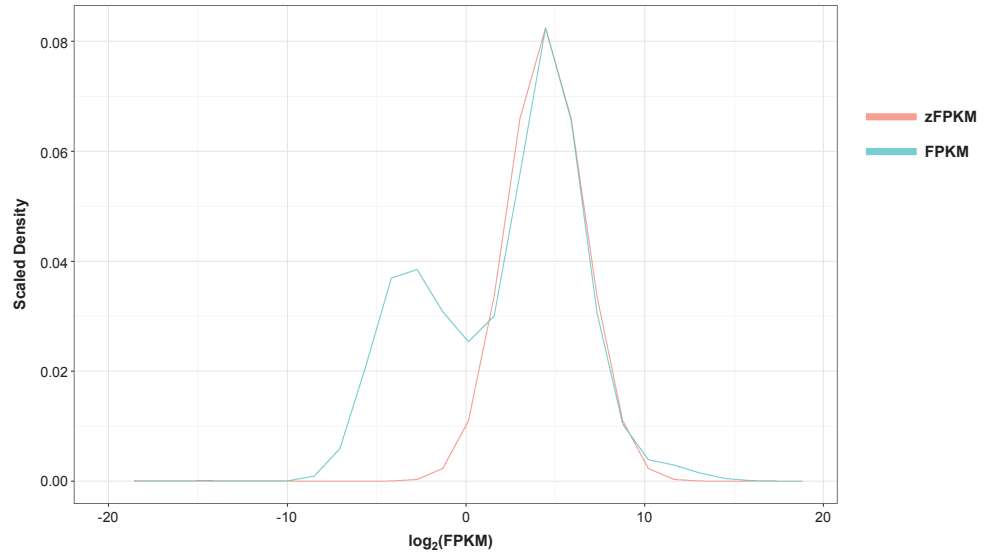


Figure 3.5: Read distribution within gene regions for all non trans-spliced or miRNA genes analysed in cell and EV samples. Regions include coding regions (CDS), 3-prime untranslated regions (3'UTRs), 5-prime untranslated regions (5'UTRs) or intronic regions (introns). The statistical analysis is shown in **Supplementary Table 7.5**. Mean ± standard deviation is shown.[28]

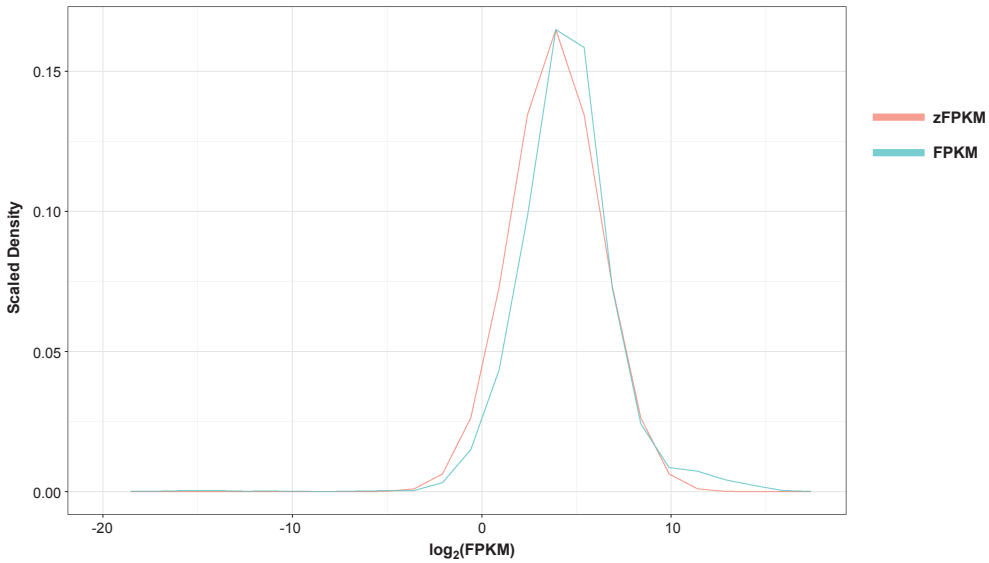
3.2.4 Comparing *Drosophila* cell and EV RNA profile highlighted the presence of EV-specific RNA cargo

FPKM values (Fragments Per Kilobase of transcript per Million mapped reads) were averaged across the three biological replicates for cellular and EV RNAs. Due to the differences in the $\log_2$FPKM distributions observed for cells and EVs, likely due to the higher read depth achieved in cell samples compared to EV samples, FPKM values were transformed into zFPKM values, resulting in normal distributions (**Figure 3.6**). This transformation allowed the biologically relevant genes in cells and EVs to be accurately separated from the ultralow-expression, noisy genes [409]. Only genes with a zFPKM greater than -3 were considered active and carried forward for further analysis. A secondary filter was added to ensure that active genes met the required read count cut-offs defined in **Section 2.6.3**.

[28]Data generated by Elizabeth Dellar



(i) Frequency distribution of average Cell FPKM values and these values after the zFPKM transformation.



(ii) Frequency distribution of average EV FPKM values and these values after the zFPKM transformation.

Figure 3.6: zFPKM transformation of average Cell and EV FPKM values. Note how the zFPKM transformation transforms the datasets into normal distributions, allowing the -3 zFPKM cut-off to be applied to identify 'active genes.' Graphs adapted from those generated by the zFPKM R studio package.

The active gene lists generated for the cell and EV samples were then compared, allowing genes to be defined as 'Cell and EV', 'Cell only' or 'EV only'. The number of genes within each of these groups are shown in **Table 3.2**. This comparison showed that *Drosophila* cells and EVs share the majority of their active gene content (84.24 ± 0.40%). A smaller proportion of these active genes were present in either cell or EV samples only (12.37 ± 0.37% and 3.39 ± 0.20%, respectively).

Group	Total number of active genes
Cell and EV	6842
Cell only	1005
EV only	275

Table 3.2: Number of active genes in the three groups: 'Cell and EV', 'Cell only' and 'EV only'.

3.2.5 Cellular and EV RNAs were predominately protein-coding

To identify differences in the types of RNA present within *Drosophila* cells and EVs, transcript types encoded by the active genes within the 'Cell and EV', 'Cell only' or 'EV only' groups were determined. **Figure 3.7** describes the prevalence of each of the following transcript types within these groups:

- mRNA - protein-coding messenger RNA
- ncRNA - non-coding RNA
- snoRNA - small nucleolar RNA
- pseudogene - non-functional gene copies of functional genes
- snRNA - small nuclear RNA
- tRNA - transfer RNA

Results showed that the majority of genes in all three groups were protein-coding (Cell only = 77.31%, Cell and EV = 96.33%, EV only = 84%). A higher proportion of genes encoding ncRNAs were present in the 'Cell only' versus the 'EV only' category (Cell only = 18.51%, EV only = 10.18%). Comparable prevalence patterns between 'Cell only' and 'EV only' genes were observed for other non-coding or regulatory RNAs such as snoRNA (Cell only = 2.39%, EV only = 2.55%), pseudogenes (Cell only = 1.29%, EV only = 1.89%) and snRNA (Cell only = 0.5%, EV only = 1.09%). Interestingly, these non-coding and regulatory RNA types were

often not shared between cells and EVs (Cell and EV: snRNA = 0.16%, pseudogenes = 0.28%, snoRNA =1.4%, ncRNA = 1.8%). tRNAs were only observed in the 'EV only' group. Taken together, these results showed that the majority of cell and EV RNA content was coding RNA and a large proportion of these protein-coding genes were shared between cell and EV fractions (86.74%, **Figure 3.8**).

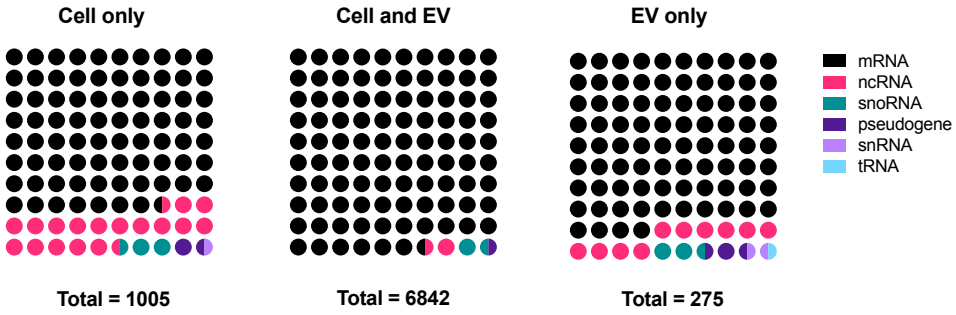


Figure 3.7: Percentage prevalence of transcripts within each transcript type category for the 'Cell only', 'Cell and EV' and 'EV only' active gene groups. One circle represents 1%. Percentage data shown in **Supplementary Table 7.6**.

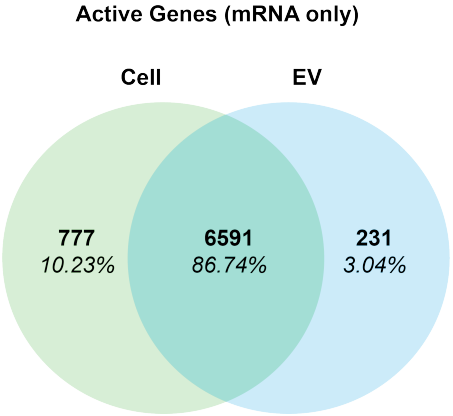


Figure 3.8: Active genes encoding mRNA transcripts within the 'Cell only', 'EV only' and 'Cell and EV' groups. Shown in bold are counts and shown in italics is this count as a percentage.

3.2.6 Specific transcripts are enriched in EVs compared to the cell

Despite a high proportion of protein-coding genes present within both cells and EVs (86.74%), 3.04% of mRNAs were exclusive to EVs (**Figure 3.8**). Differential expression (DE) analysis of the 6591 protein-coding genes shared between cells and EVs (as described in **Section 2.6.3**) identified 83 mRNAs significantly enriched in EVs compared to cells (**Figure 3.9**). To enrich the downstream analysis, these 83 genes were added to the 231 'EV only' protein-coding genes to generate an EV only plus EV enriched gene grouping, referred to as the 'EV+' transcripts (**Table 3.3**). Downstream analysis focused on investigating these mRNAs to identify possible features which may enhance their EV-loading.

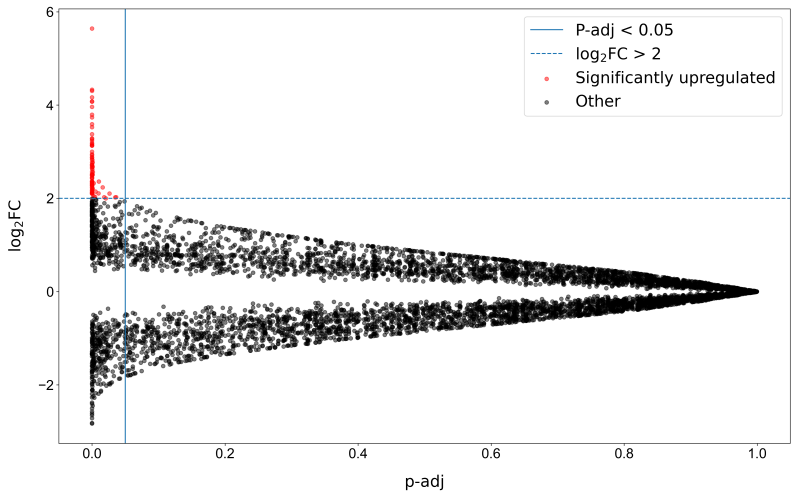


Figure 3.9: Differential expression analysis of protein-coding genes within the 'Cell and EV' active gene group. The 83 genes significantly upregulated in EVs are shown in red. The p-adj cut off of < 0.05 is highlighted by a continuous blue line. The $\log_2$Fold Change (FC) cut off of > 2 is highlighted by a dashed blue line.

Group	Number of protein-coding genes
Cell only	777
EV only plus EV enriched	314

Table 3.3: Number of active protein-coding genes in the 'Cell only' and 'EV only plus EV enriched' groups.

Gene ontology (GO) analysis (described in **Section 2.6.3**) highlighted a number of predicted biological processes, molecular functions and cellular components related to the genes encoding 'EV+' transcripts (**Figure 3.10**, **Supplementary Tables 7.7 - 7.9**). The most prevalent biological processes were related to development and pattern specification, roles known to be associated with EVs [139]. Interestingly, biological processes associated with cytoskeleton function and cell communication were also highly prevalent. This was also reflected in the enriched molecular functions related to acting binding. These results highlighted predicted roles for these gene products in cytoskeleton processes, perhaps defining mRNA or protein localisation within the cell. Encouragingly, the plasma membrane, cell projections and cell-cell junctions were enriched cellular compartment GO terms, suggesting a possible localisation of these gene products to membrane surfaces, perhaps enhancing their mRNA EV incorporation.

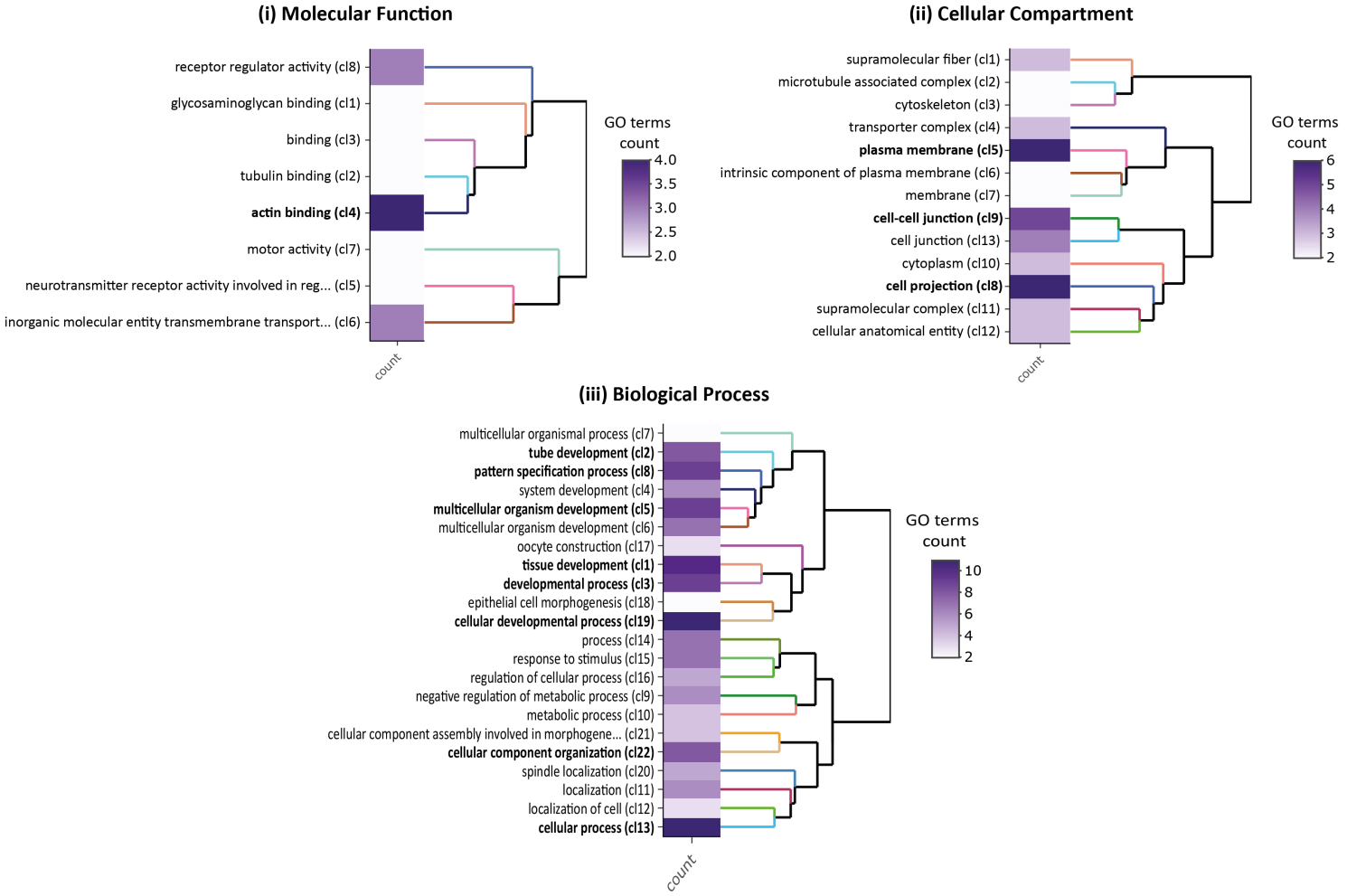


Figure 3.10: Significantly enriched (i) molecular function, (ii) cellular component and (iii) biological process gene ontology (GO) terms associated with genes encoding 'EV+' transcripts. GO terms were clustered and summarised using heatmaps. Incomplete text in figure reads: 'neurotransmitter receptor activity involved in regulation of postsynaptic membrane potential', 'inorganic molecular entity transmembrane transporter activity' and 'cellular component assembly involved in morphogenesis.' Cl = cluster. Bolded text indicates highly prevalent GO term clusters.

3.3 Novel 3'UTR motifs identified within EV-enriched mRNAs

The presence of EV-enriched mRNAs related to particular predicted functions suggested the possible existence of an mRNA sorting mechanism in *Drosophila*. mRNA EV incorporation may be controlled via the regulation of mRNA sub-cellular localisation, altering the likelihood for mRNAs to be incorporated into MVBs or released within a microvesicle at the plasma membrane [237]. The 3'UTR includes sequences which localise mRNAs in specific sub-cellular localisations, often upon binding to RNA binding proteins [428, 429]. Indeed, sequence motifs have been shown to be enriched within the 3'UTR of EV-enriched RNAs in mammalian systems [234, 245–247, 249], for example, the 25-nt 3'UTR motif enriched in microvesicle mRNAs previously which can enhance mammalian mRNA EV-loading [248]. Therefore, downstream analysis focused on probing the 3'UTR of *Drosophila* EV mRNAs for the presence of novel 3'UTR motif sequences which may modulate EV RNA incorporation in this model organism. Although this analysis focused on motif discovery within the 3'UTR, novel motifs may remain undiscovered within the 5'UTR and CDS regions and this is an area requiring further investigation.

3.3.1 Discovery of novel *Drosophila* EV-enriched motifs

The MEME Suite package [414] was utilised to discover and compare motifs as described in **Section 2.6**. A differential enrichment analysis was carried out using MEME to identify motifs (up to 25-nt in length) enriched in the 3'UTR regions of the 'EV+' mRNAs, compared to the same region in 'Cell only' mRNAs. Only 3'UTRs over 30-nt were utilised to ensure that the 3'UTR sequence length was always longer than the maximum motif length to be discovered (25-nt). The number of mRNAs within each category is shown in **Table 3.4**. FIMO was then used to scan the same sequences to identify only significant occurrences of these motifs.

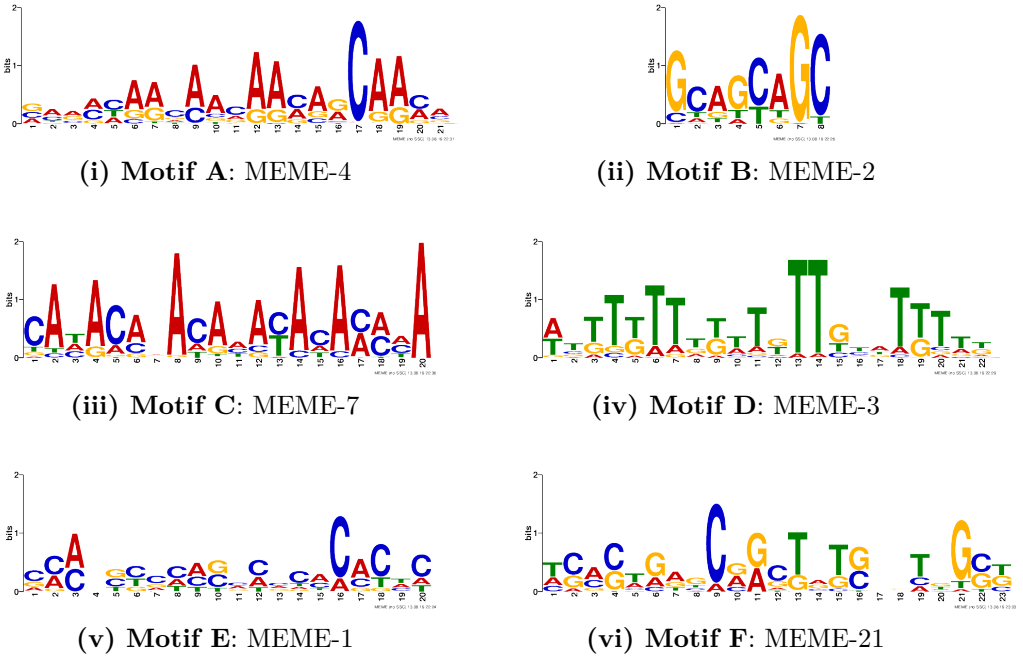
Group	**Total no. of 3'UTR sequences**	**No. of 3'UTR sequences ≥ 30 nucleotides**
Cell only	1407	1330
EV only plus EV enriched	998	979

Table 3.4: Number of mRNA transcripts with a 3'UTR sequence available and those 3'UTR sequences greater than 30 nucleotides long for the 'Cell only' and 'EV only plus EV enriched' groups.

For each motif, its prevalence within the 'EV+' and 'Cell only' mRNAs was compared. Out of the thirty motifs identified by MEME, only six were highly prevalent in the 'EV+' mRNAs upon the FIMO search and sequence logos for these motifs are shown in **Figure 3.11**. Results displayed in **Figure 3.12** showed that MEME-7 (motif C) had the highest prevalence within 'EV+' transcripts (50.90%), followed by MEME-4 (47.7%, motif A), MEME-3 (39.08%, motif D), MEME-1 (35.27%, motif E), MEME-2 (18.54%, motif B) and finally, MEME-21 (9.12%, motif F). A small percentage of 'Cell only' transcripts presented these motifs; only 6.18% (motifs A and C), 6.11% (motif D), 5.12% (motif E) and 2.35% (motif B). Interestingly, motif F was completely absent from 'Cell only' RNAs. To ensure that prevalence levels were not biased due to specific genes encoding for many transcripts which presented similar motif patterns, these percentage prevalence values were also determined at the gene level, where similar patterns were observed (**Supplementary Table 7.10**). In addition, to ensure abundance values were more broadly representative, all mRNAs at least 2-fold altered in EVs relative to cells (regardless of gene grouping) were assessed,[29] where once again, similar patterns were observed (**Supplementary Figure 7.12**).

As described in **Section 2.6** and **Figure 2.5**, a gene was carried forward in the pipeline for motif discovery if it met the read count criteria (EV: ≥ 1 read, Cell: ≥ 5 reads) in at least one of the three replicates. To confirm that the motifs identified here were representative of all three replicates, FIMO was successfully used to show the enrichment of these motifs in genes meeting a more stringent criteria, where genes met the read count criteria in all three replicates (data not shown).

[29]by Elizabeth Dellar



(i) Motif A: MEME-4 **(ii) Motif B**: MEME-2

(iii) Motif C: MEME-7 **(iv) Motif D**: MEME-3

(v) Motif E: MEME-1 **(vi) Motif F**: MEME-21

Figure 3.11: Sequence logos for six motifs identified via MEME and FIMO analysis as significantly enriched in 'EV+' transcripts when compared to 'Cell only' transcripts.

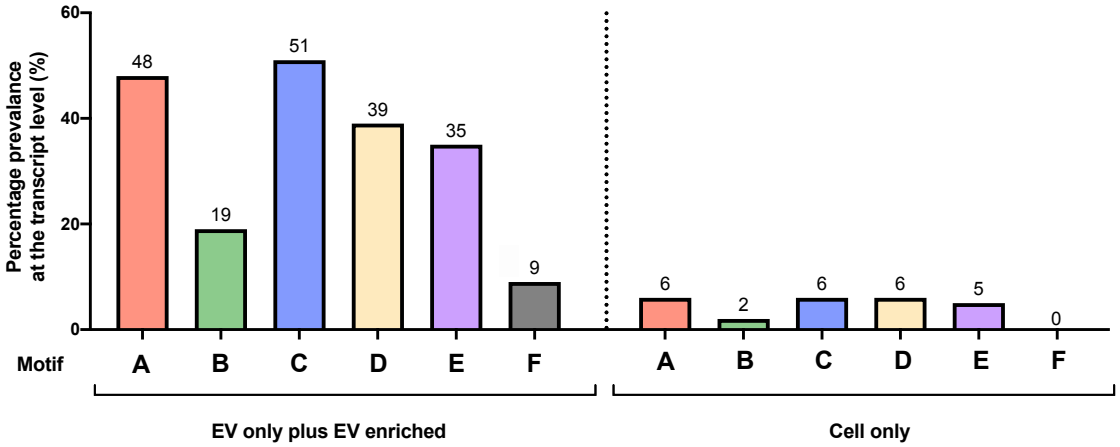


Figure 3.12: Percentage prevalence of motifs within the 3'UTR of 'EV+' and 'Cell only' active protein-coding transcripts. Data visualised from that shown in **Supplementary Table 7.10**.

3.3.2 The six novel *Drosophila* EV-enriched motifs were not analogous to one another

Six *Drosophila* motifs were shown to be enriched in the 3'UTRs of 'EV+' mRNAs when compared to the same region of 'Cell only' mRNAs (**Figure 3.11**). A MEME Suite MAST alignment (described in **Section 2.6.3**) generated a pairwise similarity matrix, where a value of greater than 0.6 highlights motifs with a high degree of sequence similarity. No motif pairs met this cut-off, suggesting that the six novel *Drosophila* motifs (A-F) were not analogous to one another.

			Motif					
		Width	**A**	**B**	**C**	**D**	**E**	**F**
Motif	**A**	21		0.49	0.29	-0.01	0.25	0.28
	B	8	0.49		0.2	0.25	0.22	0.4
	C	20	0.29	0.2		0.02	0.33	0.08
	D	22	-0.01	0.25	0.02		0	0.12
	E	20	0.25	0.22	0.33	0		0.21
	F	23	0.28	0.4	0.08	0.12	0.21	

Table 3.5: Motif similarity matrix generated via a MAST analysis to compare novel *Drosophila* motifs A-F. A value of greater than 0.6 would suggest a high degree of similarity.

3.3.3 Novel *Drosophila* EV-enriched motifs showed no significant sequence similarity with known mammalian EV-enriched motifs

To identify any potential evolutionary conservation of the *Drosophila* EV-enriched motif sequences, they were compared to a number of previously published EV-enriched motifs (**Table 3.6**). This analysis highlighted no significant similarity between the novel *Drosophila* motifs identified in this analysis and those previously shown to be EV-enriched in mammalian systems, including the 25-nt motif and its core region [248]. These results suggested that this bioinformatic pipeline for motif discovery was successful in identifying novel motifs enriched in *Drosophila* EV RNAs.

Within this analysis, comparisons were also made between all of the previously published mammalian motifs (**Table 3.6**). Interestingly, significant similarity was identified between two previously published mammalian motifs, identified by

Temoch-Diaz et al. (2019) and the 25-nt motif identified by Faith-Bolukbasi et al. (2012) (q = 0.0352, **Figure 3.13**) [74, 248], suggesting their presence in a range of mammalian contexts.

Motif sequence	RNA type	Ref.	Associated RNA-binding protein	Associated miRNA binding site
GGAG	miRNA	[234]	hnRNPA2B1	
GGCU	miRNA	[249]	SYNCRIP	
UGGA	miRNA	[74]	Lupus La	
AAUGC	miRNA	[254]	FMR1	
ACCCTG-CCGCCT-GGACTC-CGCCTGT (25nt)	mRNA	[248]		mir-1289
CTGCC/ CTGC/ CTCCC/ CGCCC/ TGCC (core variations)	mRNA	[248]		mir-1289
ACCAGCCU	mRNA	[245, 247]	YB-1	
CAGUGAGC	mRNA	[245, 247]	NSUN2, YB-1	
UAAUCCCA	mRNA	[245, 247]	YB-1	

Table 3.6: Previously published EV-enriched motif sequences from miRNA and mRNA transcripts. Also shown are the associated RNA binding proteins or miRNAs. Adapted from a table generated by Elizabeth Dellar.

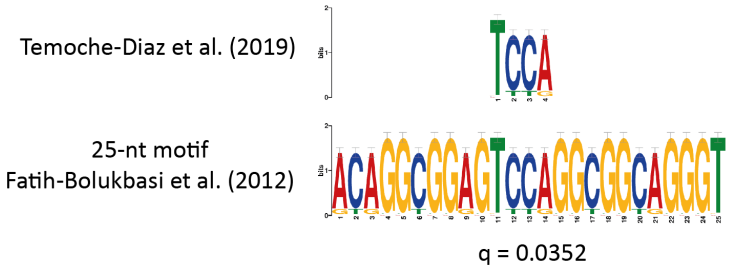


Figure 3.13: TOMTOM alignment of the previously published Temoche-Diaz et al. (2019) and the 25-nucleotide mammalian motif identified by Fatih-Bolukbasi et al. (2012) [74, 248].

3.3.4 Each motif presented a unique distribution pattern along the 3'UTR

Motifs A, B, C and D were selected for downstream analysis based on their high prevalence in 'EV+' transcripts, low prevalence in 'Cell only' transcripts and due to their high degree of sequence conservation, identified via their bit score (**Figure 3.12** and **Supplementary Table 7.10**). To determine if each motif occurred at distinctive sites along a transcript 3'UTR, motif start positions were normalised as described in **Section 2.6.3**. The unique distribution patterns along the 3'UTR for each motif are shown in **Figure 3.14**. Statistical analysis highlighted that significant differences were present between each motif, except when comparing Motif C and D (**Supplementary Table 7.11**). Motifs C and D presented median 3'UTR start positions of 56.02% and 60.69%, respectively. Motifs A and B presented more upstream median 3'UTR positions of 36.61% and 22.19%, respectively. Skewness values also reflected these patterns (**Supplementary Table 7.12**). Hits for motif B were strongly skewed toward the 5' end (0.9893 ± 0.2425, Skew ± Standard Error of Skew, $Z_{g1} = 4.08$), with motif A hits mildly skewed towards this direction (0.4169 ± 0.1431, $Z_{g1} = 2.91$). Hits for motifs C and D were only slightly skewed towards the 3' end (-0.3411 ± 0.1413, Z_{g1} = -2.25 and -0.3201 ± 0.3163, Z_{g1} = -1.01, respectively). This data showed that motifs A, C and D showed a slight tendency to appear in specific locations along the 3'UTR but motif B was preferentially localised towards the beginning of the 3'UTR sequence.

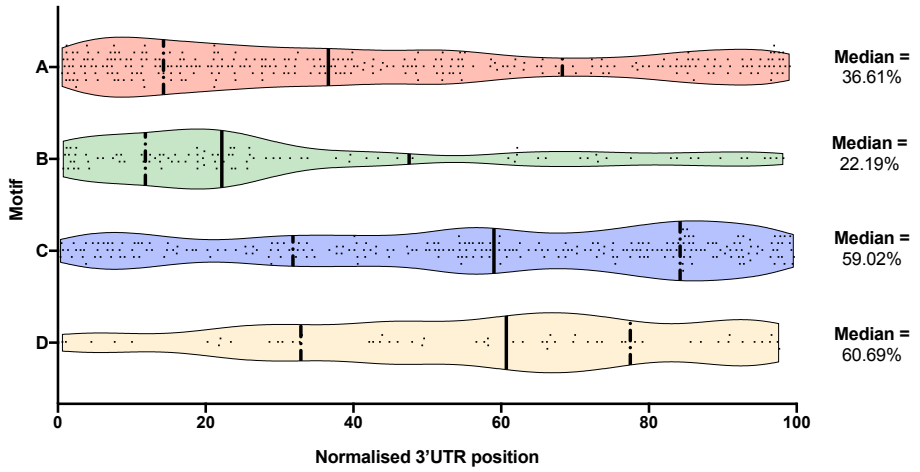


Figure 3.14: Distribution patterns of motif hits along the 3'UTRs of the 'EV+' genes. Solid lines indicate the median value and dashed lines indicate quartiles. Skew analysis and a non-parametric Dunn multiple comparison test was carried out (**Supplementary Tables 7.11 - 7.12**). All motifs, except motifs C and D, displayed significantly different distributions.

3.3.5 Motifs commonly occur in combinations within the 3'UTR

Given that motifs showed preferential distributions along the 3'UTR, it was possible that multiple motifs could be present within the same 3'UTR sequence. Motif combinations within each 'EV+' 3'UTR were identified as described in **Section 2.6.3**. Results showed that the A, B, C and D motifs commonly occurred within the same 'EV+' 3'UTR in the AC (9%, normalised transcript percentage prevalence), ABC (4.7%), ACD (4.7%), CD (4.3%) and ABCD (3.5%) combinations (**Figure 3.15**). Motifs A, C and D did occur on their own within the 'EV+' 3'UTRs (3.9%, 11% and 4.9%, respectively); however, motif B did not. Combinations containing motif B had a low prevalence overall (ranging from only 0.09% to 0.8%), except for when B was present in the ABC (4.7%, as stated) or ABCD (3.5%, as stated) combination.

A small proportion of 'Cell only' 3'UTRs contained the A, B, C and D motifs individually (4.3%, 1.2%, 5.5% and 5.7%, respectively); however, the prevalence of motif combinations in 'Cell only' 3'UTRs was low. The most common 'Cell only' motif combination was AC (1.9%). All other 'Cell only' combinations were found at levels below this, with motif combinations BD and BCD fully absent.

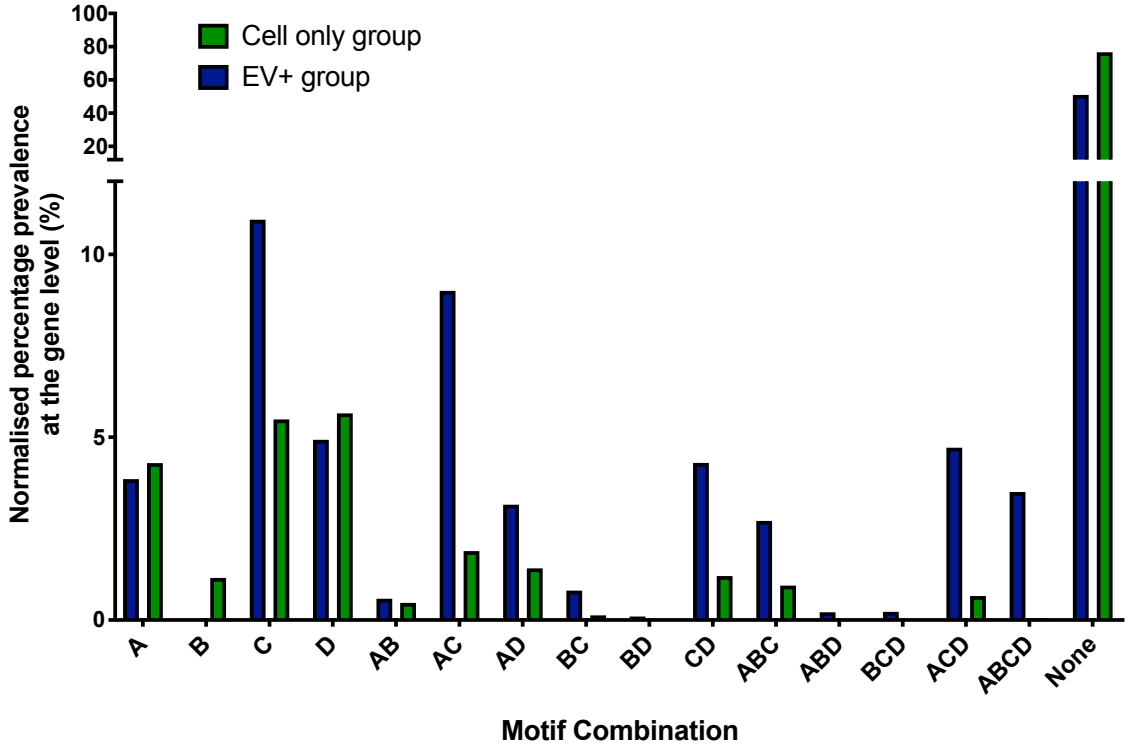


Figure 3.15: Normalised percentage prevalence at the gene level of A, B, C and D motif combinations identified within the 3'UTRs of the transcripts encoded for by 'EV+' and 'Cell only' active protein-coding genes. Prevalence levels were normalised to account for a greater number of transcripts coded for by the 'Cell only' genes compared to the 'EV+' genes.

The prevalence of each motif combination was also determined within two sets of shuffled sequences, to act as a control (the 'EV+' and 'Cell only' 3'UTRs were shuffled as described in **Section 2.6.3**). All four motifs were absent in the shuffled versions of the 'EV+' sequences. Motif A showed a small prevalence within the 'Cell only' shuffled sequences (0.13%); however, all other motif combinations were absent.

Taken together, these results suggested that EV-enriched motifs preferentially occur in specific combinations within the same 3'UTR. These results also suggested

the presence of 'booster' motifs (motifs A, C or D), which were found at high prevalence levels within 'EV+' RNAs and the presence of a 'cell-retention' motif (motif B), found at low prevalence levels within the 'EV+' RNAs, except when present in the ABC or ABCD combination.

3.3.6 3'UTR motifs were specifically arranged within EV mRNAs

Due to the enrichment of the ABCD motif combination within 'EV+' transcripts, future work focused on understanding how these motifs may function together to influence RNA EV incorporation. As a first step, the distribution of motifs along the 3'UTRs of transcripts carrying the ABCD motif combination was visualised. Start and end positions of motif occurrences were normalised to the total 3'UTR length (as described in **Section 2.6.3**) and plotted as shown **Supplementary Figure 7.13**. It was apparent from this analysis that multiple transcripts with the same parental gene displayed similar motif patterns in their 3'UTR and therefore, a shortlist of transcripts was generated by eliminating any duplicated 3'UTR motif patterns observed per gene (**Supplementary Figure 7.14**).

Using this shortlist of transcripts which carry the A, B, C and D 3'UTR motifs, the number of transcripts presenting each permutation of the ABCD combination was quantified, to determine if there was any preferential order to motif occurrences along the 3'UTR (**Table 3.7**, as described in **Section 2.6.3**).[30] Results showed that motifs A and B preferentially occurred early in the motif combination, with 100% of transcripts containing a combination with motif A or B in the first or second position. The most common motif order was BACD, present in 59.52% of transcripts analysed. This was closely followed by the ABCD (52.86%), BADC (47.62%), ABDC (also 47.62%) and DBAC (45.24%) orders. Note that due to the possibility of multiple motif occurrences within the same 3'UTR (**Figure 3.16**), it was possible for one transcript to contain multiple motif combinations and therefore, percentages for all motif combinations add up to a value greater than 100%. These results suggested

[30]by Jana Terenova

that the distribution of the novel *Drosophila* motifs along the 3'UTR was not random; restraints may exist to retain particular distribution patterns.

Order of combination	Total transcript count	Unique transcript count	Unique transcript percentage (%)
B, A, C, D	3858	25	59.52
A, B, C, D	2465	23	54.76
B, A, D, C	2431	23	54.76
A, B, D, C	1706	20	47.62
D, B, A, C	802	19	45.24
A, C, B, D	270	18	42.86
B, D, A, C	653	17	40.48
B, C, D, A	669	16	38.10
B, D, C, A	518	15	35.71
C, B, A, D	118	15	35.71
C, B, D, A	97	15	35.71
C, A, B, D	50	14	33.33
A, D, B, C	181	13	30.95
B, C, A, D	398	13	30.95
D, B, C, A	708	11	26.19
C, A, D, B	41	9	21.43
C, D, B, A	67	8	19.05
A, C, D, B	59	6	14.29
D, A, B, C	58	6	14.29
D, C, B, A	27	6	14.29
A, D, C, B	58	5	11.90
D, A, C, B	24	4	9.52
C, D, A, B	35	3	7.14
D, C, A, B	13	3	7.14

Table 3.7: Order of motif occurrences present in the shortlisted ABCD 'EV+' transcripts. Shown here are the total number of occurrences of each permutation of the ABCD combination. Also shown are the number of unique transcripts with that motif order within their 3'UTR. This number is also represented as a percentage of the total number of transcripts analysed (N=42).[31]

[31]Data generated by Jana Terenova

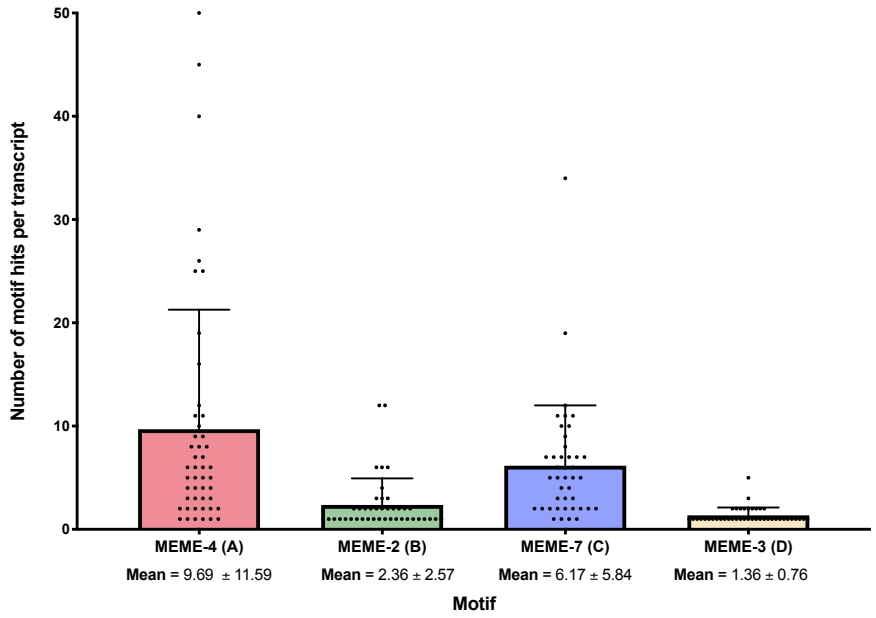


Figure 3.16: Number of A, B, C and D motif 3'UTR hits per shortlisted 'EV+' transcript. Results showed that individual motifs were repeated within the same 3'UTR.

3.3.7 RNA motifs were part of specific RNA secondary structures

The presence of preferential motif combinations and orders within the 3'UTR suggested that motifs may have at tendency to form specific secondary structures. Along these lines, previous studies have suggested, via a visual analysis of a few examples, that EV-enriched motifs were commonly located within stem-loop secondary structures [245, 246, 248]. A global secondary structure analysis of the entire 3'UTR region of shortlisted transcripts carrying the ABCD motif combination was carried out using RNAfold and the percentage of paired bases within each motif occurrence was quantified (as described in **Section 2.6.3**). Results displayed in **Figure 3.17** showed that, on average, a higher percentage of bases within motifs B and D were paired (65.25 ± 22.29% and 76.19 ± 17.05%, respectively). Motifs A and C, however, showed no clear prevalence to appear in paired versus unpaired, looped regions. A significant difference was observed when comparing all motifs, except for motif B and D (p-adj = 0.0805) (**Supplementary Table**

7.13). These results suggested that the 3'UTR motifs presented an individual tendency to occur within unpaired, looped regions.

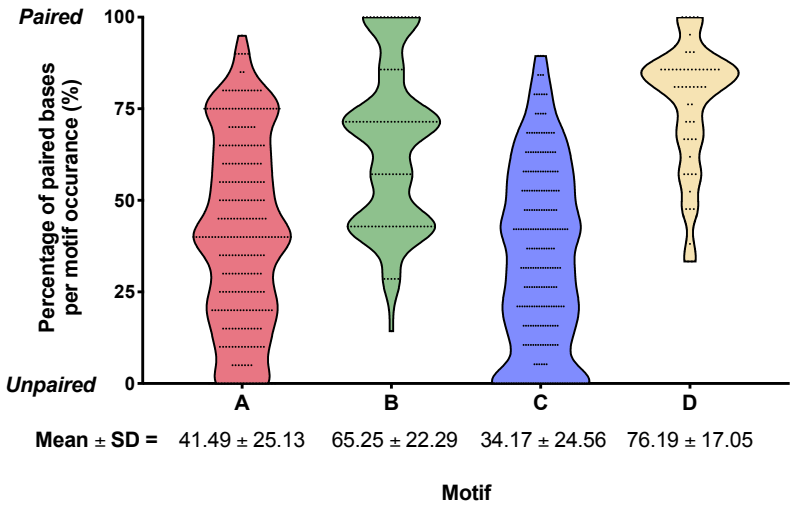


Figure 3.17: Percentage of bases paired per motif occurrence within the shortlisted ABCD 'EV+' transcript 3'UTRs. Mean ± standard deviation is displayed. Dunn's multiple comparisons statistical analysis is shown in **Supplementary Table 7.13**. Statistically significant differences were observed between each motif, except motif B and D.

3.3.8 RNA motifs potentially interact with specific RNA-binding proteins

The quantitative structural analysis suggested that each motif presented an individual likelihood to occur within unpaired secondary structures. Interestingly, interactions have been shown to occur between such RNA loop regions and RNA-binding proteins (RBPs), which in turn promote specific mRNA localisation [247, 249, 430, 431]. To determine if any of the novel *Drosophila* EV-enriched motifs showed similarity with known *Drosophila* RPB binding sites [420], a sequence similarity analysis was carried out as described in **Section 2.6.3**. This analysis highlighted that the A, C and D *Drosophila* motif sequences showed sequence similarity to the sequences of known RBP binding sites (**Figure 3.18**). Interestingly, motif B did not show sequence similarity to any of the established binding sites for *Drosophila* RBPs analysed. CNOT4 (CCR4-NOT transcription complex subunit 4, FBgn0051716) and SM (Smooth, FBgn0003435) showed a highly significant

sequence similarity with motif C ($p < 0001$, $q < 0.02$, **Figure 3.19**), highlighting a predicted interaction between motif C and these RBPs. The FMR1, SXL and RBP9 hits, along with the remaining hits depicted in **Figure 3.18**, did show significance with regards to p-values. Q-values for these hits (which account for false discovery rates due to multiple sampling, as appropriate here), however, were not significant and therefore, this should be considered a preliminary but potentially meaningful analysis that demands further biochemical confirmation. Interestingly, human orthologs of all predicted motif interacting RBPs, except for Orb2, have been identified in mRNA or protein form within the EVs released by various human or mouse cell lines (described in **Supplementary Table 7.14**).

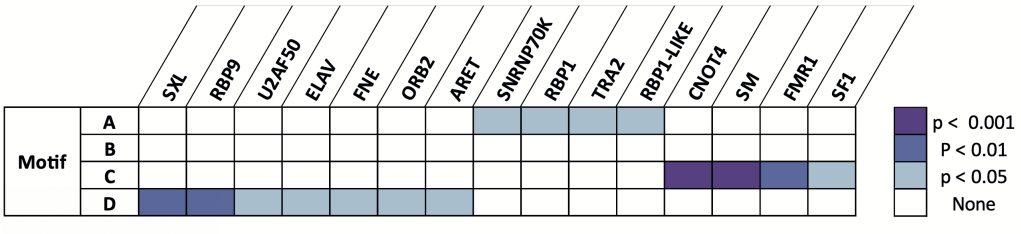


Figure 3.18: Sequence similarity hits observed between the novel *Drosophila* motifs and a database of known RBP binding sites. Colour intensity of the box depicts the significance of the sequence similarity (p-value).

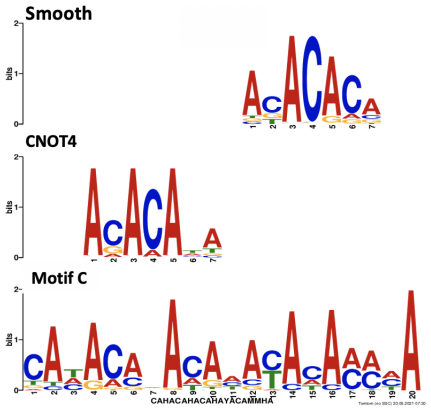


Figure 3.19: Sequence alignments for the hits observed between the novel *Drosophila* motif C and the known RNA-binding sites for Smooth and CNOT4. This sequence similarity was highly significant ($q < 0.02$), highlighting a potential interaction between motif C and these RBPs.

3.3.9 Novel *Drosophila* motifs were more prevalent in longer transcripts

When assessing transcript length profiles for the top 1000 most abundant mRNAs in cells versus EVs, no distinct differences were observed (**Figure 3.20**).[32] This result suggested that there may have been no preferential loading of shorter (< 1500 nts) RNAs into EVs, as has been previously observed [253]. The profile identified in this analysis was in line with Wei et al. (2017), who showed that cellular and EV mRNA length distributions were highly similar [215].

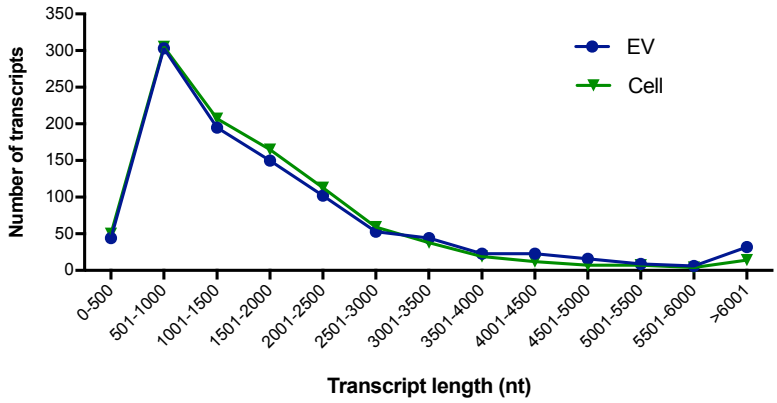


Figure 3.20: Length distribution of the median transcript length of the top 1000 most abundant mRNAs (excluding *Arc1*) detected in EV and cell samples.[33]

Following this, an assessment of transcript length (and 3'UTR length) for transcripts with and without a 3'UTR occurrence of any A, B, C or D motif combination was made. For this analysis, mean transcript lengths were determined for each *Drosophila* gene. As shown in **Figure 3.21**, genes which contained any 3'UTR combination of the A, B, C or D motifs had a significantly higher mean transcript 3'UTR length (739.1 ± 467.9 nts) compared to genes which lacked any motif occurrences (183.4 ± 123.6 nts, $p < 0.0001$). Similar results were observed when analysing total transcript lengths (**Figure 3.22**). These results suggested that the presence of the EV-enriched motifs may encourage the EV-loading of longer transcripts. It is important to note, however, that purely due to the longer

[32] by Elizabeth Dellar
[33] Data generated by Elizabeth Dellar

length of theses 3'UTR regions, the presence of an EV-enriched motif may be more likely. Future work will be required to biologically validate if these novel *Drosophila* motifs can enhance EV-loading of longer transcripts.

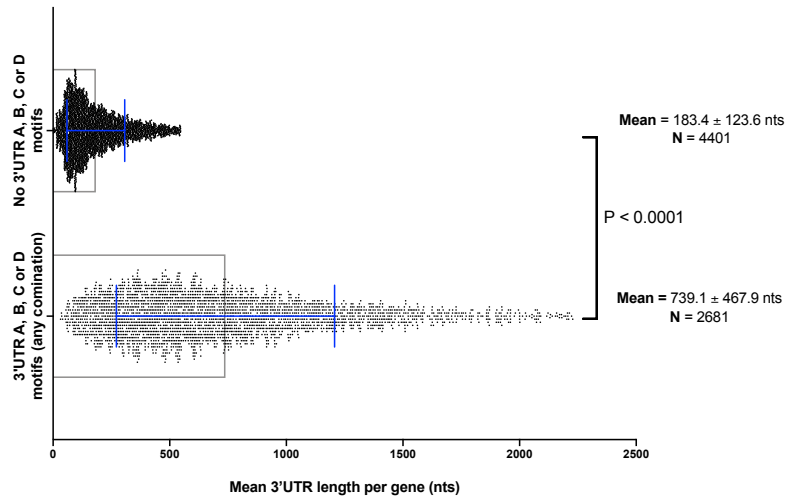


Figure 3.21: Mean 3'UTR length per gene for genes encoding transcripts with / without any combination of the A, B, C or D motifs. Statistical analysis was carried out via a two-tailed Mann-Whitney test. Outliers were identified via Graphpad (ROUT = 1%, ABCD outliers = 100, No ABCD outliers = 333). Mean ± standard deviation is shown.

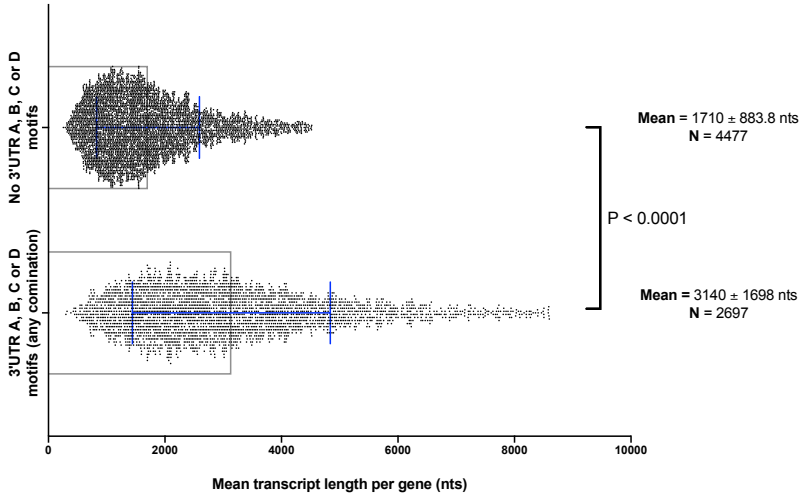


Figure 3.22: Mean transcript length per gene for genes encoding transcripts with / without any combination of the A, B, C or D motifs. Statistical analysis via a two-tailed Mann-Whitney test. Outliers were identified via Graphpad (ROUT = 1%, ABCD outliers = 84, No ABCD outliers = 257). Mean ± standard deviation is shown.

3.3.10 EV-enriched *Drosophila* motifs present in transcripts with specific gene product attribute predictions

A gene ontology (GO) analysis was carried out, as described in **Section 2.6.3**, to highlight predicted biological processes, molecular functions and cellular components related to all genes encoding transcripts carrying the 3'UTR ABCD motif combinations, regardless of their abundance within cellular or EV RNA samples (**Figure 3.23**, **Supplementary Tables 7.15 - 7.17**). This analysis highlighted that the most common predicted biological processes were associated with development. Also enriched were terms related to the regulation of signalling, as well as vesicle transport and secretion. Protein binding was the most prevalent molecular function, related to SNARE, clathrin and receptor binding; processes associated with EV release and uptake [119, 206]. The most prevalent cellular components were associated with intracellular vesicles and intracellular organelles, with these groups containing terms related to cytoskeletal functions. Taken together, these results suggested that genes encoding transcripts containing 3'UTR motif occurrences may predominantly relate to vesicle formation, function and transport, which may explain the motif EV-enrichment and potential enhancement of EV-loading. The predicted localisation of these mRNAs to particular membrane domains, and their potential relationship with cytoskeletal and actin-filament based processes, suggested they may be involved in the localised translation of proteins for membrane remodelling, perhaps for generating cell projections or modulating vesicle release [432–434]. Furthermore, the release of these mRNAs or proteins in EVs may regulate cellular development and the activity of molecular pathways in distant cells.

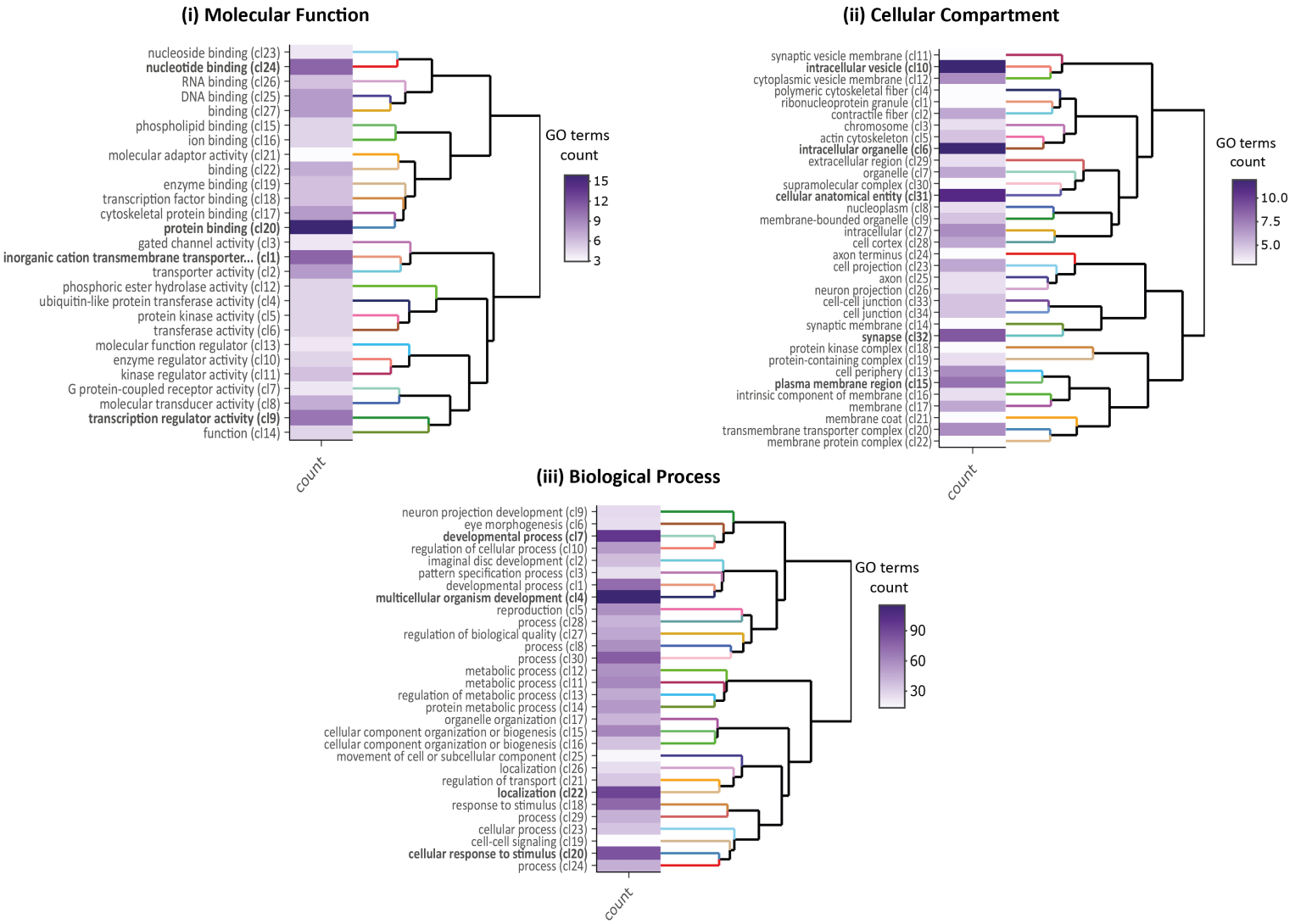


Figure 3.23: Significantly enriched (i) molecular function, (ii) cellular component and (iii) biological process gene ontology (GO) terms associated with genes encoding transcripts carrying any combination of the 3'UTR A, B, C and D EV motifs. GO terms were clustered and summarised using heatmaps. Incomplete text in figure reads: 'inorganic cation transmembrane transporter activity.' Cl = cluster. Bolded text indicates highly prevalent GO term clusters.

3.3.11 Novel motif occurrences found within mRNA coding regions and the 5'UTR

To date, many studies have focused on the discovery of 3'UTR motifs within EV-loaded transcripts, due to the role 3'UTRs play in mRNA localisation, regulation and stability. 5'UTRs, however, have also been shown to play roles in these processes, with RNA binding proteins also binding to these mRNA regions [435–437]. With this in mind, the prevalence of the 3'UTR motifs within the 5'UTR and coding regions was also determined. As before, prevalence levels were determined at the transcript and gene level, to ensure results were not distorted due to the presence of multiple transcripts from the same parental gene showing similar motif patterns (**Supplementary Table 7.18**).

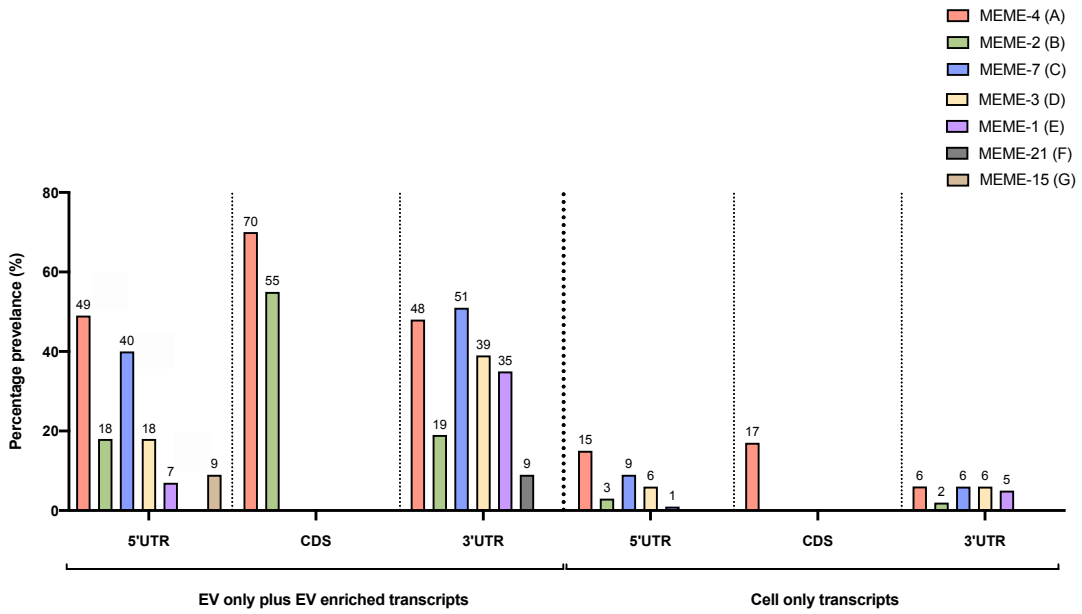


Figure 3.24: The percentage prevalence of 3'UTR, 5'UTR and CDS *Drosophila* motif occurrences determined within the 'EV+' and 'Cell only' active mRNA transcripts. Mean values are displayed above the bars. Data shown in **Supplementary Table 7.18**.

As shown previously, all motifs present in the 3'UTR were more prevalent in the 'EV+' transcripts, compared to the same region of 'Cell only' transcripts (**Figure 3.24** and **Supplementary Table 7.18**). Similarly, motif prevalence levels within the 5'UTR were higher in 'EV+' transcripts, compared to the 'Cell only' transcripts. Focusing on the UTRs of 'EV+' transcripts alone, results showed that motifs C,

D, E and F were more prevalent in the 3'UTR versus the 5'UTR, with similar prevalence levels of motifs A and B observed between these regions. Interestingly, motif F was unique to 'EV+' transcript 3'UTRs (9.12%). Conversely, motif G, which was previously excluded from analysis due to its absence in transcript 3'UTRs, was unique to 'EV+' transcript 5'UTRs (8.8%).

Comparing the prevalence of motifs within the 3'UTRs and CDS domains of 'EV+' transcripts alone, results showed that motifs A and B were more prevalent in the CDS region. All other motifs were either absent (motifs F and G) or present at low prevalence levels (motif C = 0.3%, motif D = 0.1%, motif E = 0.2%) in the CDS of these EV transcripts. The prevalence of motifs A and B in the CDS of 'EV+' transcripts (69.69% and 55.43%, respectively) were higher than levels observed in the CDS of 'Cell only' transcripts (17% and 0% respectively).

It has been shown that 3'UTRs are, on average, longer than 5'UTRs in many species, with 3'UTRs also displaying greater length variability [438]. Comparisons of 5'UTR, 3'UTR and CDS lengths were made between the 'EV+' and 'Cell only' transcripts (**Figure 3.25**). No significant differences were observed in the 5'UTR, 3'UTR or CDS lengths of 'EV+' transcripts, suggesting that the differences in motif prevalence levels between these regions were likely not disturbed due to differences in sequence length.

Significant differences were observed between 5'UTR, 3'UTR and CDS lengths in the 'Cell only' transcripts. 'Cell only' 3'UTRs were significantly longer than 'Cell only' 5'UTRs, with these 3'UTRs also displaying greater length variability, following the patterns previously observed [438]. Importantly, the 5'UTR, 3'UTR and CDS sequences of 'Cell only' transcripts were significantly shorter than the same regions in 'EV+' transcripts. This result suggested that increased length of 'EV+' transcripts may enhance the likelihood of motifs being present, as observed in **Section 3.3.9**. Tushev et al. (2018) suggest that longer sequences may be more 'regulatory competitive' due to an increased likelihood to contain *cis*-acting factors [439]. This is a feature which could potentially encourage mRNA EV-loading.

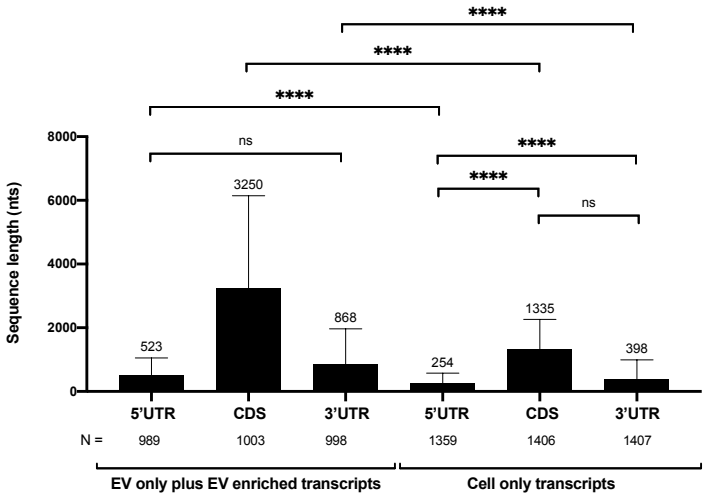


Figure 3.25: Sequence length of 3'UTR, 5'UTR and CDS *Drosophila* regions determined for the 'EV+' and 'Cell only' active mRNA transcripts. Mean + standard deviation is displayed, with mean values shown above the error bars. Statistical analysis was carried out via a Dunnett's multiple comparisons test. **** indicates $p < 0.0001$.

In summary, this analysis highlighted that for 'EV+' transcripts, all motifs were more prevalent within UTRs, except for motifs A and B which were more prevalent in the CDS region. Of the motifs prevalent in both the 3' and 5' UTR, all were significantly more prevalent in 3'UTR, except motifs A and B which appeared with equal frequencies in both regions. Intriguingly, motifs F and G which were largely excluded from downstream analyses due to their low prevalence within the 3'UTR, were in fact unique to 3'UTR and 5'UTR regions of 'EV+' transcripts, respectively. In the future, it would be interesting to specifically investigate the influence of these motifs on mRNA EV-loading.

3.4 Discussion

3.4.1 Characterisation of *Drosophila* S2R+ cell and EV RNAs

This chapter described the bioinformatic analysis of *Drosophila* S2R+ cell and EV RNAs. This analysis was carried out to identify EV-enriched RNA features which may encourage mRNA incorporation into EVs. An improved understanding of these fundamental processes may highlight novel targets for modulating cell communication systems for therapeutic purposes.

RNA-sequencing was carried out on RNA isolated from *Drosophila* S2R+ EVs and their parental cells. All A260/280 ratios showed an acceptable level of RNA purity, widely accepted as ∼1.8 to 2.1 [440, 441]. Cellular RNA A260/230 ratios were ∼2, indicative of pure RNA. EV RNA A260/230 values were more variable; however, a consensus value for the lower limit of this ratio has not been firmly established [440]. A low A260/230 ratio can reflect sample contamination with EDTA, carbohydrates or phenols; however, it can also be influenced by a low starting sample amount in the RNA extraction procedure [440, 442]. Indeed, for EV replicate two, where a poor A260/230 ratio was observed, the RNA concentration was low (8.25 ng/ul, Qubit).

Possible fragmentation of EV-loaded RNAs

Bioanalyser profiles for cell and EV RNA samples showed that small RNAs (up to 500 nucleotides) were the most abundant RNA species, with ribosomal RNA peaks largely absent. Ribosomal RNA is believed to be depleted from EVs [443], with other studies using mammalian models [68, 70, 264] and *Drosophila* S2R+ cells [366] reporting a lack of large 18s and 28s rRNA Bioanalyser peaks. A high proportion of RNA-sequencing reads, however, aligned to 18s and 28s rRNA regions, an observation also made by Lefebvre et al. (2016) when studying *Drosophila* S2R+ EV RNA [366]. Lefebvre et al. (2016) proposed that EV-loaded rRNAs may be fragmented, explaining the diminishment of the Bioanalyser rRNA peak but the presence of rRNA sequencing reads. The presence of EV loaded rRNA (as well as tRNAs, vault-RNAs or other non-coding RNAs) has been reported in

Drosophila, mammalian, parasite and urine derived EVs [210, 211, 223, 230, 231, 366, 444–446]. These fragments have the potential to be involved in translational inhibition, guidance of mRNA cleavage or promotion of morphological changes in recipient cells [447, 448].

Focusing on all non trans-spliced or miRNA genes, cell and EV RNA reads predominantly aligned to coding regions (approximately 69 - 74%), with the second most prevalent region being the 3'UTR (approximately 11 - 12.5%). Interestingly, a significant decrease in the proportion of reads aligning to coding regions was observed in EV RNA when compared to cellular RNA, with the proportion of reads aligning to UTR and intron regions tending to increase in EVs. These results suggested that in addition to rRNA fragments, UTR or intron RNA fragments may too be loaded into *Drosophila* EVs. The presence of mRNA fragments has been reported [210, 449, 450], with studies showing that mRNA fragments within EVs were predominantly derived from the 3' end of the transcript [210, 215]. Post-transcriptional cleavage of mRNAs has been proposed as a mechanism to generate 3'UTR fragments enriched in regulatory elements [451]. These fragments may act independently to carry out non-coding functions across multiple species [452]; providing competitive binding sites, affecting the stability of cellular mRNAs or altering gene expression by interfering with the RNA-induced silencing complex (RISC) [210, 215].

Alternatively, a tendency towards a higher proportion of UTR or intronic reads within EVs may be the result of un-spliced RNA EV-loading, perhaps via an overlap between RNA processing and RNA EV-loading. This is potentially evidenced by the EV-loading of the RNA processing proteins, Ago and Dicer, in specific contexts [334]. In addition, changes in polyadenylation and deadenylation patterns between cell and EV RNAs may also alter the balance of reads aligning to coding regions. It is known that polyadenylation and deadenylation can alter mRNA turnover and stability [453–455], both of which can potentially influence mRNA EV-loading. Read coverage comparisons at the transcript level are required to uncover if full-length, fragmented or un-spliced transcripts become EV loaded; however, further RNA-sequencing experiments are required to attain the appropriate read

depth for transcript-level resolution [456]. This will be an interesting avenue of future research, with the potential to highlight novel mechanisms by which cells regulate gene expression at a distance.

Drosophila S2R+ cells and EVs carry distinct RNA cargo, with EV RNA cargo more variable than that of parental cells

A Principal Component Analysis (PCA) of cell and EV RNA normalised read counts highlighted distinct EV and cell RNA profiles. The close clustering and strong Spearman correlation of cellular RNA replicates suggested a high degree of consistency in cellular RNA content. EV RNA replicates, however, displayed a higher degree of variation. Similar observations have been reported in mammalian RNA-sequencing datasets [216, 224]. This result suggested that EV RNA cargo is distinct from the RNA profile of the cells of origin, with EV RNA content highly dynamic, likely changing as a result of biological or environmental conditions.

Cellular and EV RNA display distinct transcript profiles

Although the majority of active RNAs were common to both S2R+ cell and EV samples, some differences were apparent. Cells and EVs appeared to have distinct ncRNA profile, with few ncRNAs shared between both fractions. EV-loading of ncRNAs, such as miRNAs, has been related to cellular concentrations [225]; however, the presence of EV-specific ncRNAs suggested that active loading mechanisms may also be involved.

Drosophila cells and EVs also displayed unique profiles for alternative ncRNAs such as snRNA, snoRNA and pseudogenes. snoRNAs are involved in ribosome biogenesis and function and are mainly localised to the nucleus [457]; however, they are shuttled to the cytoplasm in response to cellular stress [458, 459]. EV-loading of snoRNAs (or fragments thereof) is affected by LC3-dependent EV-loading and secretion (LDELS), a discrete secretory autophagy pathway [177]. EV-mediated snoRNA release has been shown in the context of inflammation and was proposed as a mechanism by which homeostatic processes are regulated at a distance [178]. EV-loaded pseudogenes may be an additional mechanism by which gene expression

is altered in recipient cells, by providing competitive binding sites for regulatory miRNAs [460]. In the future, it would be interesting to further explore the EV-specific loading of these alternative non-coding RNAs; determining the mechanism of loading, the influence of cellular stress and their effect on recipient cells.

The most common RNA type in cellular and EV RNA samples was mRNA. This was in line with Lefebvre et al. (2016) who showed that mRNAs accounted for 74.5% of the top 1,000 most abundant transcripts present within *Drosophila* S2R+ EVs [366]. Both EV-specific and EV-enriched mRNAs were identified, providing further evidence to suggest that *Drosophila* EVs contained distinct mRNA cargo which did not fully reflect the mRNA profile of parental cells. Much less is known about the process of mRNA EV incorporation, with many studies previously focusing on ncRNA cargo, particularly miRNAs. Therefore, downstream analysis focused on uncovering mRNA features which may facilitate their EV-loading and long-range transfer.

3.4.2 Mechanisms of selective EV-loading in *Drosophila melanogaster*

The extent to which active and regulated process are involved in EV RNA incorporation remains a highly debated topic, with some studies suggesting that EV RNA largely reflects the RNA profile of donor cells [225] and others highlighting the non-random loading of particular RNAs [216, 226–228]. More recently, it has been suggested that both non-selective and selective mechanisms for RNA EV-loading are present [78], with RNAs loaded via either mechanism having the potential to be functionally relevant [229]. Gene ontology analysis in this study highlighted predicted cellular processes which EV-specific and EV-enriched ('EV+') mRNAs may be involved in, such as development, pattern specification and regulation. Some suggest that the EV-enrichment of specific mRNAs with defined functions may in itself suggest the existence of an active sorting mechanism [68, 461].

Proposed mechanisms of RNA EV-loading were previously introduced in **Section 1.2.4**. In summary, specific EV-enriched RNA motifs can modify RNA secondary

structures [234, 245–249, 462] to encourage lipid-raft or RNA binding protein interactions, subsequently influencing RNA EV-loading [237]. 3' and 5' regions of mRNAs are 'hotspots' for post transcriptional modifications (PTMs) and for *cis*-acting factors such as sequence motifs, which play roles in recruiting *trans*-acting factors such as RBPs or miRNAs [233, 463]. For this reason, motif searches were carried out in the *Drosophila* 3'UTR of EV-enriched transcripts to identify potential features which regulate their EV incorporation.

EV-enriched motifs present within the 3'UTR of *Drosophila* EV mRNAs

MEME Suite analysis uncovered six novel motifs which were enriched in the 3'UTRs of *Drosophila* EVs relative to the cell. Four motifs, A, B, C and D, were selected for downstream analysis due to their high abundance in EV transcripts, high degree of sequence conservation between transcripts and their diminishment in cells. The distribution pattern of these four novel *Drosophila* motifs along the 3'UTR of the 'EV+' transcripts revealed that each motif had a preferential location. Motif B showed a strong preference to occur upstream in the 3'UTR, motif A showed a slight preference to occur in these upstream regions, whilst motifs C and D occurred throughout the 3'UTR. This result introduced the possibility that multiple motifs may be present within a single 3'UTR sequence.

Novel *Drosophila* motifs preferentially occur in combinations along the 3'UTR

Along these lines, combinations of the motifs within the 3'UTR were identified, highlighting that motifs frequently occurred together, with the most commonly observed combinations being AC, ACD, CD and ABCD. This result reflected mammalian cell culture studies which showed that EV-enriched motif combinations were preferred over single or double motif occurrences [245]. The combinatorial nature of EV-loading motifs appears reminiscent of regulatory element organisation along DNA to recruit transcription factors [245, 464]. In addition to occurring in combinations, results showed that motifs were present multiple times per 3'UTR, for example, motif A was observed, on average, approximately ten times within a

single 3'UTR sequence. This was once again in line with mammalian studies which proposed that short motifs may have weak localisation functions when acting alone, but multiple tandem copies may function together to facilitate finer control [245].

Interestingly, motif B never occurred alone in 'EV+' 3'UTRs and was only observed in the presence of other motifs. These results suggested the presence of EV-enriched (e.g. AC, ACD, CD or ABCD) and EV-depleted (e.g. B, BD, BCD or ABD) motif combinations. It is possible that in the same way that EV-enriched motifs may encourage EV-loading of RNAs, EV-depleted motifs may encourage cellular retention. Cell-retention motifs have previously been reported in a mammalian model by Villarroya-Beltri et al. (2013), who studied miRNA EV-loading [234]. In the future, it will be interesting to determine the biological influence of EV-enriched and EV-depleted motifs, as well as the importance of their combinatorial nature. Investigations to explore these areas will be a focus of **Chapter 5**.

Novel *Drosophila* motifs occur in characteristic secondary structures

The presence of preferential motif orders and combinations within the 3'UTR suggested that these motifs may contribute to a particular secondary structure. Previous analyses have studied the secondary structure of short motif carrying regions via a visual analysis, often showing paired stem regions containing loops or bulges [245–248, 462]. The development of a quantitative approach to analyse RNAfold predictions of the entire 3'UTR secondary structure allowed additional factors such as context of the motif occurrence, spacing and location to be accounted for. These results showed that each motif had an individual preference to occur within paired versus unpaired regions, with motifs B and D predominantly occurring within paired regions. These results suggested that the primary sequence of motifs, in the wider context of the entire 3'UTR, were important to define RNA secondary structure; mirroring the results obtained in mammalian models. The exact mechanisms by which RNA sequence and structure influence EV-loading remains poorly understood and will be an important aspect of future investigations.

Novel *Drosophila* motifs are predicted to bind to a unique repertoire of RNA-binding proteins

RNA motifs and the resultant secondary structure may influence RNA EV-loading via their interaction with RBPs [175, 234, 245, 247, 249, 462]. The *Drosophila* EV-enriched motifs A, C and D displayed similarity with the binding sites of a unique repertoire of RBPs, suggesting potential RNA-RBP interactions. Significant similarity was detected between motif C and the binding site for CNOT4 (CCR4-NOT transcription complex, subunit 4). Beyond the role of the CCR4-NOT complex as a positive regulator of the Jak/Stat pathway [465, 466], it is believed to play a role in protein ubiquitination [467, 468] and mRNA deadenylation [453, 469]. CNOT4, therefore, may be involved in mRNA binding via motif C, potentially influencing mRNA modifications to regulate RNA stability and EV-loading.

Significant similarity was also detected between motif C and the binding site for SM. The ExoCarta database [173–175] showed that hnRNPL protein (heterogeneous nuclear ribonucleoprotein L), the closest human ortholog of SM, is present within EVs released by various cell types [179, 470, 471]. Indeed, human orthologs of all predicted motif interacting RBPs, except for Orb2, have been identified in either mRNA or protein form within mammalian EVs (**Supplementary Table 7.14**). hnRNPL has been shown to regulate mRNA translation [472–475]. Encouragingly, hnRNPL is also likely to bind a mammalian EV-enriched motif [476], suggesting possible conservation of this function between species. Motif C, therefore, may bind SM to alter mRNA translation, modulating mRNA turnover and subsequently influencing mRNA EV incorporation. These studies highlight the interplay between mRNA, RBPs and miRNAs in RNA regulation.

The MEME Suite tool, Tomtom, was harnessed for this analysis, where each EV-enriched motif was compared with the *Drosophila* RBP binding-site database. The most similar pairs were reported, along with estimates of the statistical significance of each match [414, 420, 421]. These are predicted interactions and experiments described in **Section 3.5** aim to biochemically characterise the RBPs interacting

with the *Drosophila* motifs, to biologically validate their function and potentially identify novel RBPs involved in RNA EV-loading.

Interestingly, motif B did not show sequence similarity with any of the *Drosophila* RBP binding-sites investigated. Motif B may play an alternative role by altering RNA stability or mediating miRNA interactions. Motif B may also interact with a currently unidentified RBP, as shown for mammalian motifs which were distinct from any known RNA elements [245]. Alternatively, this motif may bind miRNAs, alternative regulators of EV-RNA loading [248]. Experiments described in **Section 3.5** aim to identify miRNA interactions with the novel *Drosophila* motifs to potentially uncover an additional layer of control in mRNA EV-loading in this model organism.

Novel *Drosophila* motifs predominantly occurred in untranslated regions

Previous studies have reported that the effect of EV-enriched motifs was strongly dependent on the motif location [245, 247, 462]. Furthermore, RBPs have been shown to bind at both 3' and 5' UTRs to modulate translational repression [477–479]. Therefore, the prevalence of motif occurrences throughout the mRNA was assessed. Encouragingly, regardless of transcript location, motifs were more prevalent in the 'EV+' transcripts. Motifs were identified in both the 3' and 5' UTRs, with the majority of motifs more prevalent in the 3'UTR. These results support the hypothesis that the same RNA binding proteins may bind to both the 3' and 5' UTR.

Only motifs A and B were more prevalent in the CDS compared to the 3'UTR in 'EV+' transcripts. This result supported the model proposed by Singh et al. (2015) that 5' and 3' UTRs are more accessible for making regulatory interactions. The CDS is compacted to protect the open reading frame (ORF) and allow the mRNA and protein complex (ribonucleoproteins or mRNPs) to move around the cell [463]. Motifs A and B may play roles beyond *trans*-factor binding, for example, altering RNA stability or defining RNA secondary structure. In accordance with this hypothesis, motif B did not show any sequence similarity with the binding sites of known *Drosophila* RBPs. In the future, biological validation is needed to

assess the effect of motif location on the likelihood that an mRNA becomes EV incorporated. In addition, probing the 5'UTR and coding sequences individually for the presence of alternative EV-enriched motifs may identify additional motifs involved in enhancing mRNA EV-loading.

Novel *Drosophila* motifs may encourage the loading of longer transcripts

Results in this chapter highlighted that, on average, transcripts carrying any combination of the A, B, C or D motifs were approximately two times longer than those lacking any 3'UTR motif occurrences, with these transcripts displaying 3'UTRs approximately four times longer. The mean transcript length for transcripts carrying motif combinations was approximately 3000 nts, much longer than the typical EV RNA length of 200 to 1000 nucleotides [224, 235, 480]. Studies have often focused on small sequencing RNA datasets [213, 481], meaning they do not account for the entire RNA population. This analysis utilised a total RNA sequencing dataset and therefore, a fairer assessment of transcript length could be made. Indeed, this analysis highlighted no distinct differences in the transcript length profiles for the top 1000 most abundant cell and EV genes, in line with results observed by Wei et al. (2017) [215].

Transcripts with long 3'UTRs may be inherently more likely to carry motif sequences, potentially increasing their likelihood to be incorporated into EVs. This in itself may be a potential loading feature, as described by Tushev et al. (2018), who showed that 3'UTR lengths may be a means to alter intra-cellular distribution and effect the 'regulatory competitiveness' of an mRNA isoform [439]. Repeated motif elements within the same 3'UTR may encourage the EV-loading of longer transcripts.

Carrying out an isoform specific RNA-sequencing analysis would allow full-length transcript reconstruction to be carried out, to ensure that RNA fragments alone are not being detected. Furthermore, as different transcript isoforms from the same gene presented different motif combinations, RNA-sequencing at the isoform level may further emphasise the influence of 3'UTR motif combinations on EV RNA abundance. Due to the short read lengths generated during these experimental

processes and low read coverage, however, analysis at the transcript isoform level was not completed as reconstructing full-length transcripts, especially for low abundance transcripts, is challenging and may not yield definitive results [482]. **Section 3.5** describes possible confirmatory experiments to overcome these caveats.

Transcripts containing the 3'UTR EV-enriched motifs showed distinct gene ontology profiles

Genes encoding transcripts carrying 3'UTR motif combinations showed an enrichment in cellular component GO terms related to vesicles, intracellular organelles (such as the endosome) and the plasma membrane highlighting their potential relationship with EV-loading. GO terms related to synapses were also enriched, in concordance with Lefebvre et al. (2016) who suggested that EV-loading of neuronal mRNAs was a mechanism by which cells can reduce transcriptional noise or decrease aberrant protein expression [366]. They also hypothesise that, like *Arc1* mRNA targeting to neuronal dendrites via 3'UTR motifs [483], *cis*-acting regulatory RNA elements may be involved in the membrane targeting of neuronal mRNAs. The bioinformatic analysis described in this chapter provided evidence to suggest that this hypothesis may indeed be true. Biological validation that the novel *Drosophila* motifs do encourage mRNA EV-loading will provide further evidence to support this hypothesis.

Genes encoding transcripts carrying 3'UTR motif combinations were enriched for predicted molecular functions related to the cytoskeleton. The association between mRNA and the cytoskeleton is a key process to determine mRNA localisation for many cellular functions and it is often controlled via 3' localising sequences and RBP binding [484–487]. mRNAs localise to specific cellular compartments where they are required to facilitate protein sorting as mRNA transport presents a number of advantages over protein transport, such as reduced transport costs, lower risk of off-target protein activities and fine control of protein expression in both space and time [485]. mRNA localisation within the cell is achieved via machinery such as shuttling endosomes and SNAREs [488, 489], also key machinery in EV biogenesis. Therefore, mRNA localisation and mRNA EV-loading may be intricately linked,

with *cis*-acting factors, such as 3'UTR motifs, playing important roles in both cellular processes. It is also possible that mRNAs get incorporated into EVs as a secondary consequence of their cellular localisation.

The most enriched biological processes related to the genes encoding motif carrying transcripts in this analysis were associated with the regulation of development and signalling, as well as the regulation of cell communication. Motif carrying transcripts may carry out important long-range communication functions related to development. EVs have been shown to play important roles in development [167], for example, in *Drosophila* synaptic growth [34, 125] and wing disc development [47, 362, 490], as well as in *C. elegans* cellular adhesion and morphogenesis [140].

Regardless of whether mRNA EV incorporation is a mechanism by which cells release unwanted transcripts, a secondary consequence of their cellular localisation or an active loading processes to enrich EVs with mRNAs to carry out specific functions, EV mRNAs can have significant consequences on recipient cells. Furthermore, the mRNA features which facilitate their loading, in any of these capacities, are not fully understood. Gaining a deeper knowledge of these processes may improve the possibilities to adapt mRNA sequences to enhance the EV incorporation of customised cargos. Therefore, it is important that the function of these bioinformatically predicted EV-enriched motifs to modulate mRNA EV-loading is biologically validated in cell culture and *in vivo* and this will be a focus of **Chapter 5**.

3.5 Future Work

3.5.1 Determining RNA binding protein and miRNA interactions with EV-enriched motifs

The analysis described within this chapter highlighted RBPs whose binding site showed significant similarity to EV-enriched motif sequences. Biochemical validation of these predicted RBP-motif interactions is now required. The methods used to confirm these interactions will be RNA-centric, where the focus is on proteins associating with an RNA of interest [491]. A simple *in vitro* method to achieve this is via 5' or 3' RNA biotinylation, where biotinylated RNA bound to streptavidin beads is washed with cellular extract. Proteins which bind can be subsequently analysed via western blot or proteomics [491, 492]. Castello et al. (2013, 2017) present an exciting RNA and RBP UV cross-linking tool which could be used together with the transfection of constructs encoding an mRNA containing various 3'UTR motif combinations (described in **Chapter 5**), without the need for additional tags, to identify motif specific RNA binding proteins with minimal false-positive contamination [493, 494]. These experiments may validate the RBPs predicted to bind to motif regions via bioinformatic analysis or potentially highlight novel RNA interacting proteins with possible roles in modulating RNA EV-loading.

An interesting approach developed by Theil et al. (2019) utilised vIPR (*in vivo* Interactions by Pulldown of RNA), proteomics and small RNA-sequencing to identify both RBPs and miRNAs bound to an mRNA of interest [495]. A complementary bioinformatic analysis of motif containing transcripts, using tools such as TargetScan, may identify the presence of miRNA binding sites within motif regions [496]. Together, novel *trans*-acting factors involved in modulating RNA EV-loading may be identified, potentially providing novel ways to manipulate the EV-loading of customised RNA cargos.

3.5.2 Biological validation that the novel *Drosophila* EV-enriched motifs encourage the EV-loading of longer transcripts

Motif carrying transcripts were, on average, 3000 nts long, which was approximately twice as long as transcripts which lacked any 3'UTR motif occurrences. Intact EV RNA transcripts are generally thought to be 200 to 1000 nucleotides long, or otherwise likely to be present as RNA fragments [215, 224, 235, 480]. Whilst Bioanalyser profiles of EV RNA samples showed that EV RNA cargo was predominantly below 500 nucleotides in length, the presence of longer RNA transcripts cannot be confidently ruled-out as they may be present, but not in a high enough abundance to result in clear peaks. The length distribution profiles for the top 1000 most abundant cellular and EV mRNAs were highly similar; however, read coverage analysis of transcripts of interest is required to confirm if the full-length transcripts become loaded, rather than shorter RNA fragments. In addition, qPCR analysis of EV RNA can be used to validate that the motif combinations encourage the EV-loading of long transcripts, using multiple primers specifically designed to cover the entire transcript length, to confirm that the entire transcripts are EV-loaded, and not fragments thereof.

3.5.3 Comparison of surface EV RNAs with those loaded within the EV lumen

A highly debated topic in the EV field is whether RNAs are present within the EV lumen or are non-vesicular, perhaps via an interaction with RNA binding proteins [195]. EV purification methods can result in the isolation of molecules, such as RBPs, which are not associated with EVs [497]. For example, some non-vesicular miRNAs have been found extracellularly in association with Argonaute 2 (Ago2), which prevents miRNA RNase degradation [257, 276]. The analysis described in this chapter utilised the size-exclusion chromatography (SEC) method of EV isolation, which does allow better separation and EVs from protein contaminants [498, 499]. Previous studies have shown distinct RNA signatures for EV versus RNP-associated

RNA, with RBPs predominantly interacting with ncRNAs [78, 215], which were not the focus of this investigation. Nevertheless, improved methods are required to distinguish between these two extracellular RNA populations.

A protocol described by Ter-Ovanesyan (2019) can be used to distinguish between surface and luminal EV RNAs. Proteinase and RNase treatment of EV preparations is used to degrade proteins which protect surface RNAs, yielding purely luminal EV RNA for downstream analysis [500]. Future work will focus on carrying out these treatments on *Drosophila* EVs prior to RNA-sequencing to validate that the novel EV-enriched motifs identified in this analysis are bona fide EV-loading motifs and not associated with non-vesicular mRNAs. Additionally, an *in situ* hybridisation-electron microscopy (ISH-EM) method [501] could be applied to detect luminal EV RNA, with the high resolution perhaps allowing the determination of whether specific RNAs are located inside or at the surface of EVs [502].

3.6 Chapter Summary

This chapter described the bioinformatic analysis of *Drosophila* cellular and EV RNAs. Aside from ribosomal RNAs, the most common type of RNA present within both cellular and EV fractions was protein-coding. Although *Drosophila* cellular and EV RNA abundance levels were strongly correlated, EV-enriched and EV-specific protein-coding genes were identified, suggesting active-loading mechanisms may exist to encourage RNA EV-loading. One possible mechanism to encourage EV-loading is the presence of *cis*-acting factors, such as RNA motifs, particularly within mRNA untranslated regions. Comparative analysis of the 3'UTRs of EV-enriched and EV-specific mRNAs, versus the same region with cell-specific mRNAs, identified a number of EV-enriched motifs. These motifs displayed distinct distribution patterns along the 3'UTR; preferentially occurring in combinations and tending to occur in particular secondary structures. Motifs shared sequence similarity with RBP binding-sites, suggesting their possible involvement in the recruitment of *trans*-acting factors. EV RNA motif features such as combinations, preferential locations, secondary structures and repetitions, have been described separately in numerous mammalian systems. This *Drosophila* study identified these EV RNA features in a single analysis, suggesting possible mechanistic conservation of their function.

Motifs were prevalent within longer transcripts, suggesting they may encourage the EV incorporation of long transcripts. Further validation of full transcript loading, rather than RNA fragments, is now required. Gene ontology analysis highlighted that motif carrying transcripts were predicted to be associated with cellular components such as membranes and vesicles, with these transcripts also predicted to be involved in developmental, signalling and cytoskeleton pathways. This result suggested that the presence of motif sequences may affect the EV incorporation of specific mRNAs with particular functions. Experiments described in **Chapter 5** will focus on the biological validation of these EV motifs. This validation is essential to determine the effectiveness of these motifs to modulate long-range mRNA transfer and identify their potential to encourage the loading of alternative customised cargoes.

4

Developing novel tools to study cell-to-cell communication

Contents

4.1 Introduction 126
4.2 Rational design of the Cre-LoxP based reporter system in *Drosophila* 126
4.3 Validation of the Cre-LoxP based reporter system in cell culture 128
4.3.1 Cre-LoxP reporter system functionality in *Drosophila* SR2+ cells 128
4.3.2 Validation of *cre* mRNA EV-loading in *Drosophila* cell culture 129
4.3.3 The *Drosophila* Cre-LoxP system is functional in cell culture 136
4.4 Validation of the Cre-LoxP reporter system *in vivo* . . 138
4.4.1 Transfer of Cre detected within complex tissues using the *in vivo Drosophila* Cre-LoxP reporter system 139
4.4.2 Cre transfer across the anterior/posterior boundary of the imaginal wing disc was limited to early larval development 140
4.4.3 Cre transfer observed between different cell types in close proximity 142
4.4.4 The mammalian 25-nt EV motif did not alter reporter signalling *in vivo* 143
4.5 Discussion 146
4.5.1 A novel Cre-LoxP based reporter system to study EV mediated communication in *Drosophila* 146
4.5.2 Comparison of the *Drosophila* Cre-LoxP reporter system with other reporters of EV-mediated transfer 152
4.5.3 The mammalian 25-nt EV-loading motif did not influence *cre* mRNA loading in *Drosophila* 153
4.6 Future work 155
4.6.1 Further validation of *cre* mRNA EV-loading and transfer 155
4.7 Chapter Summary 158

4.1 Introduction

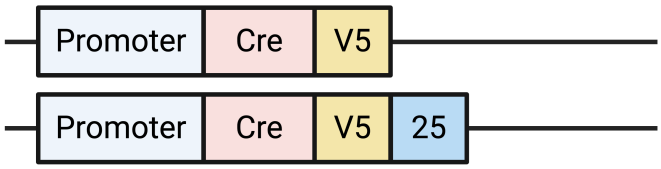
As demonstrated in **Chapter 3**, a bioinformatic comparison of cellular and EV mRNA cargo revealed the presence of EV-enriched mRNA features. Limited tools currently exist to biologically demonstrate the ability of these features to facilitate mRNA EV-loading, especially in complex *in vivo* scenarios. Therefore, encouraged by the success of the Cre-LoxP based reporter system to study EV mRNA transfer in mammalian models [309–312], a similar system was developed for use in *Drosophila melanogaster*. Using this *Drosophila* model, it was possible to circumvent some of the genetic constraints of mouse models, whilst also taking advantage of their easily accessible larval tissues. In addition, insights gained using the *Drosophila* model may provide an evolutionary perspective regarding EV-mediated cell communication mechanisms.

4.2 Rational design of the Cre-LoxP based reporter system in *Drosophila*

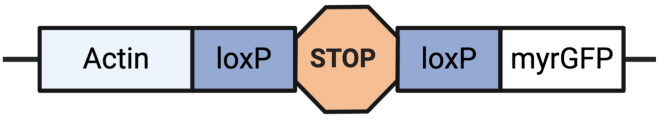
Plasmids encoding Cre recombinase (Cre) under the control of a universally expressed ubiquitin (Ubi) promoter or an artificial upstream activated regulatory sequence (UAS) promoter were generated, to allow constitutive or conditional (via the Gal4 transcription factor) Cre expression, respectively. Within these constructs, Cre was tagged at the C-terminus with the epitope V5 (Cre-V5), to allow visualisation of its expression via immunolabeling.

An additional version of the Cre construct was generated to contain a 25-nucleotide (25-nt) mammalian EV-enriched sequence [248] at the 3-prime end (Cre-V5-25) (**Figure 4.1i**). The ability of this mammalian motif to modulate mRNA EV-loading could then be assessed in the *Drosophila* system. The 25-nt motif was inserted after the stop codon so that it would only be present in *cre* mRNA and not in the encoded protein. Cre constructs were also generated to contain the EV-enriched motif combination identified in **Chapter 3** at their 3-prime end and results obtained using these constructs are described in **Chapter 5**.

Alongside these Cre constructs, the complementary reporter plasmid was developed to contain a ubiquitously expressed actin promoter and a loxP flanked stop-cassette upstream of a myristoylated green fluorescent protein (myr-GFP) coding sequencing (**Figure 4.1ii**). In the absence of Cre, the stop-cassette remains intact and terminates transcription before the myr-GFP sequence is reached, meaning the cell remains GFP negative (Reporter^{GFP-}). In the presence of Cre recombinase, recombination causes the excision of the stop-cassette and the subsequent expression of myr-GFP, forming Reporter^{GFP+} cells (**Figure 4.2**).



(i) Cre recombinase constructs containing a V5 tag for visualisation via immunostaining. Also generated were constructs with an additional 3-prime 25-nucleotide EV motif (represented as '25'). Constructs either contain a Ubiquitin (Ubi) or upstream activated regulatory sequence (UAS) as the promoter.



(ii) Reporter construct containing a loxP flanked stop-cassette upstream of a myristoylated green fluorescent protein (myr-GFP) sequence. This construct is under the control of an actin promoter.

Figure 4.1: Cre and reporter construct design.

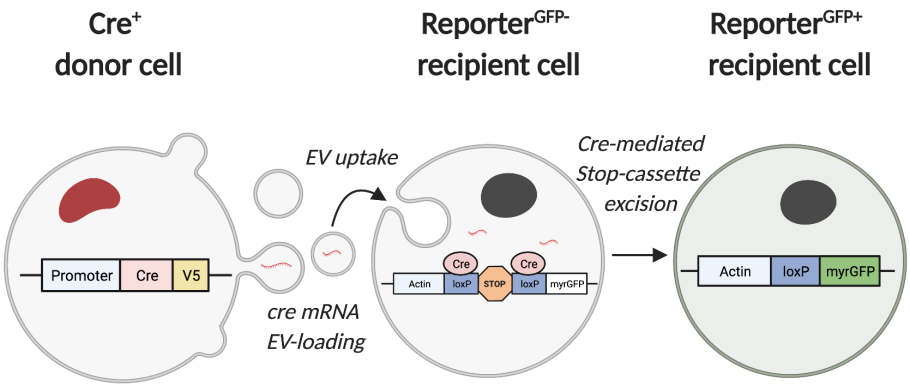


Figure 4.2: Schematic diagram to show EV transfer between a Cre donor cell and a recipient reporter cell. EVs loaded with *cre* mRNA are transferred to recipient cells, resulting in the production of Cre protein, excision of the loxP flanked stop cassette and the switch on of the myr-GFP signal. This process results in the conversion of the recipient cell into the Reporter^{GFP+} form.

4.3 Validation of the Cre-LoxP based reporter system in cell culture

4.3.1 Cre-LoxP reporter system functionality in *Drosophila* SR2+ cells

To determine the functionality of the Cre-LoxP based reporter system in cell culture, co-transfection experiments were carried out using *Drosophila* S2R+ cells. Within a single culture, cells were transfected with the reporter construct and either the Ubi Cre-V5 or Ubi Cre-V5-25 construct. The localisation of Cre protein is predominately nuclear. Within this single culture, cells were observed in three cell states: Cre^{+} alone (visible via nuclear anti-V5 immunolabeling), Cre^{+} and Reporter^{GFP+} (visible via nuclear anti-V5 immunolabeling and membrane GFP signal) or Reporter^{GFP+} alone (visible via membrane GFP signal) (**Figure 4.3i - iii**, respectively). These different cell states were due to the uptake of either just one or both plasmids during the co-transfection culture period. The appearance of Reporter^{GFP+} cells alone suggested that the Cre-LoxP system may detect the transfer of *cre* between cells in culture.

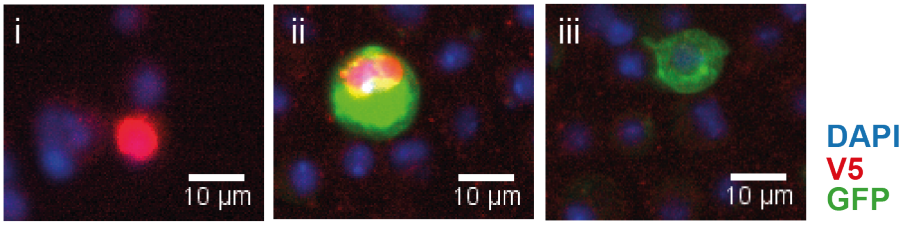


Figure 4.3: Cell states observed during a co-transfection experiment within a single culture. Cells were transfected with the Ubi Cre-V5-25 and Actin-loxP-myrGFP reporter constructs. (i) Some cells had taken up the Cre construct alone (visible via nuclear anti-V5 immunostaining). (ii) Other cells had taken up both the Cre and reporter constructs (visible via nuclear anti-V5 immunostaining and membrane GFP signal). (iii) Some cells had taken up the reporter construct alone, yet the stop-cassette was successfully excised (visible via membrane GFP signal).

4.3.2 Validation of *cre* mRNA EV-loading in *Drosophila* cell culture

To confirm that the Cre-LoxP based reporter system was sensitive to detect EV-mediated transfer of *cre* mRNA, RNA was isolated and quantified from purified EVs obtained from the conditioned media of Cre-transfected cells. To determine the biological influence of the mammalian EV motif in a *Drosophila* model, *cre* mRNA levels were compared in the presence and absence of the 25-nt EV motif.

Characterisation of EVs released by Cre transfected cells

Cell culture conditions and EV extraction protocols were optimised for use with *Drosophila* S2R+ cells as described in **Supplementary Section 7.1.2**. Cells were transfected with either the Ubi Cre-V5, Ubi Cre-V5-25 or Ubi control (not expressing Cre) plasmid. A media control (media conditioned in the absence of cells) was also included as advised by the International Society for Extracellular Vesicles (ISEV) [337]. A cell viability assay confirmed greater than 70% cell viability for each condition, ensuring that EV extractions were not largely contaminated with apoptotic bodies or cell debris (**Figure 4.4**).

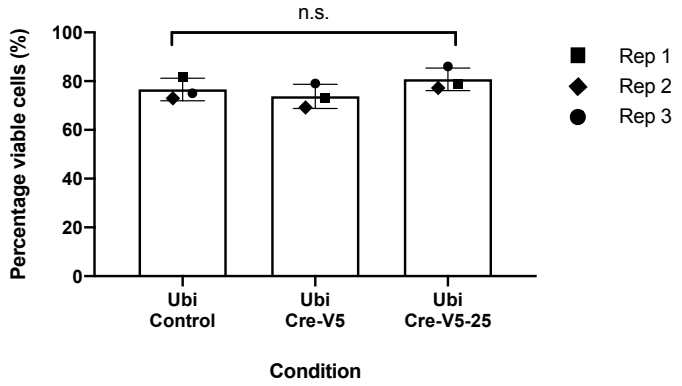


Figure 4.4: Percentage cell viability after a 48 hour conditioning period for EV extraction using Ubi control, Ubi Cre-V5 and Ubi Cre-V5-25 transfected cells. Mean ± standard deviation of 3 biological replicates shown. A Kruskal-Wallis test revealed no significant difference between all three conditions.

EVs were isolated from the conditioned media via SEC. No significant difference was observed in the particle concentrations obtained in cell conditioned media EV preparations. There was an expected reduction the particle concentrations observed in the media control condition (Media control = 0.32 ± 0.17 particles x 10^{10}/ml, Ubi control = 2.94 ± 2.47, Ubi Cre-V5 = 2.44 ± 2.01, Ubi Cre-V5-25 = 1.8 ± 1.21, **Figure 4.5**). This result confirmed that the majority of particles isolated in experimental conditions were released by the S2R+ cells, and that Cre expression did not significantly alter particle concentrations.

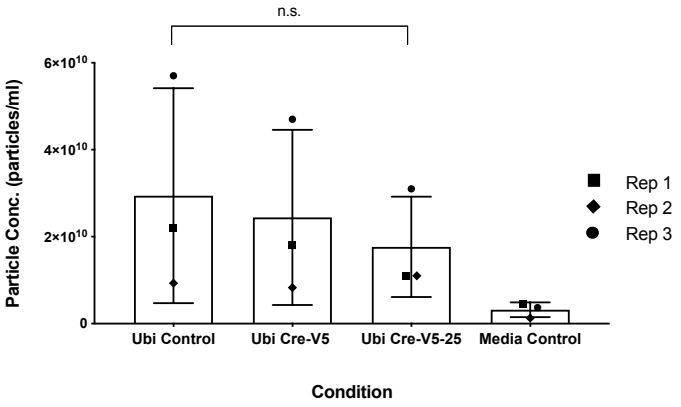


Figure 4.5: NTA particle concentrations of EVs extracted from Ubi control, Ubi Cre-V5, Ubi Cre-V5-25 and media control conditions (pooled SEC fractions 6 to 10). Mean ± standard deviation of three biological replicates is shown. A Dunn's multiple comparisons test revealed no significant difference for cell conditioned media.

Particle size distributions (also determined via ParticleMetrix NTA) showed similar profiles for the Ubi Cre-V5, Ubi Cre-V5-25 and Ubi control conditions. Particles were predominantly within the 45 to 300 nm size range, in line with the proposed EV size range [79] (**Figure 4.6**). Estimated mean particle size for EV preparations generated from cell conditioned media ranged from 110.1 ± 7.78 nm (Ubi control) to 115.7 ± 5.52 nm (Ubi Cre-V5-25). Particles from the media control condition appeared larger in size (134.6 ± 11.51 nm on average) (**Figure 4.7**). An absence of particles larger than 400 nm provided further evidence to suggest that EV samples were not contaminated with high levels of apoptotic bodies [79] (**Supplementary Figure 7.15**). Altogether, these results confirmed that protocols successfully isolated particles within the exosome and microvesicle size range. Results also showed that Cre expression did not alter the profile of EVs released.

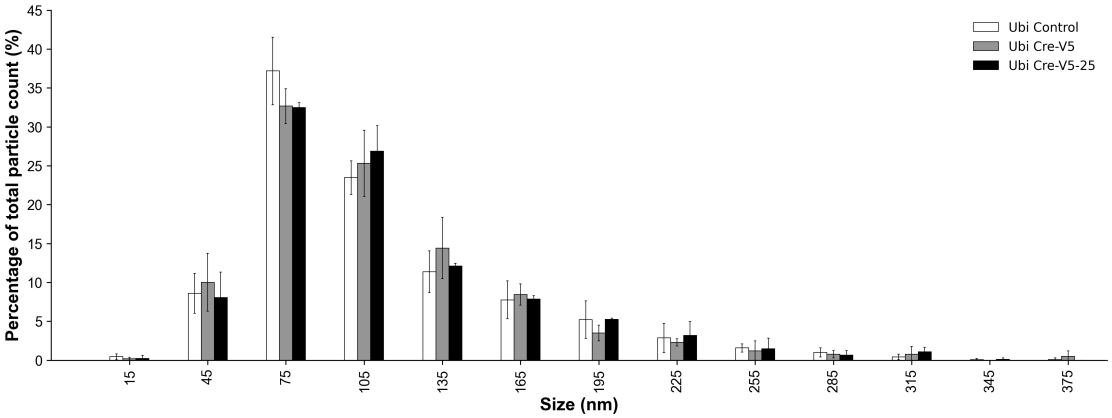


Figure 4.6: Size distribution profile (15 to 375 nm) determined via ParticleMetrix NTA of EVs extracted from Ubi control, Ubi Cre-V5 and Ubi Cre-V5-25 transfection conditions. Data shown for pooled fractions 6 to 10 after SEC extraction. Mean ± standard deviation of three biological replicates shown. Particle counts in each size window were normalised to total number of particles per sample to give a percentage of total particles value (%).

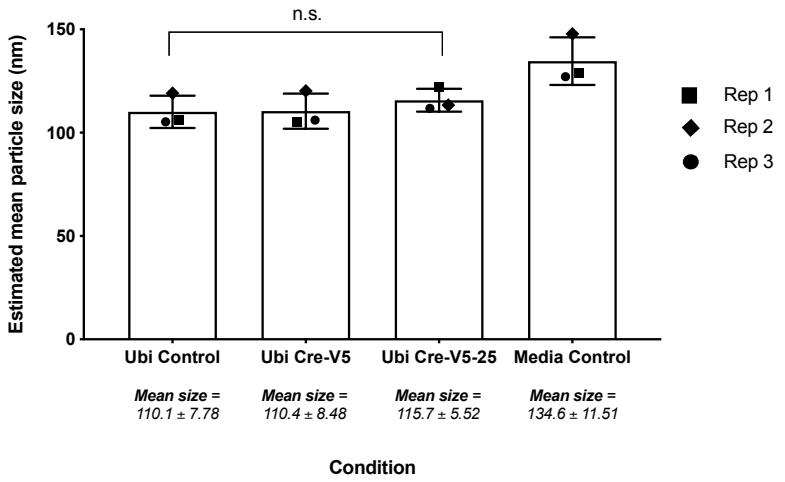


Figure 4.7: Estimated mean particle size (and standard deviation) determined using NTA size distribution data from three biological replicates. A Dunn's multiple comparisons test revealed no significant difference between the Ubi control, Ubi Cre-V5 and Ubi Cre-V5-25 conditions.

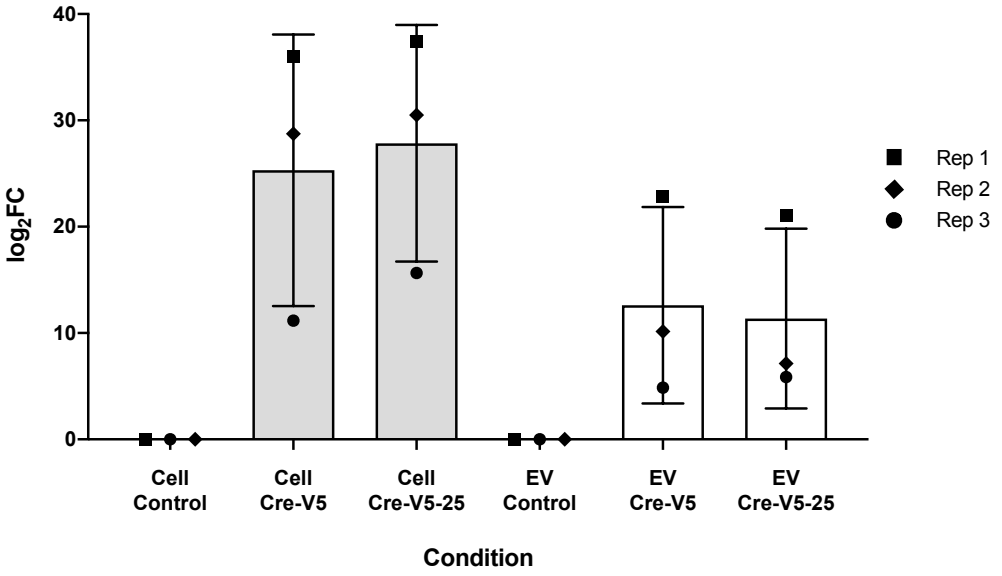
cre mRNA present in the EV RNA profile from Cre transfected cells

EVs were successfully isolated from the conditioned media of Cre transfected cells. Following this, quantitative real-time PCR (qPCR) was used to confirm the presence of *cre* mRNA from RNA extracted from these EV fractions. Log_2 Fold Change (Log_2FC) values for *cre* mRNA within cell and EV RNA fractions are shown in **Figure 4.8i** (normalised to the housekeeping gene, *Arc1*). The housekeeping transcript *Arc1* was selected due to its successful use by Lefebvre et al. (2016) when studying *Drosophila* S2R+ EV mRNA [366]. This transcript also presented high FPKM values in the *Drosophila* S2R+ cell and EV RNA samples analysed via the RNA-seq analysis described in **Chapter 3** (Cell FPKM = 277.6 ± 22.5, EV FPKM = 13451.1 ± 2553.9, Mean ± standard deviation).

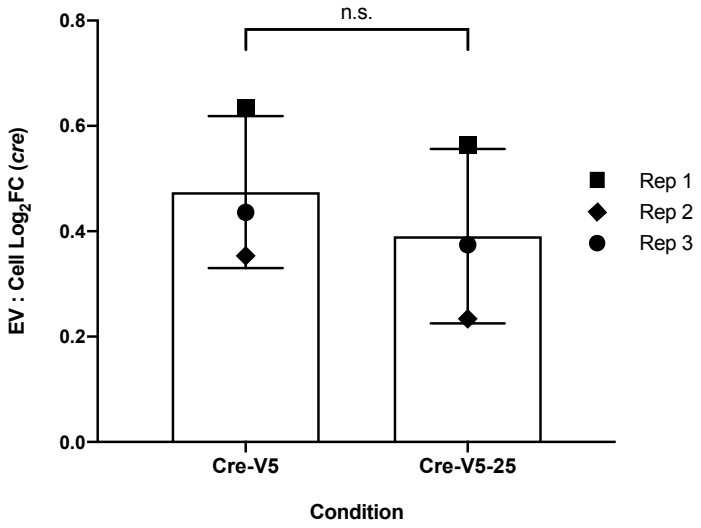
In cell and EV fractions isolated from the Ubi Cre-V5 or Ubi Cre-V5-25 conditions, *cre* was successfully detected (Log_2FC (cell): Ubi Cre-V5 = 25.30 ± 12.77, Ubi Cre-V5-25 = 27.85 ± 11.12 / Log_2FC (EV): Ubi Cre-V5 = 12.62 ± 9.232, Ubi Cre-V5-25 = 11.36 ± 8.451). Media control and Ubi control conditions presented high Ct values for *cre* (**Figure 4.9**) indicative of an absence of *cre* mRNA. Media control samples also presented high Ct values for *Arc1*, suggesting the RNA

isolated was not contaminated with RNA present in the media. Comparing EV to cell Log_2FC ratios to account for changes in *cre* levels in the parental cells (likely due to different transient transfection rates between replicates), no significant difference was observed between levels of EV-loaded *cre* mRNA upon the addition of the 25-nt EV motif (EV:Cell Log_2FC ratio: Cre-V5 = 0.4744, Cre-V5-25 = 0.3907, p = 0.7, **Figure 4.8ii**). These results suggested the function of the mammalian 25-nt motif to enhance EV-loading of mRNA may not be effective in *Drosophila*; however, further investigations are required to fully understand the biological significance of this motif in a *Drosophila* model.

Although the replicates shown in **Figure 4.8** are variable, assessing the raw Ct values (**Figure 4.9**) provided context for these results. Whilst levels of *Arc1* remained highly consistent between parental cell replicates, cellular *cre* mRNA levels were more variable, likely the result of variable levels of transient Cre transfection between each replicate. Some *Arc1* variability was observed between EV replicates, however, variability was also observed for this gene in the RNA-seq analysis; following the general trend that EV RNAs showed a higher degree of variability between replicates compared to the parental cell (**Section 3.2.1**).



(i) Log_2FC values for *cre* mRNA levels in cell and EV RNA extracted from Cre transfected cells or the EVs they release.



(ii) EV to cell ratio of Log_2FC values, showing no significant change in the *cre* mRNA levels present in EV loaded mRNAs (via a two-tailed Mann-Whitney test). This analysis accounted for differences in Cre transfection in parental cells.

Figure 4.8: qPCR data showing *cre* mRNA levels, using *Arc1* as the housekeeping gene. Mean ± standard deviation of three biological replicates shown. These results highlighted that the addition of the mammalian 25-nt motif to *cre* mRNA did not enhance its EV abundance.

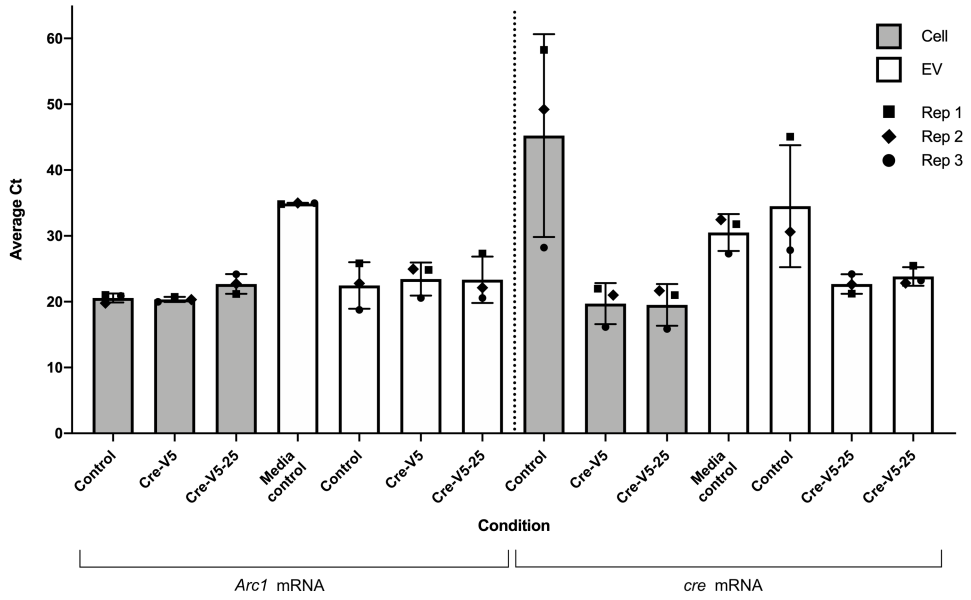


Figure 4.9: Raw Ct values for qPCR data showing *cre* mRNA levels, using *Arc1* as the housekeeping gene. Mean ± standard deviation of three biological replicates shown. Each point is an average of one or two technical qPCR replicates.

An RNase protection assay was carried out[34] to confirm the presence of luminal *cre* mRNA, as opposed to non-vesicular mRNA. Isolated EVs were treated with RNase, in the presence and absence of a lysis agent, prior to RNA extraction [502, 503]. In the absence of the lysis agent, EV surface RNA is degraded by the RNase, but EV-loaded RNA is protected by the EV membrane. In the presence of the lysis agent, the EV membrane is broken down and luminal RNA is exposed and degraded. Encouragingly, a preliminary experiment highlighted that a proportion of *cre* mRNA was protected from RNase treatment (**Figure 4.10**). Therefore, it is likely that a proportion of this mRNA is present within the EV lumen.

[34]by Elizabeth Dellar

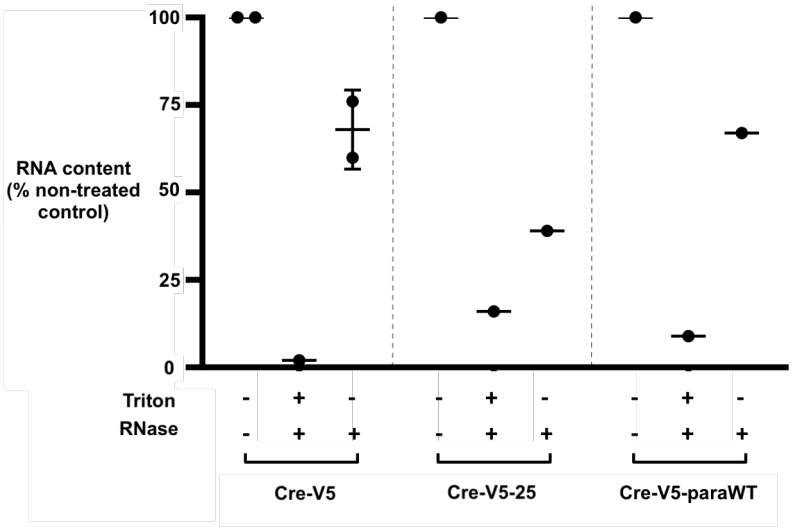


Figure 4.10: RNase protection assay to detect luminal *cre* EV mRNA. Relative abundance of *cre* mRNA (determined via qPCR) detected in the presence of RNase with or without Triton, relative to untreated (-/-) conditions. Each datapoint represents a biological replicate. Shown here are the protection values obtained using Ubi-Cre-V5, Ubi-Cre-V5-25 and Ubi-Cre-paraWT constructs. Development and validation of the Ubi-Cre-paraWT construct is described in **Chapter 5**.[35]

4.3.3 The *Drosophila* Cre-LoxP system is functional in cell culture

Although a qPCR analysis of *cre* mRNA cargo highlighted no EV enrichment upon the inclusion of the mammalian 25-nt EV-motif, the biological effect of this motif could be determined using the Cre-LoxP reporter system. To this end, the number of cells presenting each cell status was quantified during co-transfection experiments in the absence or presence of the 25-nt motif. No significant increase in the prevalence of Cre^+ and co-transfected cells (Cre^+ and $Reporter^{GFP+}$) was observed upon the addition of the 25-nt motif. There was a tendency for the the level of $Reporter^{GFP+}$ alone cells to increase (Ubi Cre-V5: 2.9 ± 2.4 $Reporter^{GFP+}$ cells/frame, Ubi Cre-V5-25: 4.26 ± 2.6, **Figure 4.11**) upon the inclusion of this mammalian EV motif; however, this increase was also not significant ($p = 0.0517$). This result was in

[35]Data generated by Elizabeth Dellar

concordance with the qPCR analysis, suggesting that the inclusion of the mammalian 25-nt EV motif may not modulate *cre* mRNA transfer in a *Drosophila* model.

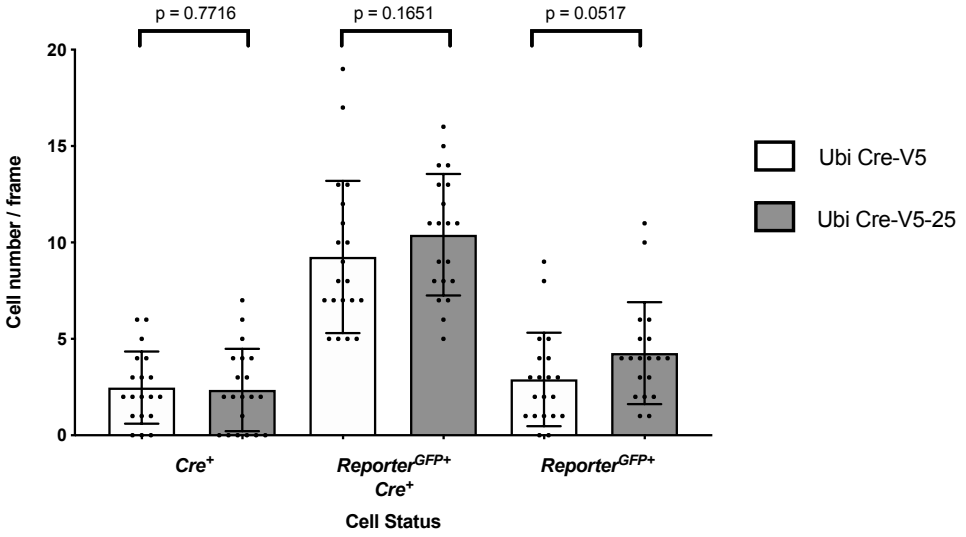


Figure 4.11: Within the same culture, cells were transfected with the Reporter plasmid, together with either the Ubi-Cre-V5 or Ubi-Cre-V5-25 plasmid. The number of cells presenting three different cell states were then quantified within this culture: Cre^{+} alone, $Reporter^{GFP+}$ and Cre^{+} or $Reporter^{GFP+}$ alone. Addition of the 25-nt EV-loading motif to *cre* mRNA resulted in no significant increase in the number of $Reporter^{GFP+}$ alone cells during co-transfection experiments. 20 fields of view were analysed (10 per technical replicate). Mean ± standard deviation is displayed. Statistical significance was determined via two-tailed Mann-Whitney test (Outliers were removed (R=1%)).

It is important to note, however, that a number of caveats exist for these co-transfection experiments. Firstly, because Cre and reporter plasmids were transfected simultaneously, cells which appeared $Reporter^{GFP+}$ alone may have been cells transfected with both the Cre and reporter plasmids, with Cre levels too low to be detected via immunolabeling. Secondly, there was a possibility for cell fusion or direct cell contact events to occur, also facilitating *cre* transfer. To circumvent these limitations, the power of the *Drosophila* model was harnessed to generate a more effective *in vivo* model of cell communication.

4.4 Validation of the Cre-LoxP reporter system *in vivo*

To overcome some of the drawbacks related to the cell culture experiments and encouraged by the qPCR analysis highlighting *cre* mRNA within isolated EVs, the system was transferred to the *in vivo* setting. *Drosophila melanogaster* fly lines containing the UAS Cre-V5, UAS Cre-V5-25 and Actin-loxP-myrGFP constructs were generated (**Table 2.8**).

Initial tool validation was carried out using the imaginal wing disc (IWD) and eye/antennal imaginal disc (E/A ID), well-established models of cell communication and tissue patterning [346]. Using the UAS-Gal4 system, Cre expression was induced in selected groups of cells and the pattern of Reporter^{GFP+} cells in relation to these sources of Cre was followed. A series of control experiments were completed to ensure that the Cre and reporter elements worked appropriately *in vivo*. Previous studies have reported UAS promoter leakiness in the absence of a Gal4 driver [504, 505]. In the absence of a Gal4 driver, no leaky expression of UAS-Cre and no spontaneous excision of the loxP flanked stop-cassette of our reporter occurred *in vivo* (**Supplementary Figure 7.16**). This demonstrated that any Reporter^{GFP+} signal was exclusively linked to either expression or uptake of Cre (in mRNA or protein form).

For each of the Gal4 drivers used to activate Cre expression in a particular cellular domain, the restricted expression pattern of that driver was confirmed via FLP/FRT lineage tracing experiments [506]. Lineage tracing involves the permanent lacZ labelling of all cells which produced Gal4 at any point during their lifetime. LacZ could then be identified via β-galactosidase immunolabeling. Initial *in vivo* experiments in the IWD were carried out using the ptcGal4 driver. Previous studies have confirmed the exclusive expression of this driver in the IWD anterior, in all anterior cells during early development and later refined to a narrow strip of cells adjacent to the anterior/posterior border [504, 507–509]. Lineage tracing carried out in this analysis highlighted a similar pattern (**Figure 4.12**).

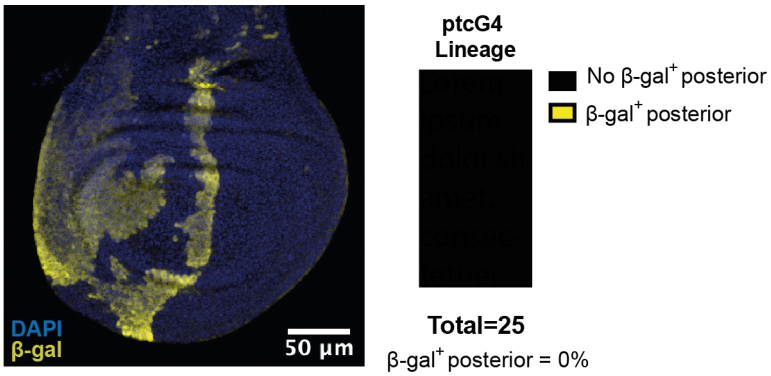


Figure 4.12: Lineage tracing of the ptcGal4 driver in the IWD via β-galactosidase (β-gal) immunolabeling. Also shown are the proportion of IWDs showing none or β-gal positive staining in the posterior. Genotype: UAS Flp/ptcGal4; Actin-FRT-stop-FRT-lacZ/TubGal80^{ts}, UAS HisRFP.

4.4.1 Transfer of Cre detected within complex tissues using the *in vivo Drosophila* Cre-LoxP reporter system

Using the ptcGal4 driver, Cre expression was driven specifically in the anterior domain of the IWD. In line with previous studies, Cre toxicity was observed when Cre expression levels were high [367, 510]. Therefore, the Gal4 suppresser, Gal^{80}, was added to reduce Cre expression to a non-toxic levels when experiments were carried out at 25°C. Further exploration of the measures taken to overcome Cre toxicity are explored in **Section 5.3**.

Within this system, the ectopic expression of Cre under the regulation of this driver induced the appearance of Reporter^{GFP+} cells in the anterior compartment of the wing disc, but also unexpectedly in the posterior compartment (**Figure 4.13**). These results suggested Cre transfer was occurring across the anterior/posterior (A/P) boundary. Furthermore, Reporter^{GFP+} signal occurred at a distance from the A/P boundary approximately 45% of the time, creating a 'gap region,' indicated by the asterisks in **Figure 4.13** (**Supplementary Figure 7.17**). These results suggested long-range mechanisms of communication may facilitate Cre transfer in this cellular scenario.

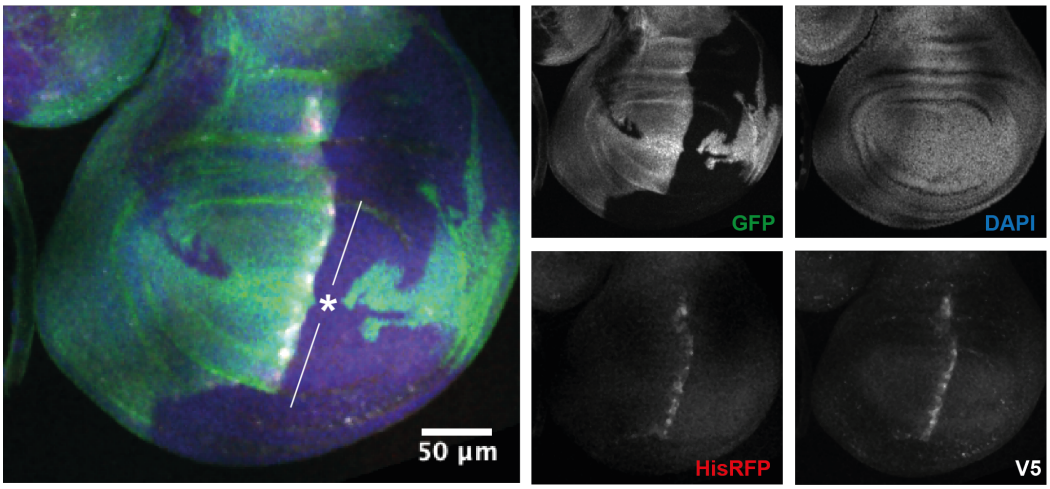


Figure 4.13: The *in vivo* Cre-LoxP system reports Cre transfer across the IWD anterior/posterior boundary. Reporter^{GFP+} cells observed in the anterior and posterior side. Posterior Reporter^{GFP+} frequently arises beyond a 'gap region,' highlighted by the white asterisk and line annotation. Genotype: UAS Cre-V5-25/+; Actin-loxP-myrGFP/ptcGal4; TubGal80^{ts}, UAS HisRFP/MKRS).

4.4.2 Cre transfer across the anterior/posterior boundary of the imaginal wing disc was limited to early larval development

Given the importance of morphogen signalling during early wing disc development, with studies presenting links to EVs in these contexts [36, 47], experiments were carried out to determine if the communication events detected using this Cre-LoxP system were limited to any particular stage of larval development. The thermosensitive (ts) Gal4 suppressor, Gal80^{ts} was used to finely control the timing of Cre expression. At 18°C this suppresser was fully active, meaning the Gal4 driver (and associated Cre expression) was switched off. At 25°C, the Gal80^{ts} activity was reduced, meaning expression of the Gal4 driver was increased and thus, Cre expression was switched on. This temperature sensitive quality was used to define two timepoints at which Cre expression would be switched on: only early in larval development or only late in larval development. The schematic in **Section 2.3.3** describes the experimental set-up (**Figure 2.2**).

Results showed that the frequency of IWD posterior Reporter^{GFP+} signal, when Cre expression occurred during early larval development only, was 33.3 ± 9.1%.

This was an increase from the 25 ± 8.2% observed when Cre expression occurred during the entirety of larval development; however, this increase was not significant (p = 0.4952, **Figure 4.14**). There was tendency for mean percentage Reporter^{GFP+} posterior area to decrease when Cre expression was limited to early development only, when compared to Cre expression throughout larval development (Cre-V5-25: 40.0 ± 15.3, Cre-V5-25 early only: 25.2 ± 19.6, **Supplementary Figure 7.18**); however, this decrease was also not significant (p = 0.2523).

A surprising observation was that no posterior IWD Reporter^{GFP+} signal was observed when Cre expression was limited to late larval development stages only (**Figure 4.14**). These results suggested that Cre transfer across the anterior/ posterior boundary of the IWD was time sensitive, occurring in early larval development stages only.

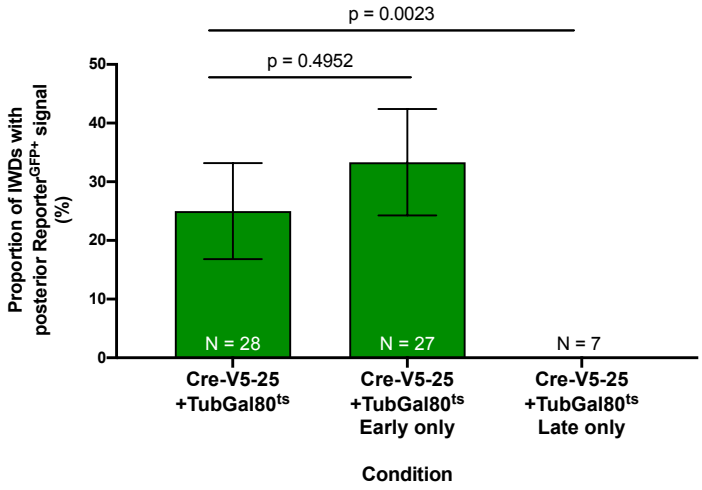


Figure 4.14: Proportion of IWDs displaying posterior Reporter^{GFP+} signal during the timeline experiment with ptcGal4, controlling Cre expression during the whole of larval development, during early larval development only or during late larval development only. Percentage proportion ± SE is shown. Statistical analysis was carried out as described in **Section 2.5.1**. This analysis highlighted that reporter signalling was limited to early developmental stages only. Genotype: UAS Cre-V5-25/+; ptcGal4/Actin-loxP-myrGFP; TubGal80^{ts}, UAS HisRFP/MKRS.

4.4.3 Cre transfer observed between different cell types in close proximity *in vivo*

Encouraged by the success of the Cre-LoxP system to detect Cre transfer within a single tissue, the tool was then evaluated to determine its effectiveness to detect Cre transfer between different cell types in close proximity. The hmlGal4 driver was used to express Cre in hemocytes, the immune surveillance cells of *Drosophila* [511]. Investigations were then carried out to determine the presence of Reporter^{GFP+} signal in eye/antennal imaginal discs. Communication between these tissues has been documented, with hemocytes playing a role in the phenomenon of apoptosis-induced proliferation [355]. Interestingly, Reporter^{GFP+} signal was observed in the eye/antennal disc epithelial cells when Cre was expressed in hemocytes (**Figure 4.15**). Transverse cross sections showed hemocytes present on the surface of the eye/antennal disc, with Reporter^{GFP+} clones present in the disc proper as well as the peripodial membrane. HmlGal4 lineage tracing confirmed this Gal4 line was not expressed in eye/antennal discs (**Figure 4.18ii**). Taken together, these results suggested that the Cre-LoxP system was sensitive to detect communication between different cell types in close proximity.

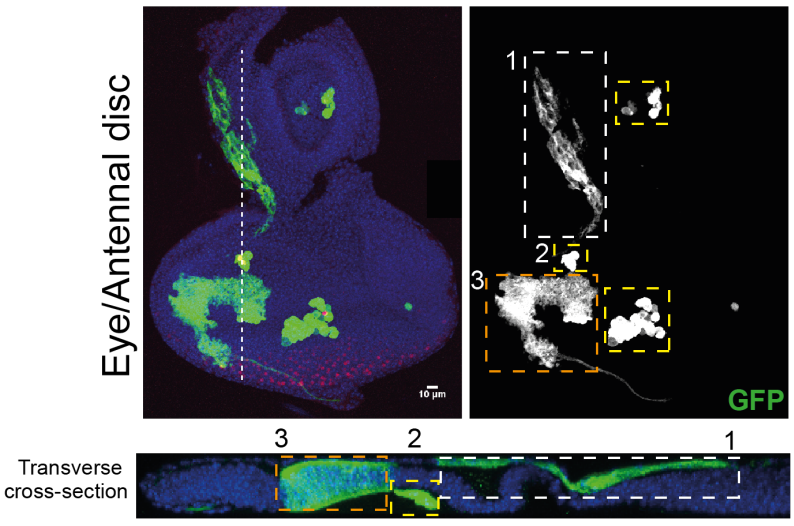


Figure 4.15: Eye/antennal disc Reporter^{GFP+} signal when Cre was expressed in hemocytes using hmlGal4. Reporter^{GFP+} cells were observed in the disc proper (orange box, 3) and in the peripodial membrane (white box, 1). Hemocytes are highlighted by yellow boxes (2). Genotype: UAS Cre-V5-25/+; hmlGal4, UAS GFP/Actin-loxP-myrGFP; droncKO, TubGal80^{ts}, UAS HisRFP/MKRS.

4.4.4 The mammalian 25-nt EV motif did not alter reporter signalling in two different *in vivo* cellular contexts

To determine if the *in vivo* system was sensitive to the effects of mRNA 3'UTR motifs, the prevalence of Reporter^{GFP+} signal observed in the presence and absence of the 25-nt EV motif in different cellular scenarios was quantified. In the IWD experimental setup described above, the addition of the 25-nt EV motif to *cre* mRNA did not significantly enhance the frequency of clones observed in the posterior compartment (Cre-V5 = 39.1 ± 5.9%; Cre-V5-25 = 25 ± 8.2%, p = 0.1606, **Figure 4.16**). Nor did the addition of the 25-nt motif alter the mean percentage posterior area becoming Reporter^{GFP+} (**Figure 4.17**). In this cellular scenario, it appears that the mammalian EV motif did not affect Cre transfer across the anterior/posterior IWD boundary.

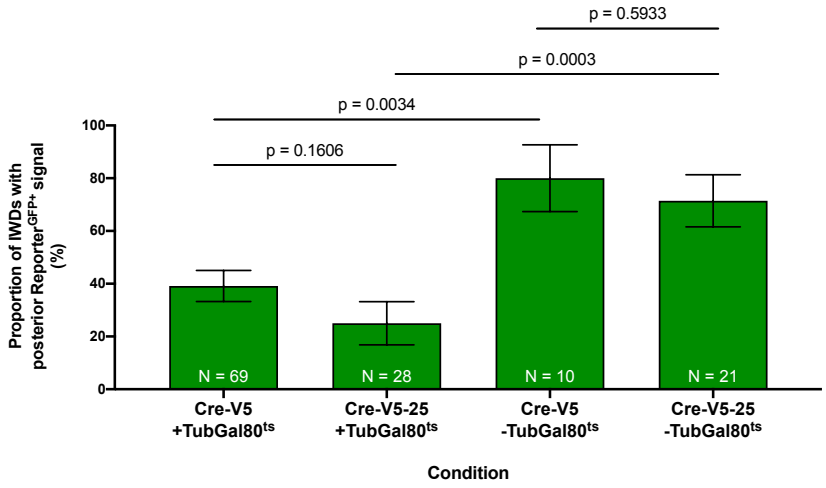


Figure 4.16: Proportion of IWD posteriors showing any Reporter^{GFP+} signal for the Cre-V5 and Cre-V5-25 conditions in the presence or absence of TubGal80^{ts}. Percentage proportion ± SE is shown. Basic genotype: UAS Cre/+; ptcGal4/Actin-loxP-myrGFP (with and without TubGal80^{ts}, UAS HisRFP on chromosome III and using UAS Cre-V5 or UAS Cre-V5-25 as described in the figure). Statistical analysis was carried out as described in **Section 2.5.1**.

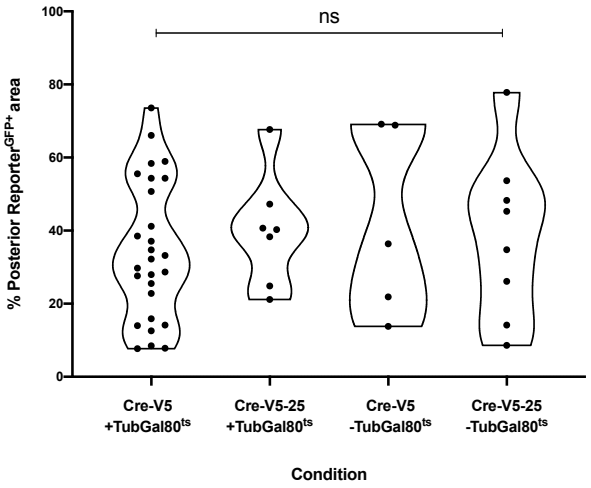


Figure 4.17: Percentage area of IWD posterior Reporter^{GFP+} signal for the Cre-V5 and Cre-V5-25 conditions in the presence or absence of TubGal80^{ts}. Statistical significance values determined via a Kruskal-Wallis test. No significant difference was observed (p = 0.8434). Genotype: UAS Cre/+; ptcGal4/Actin-loxP-myrGFP (with and without TubGal80ts, UAS HisRFP on chromosome III and using UAS Cre-V5 or UAS Cre-V5-25 as described in the figure).

When harnessing the Cre-LoxP system to study communication between hemocytes and epithelial disc tissues, the frequency of E/A IDs with Reporter^{GFP+} signal tended to decrease upon the addition of the 25-nt motif; however, this decrease was not significant (Cre-V5: 25 ± 7.21%, Cre-V5-25: 9.38 ± 5.15%, p = 0.078, **Figure 4.18**). These results suggested that differential signalling was not driven upon the incorporation of the mammalian 25-nt EV-loading motifs in multiple *in vivo* contexts in the *Drosophila* model.

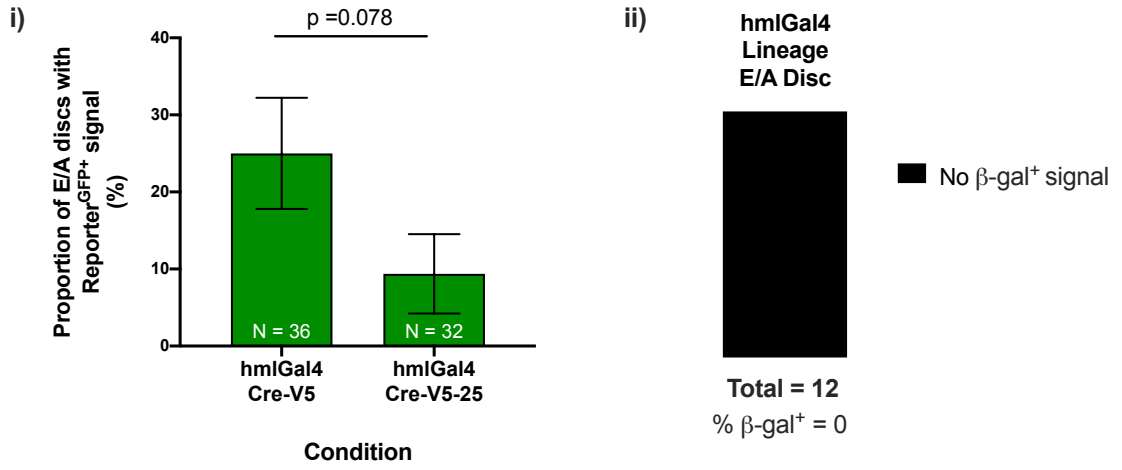


Figure 4.18: (i) Proportion of eye/antennal (E/A) discs showing Reporter^{GFP+} signal when Cre was expressed in hemocytes using hmlGal4. Shown is the proportion in the absence and presence of the 25-nt EV motif. Percentage proportion ± SE is shown. Genotype: UAS Cre/+; hmlGal4, UAS GFP/Actin-loxP-myrGFP; droncKO, TubGal80^{ts}, UAS HisRFP/MKRS (UAS Cre-V5 or UAS Cre-V5-25 was used as described in the figure). (ii) LacZ lineage tracing using the hmlGal4 driver confirms that Reporter^{GFP+} clones in the E/A disc are not due to leaky Gal4 expression. Genotye: UAS Flp/hmlGal4, UAS GFP; Actin-FRT-stop-FRT-lacZ/droncKO, TubGal80^{ts}, UAS HisRFP. Statistical analysis was carried out as described in **Section 2.5.1**.

4.5 Discussion

4.5.1 A novel Cre-LoxP based reporter system to study EV mediated communication in *Drosophila*

This chapter described the successful establishment of a Cre-LoxP reporter system to study long-range cell communication in *Drosophila melanogaster*. Initial validation experiments were carried out in cell culture before investigations were transferred to the *in vivo* setting to study cell communication within and between complex tissues.

Validation of EV isolation methods used in the *Drosophila* Cre-LoxP cell culture system

Cell culture experiments documented in this chapter confirmed that the transfection of *Drosophila* S2R+ cells with a plasmid encoding Cre did not alter cell viability or disrupt the EV size or concentration profiles obtained. The estimated mean particle size of EVs released by Cre transfected and control transfected cells (112 - 113.5 nm) were comparable to, but slightly lower than, those previously described for *Drosophila* S2R+ cell EVs (150.9 ± 3.0 nm) [366]. Differences in the hardware and software used for these separate NTA analyses may be the cause for this disparity [512]. The larger particle size observed previously may be the result of protein aggregates, commonly observed after ultracentrifugation EV isolation [366, 513–515]. EVs are less prone to aggregation after size-exclusion chromatography (SEC), the method used in this analysis; however, smaller lipoproteins which mimic the characteristics of EV may be enriched [516, 517]. Given that lipoproteins are often sourced from the serum used to supplement culture medium [46], and validation experiments in this analysis were carried out in serum free conditions, lipoproteins are unlikely to be contributing to variations in particle size.

Validation of *cre* mRNA EV-loading in the *Drosophila* Cre-LoxP cell culture system

The appearance of Reporter^{GFP+} alone cells, with undetectable levels of Cre expression, under Cre and Reporter co-transfection conditions suggested that long-range Cre transfer may have occurred, in either protein or mRNA form. One

hypothesis for the mechanism of Cre transfer between cells was EV-mediated transfer. qPCR analysis confirmed the presence of *cre* mRNA within the RNA extracted from the EVs released by Cre transfected cells. An additional RNase protection assay confirmed that a proportion of this *cre* mRNA was protected from RNase treatment and therefore, likely to be present within the EV lumen. Encouragingly, previous cell culture analyses of the mammalian Cre-LoxP systems also confirmed the presence of *cre* mRNA within EVs, with Cre protein undetectable via Western Blotting analysis or ELISA [309–312]. ISEV position statements do highlight that functional RNA transfer on an EVs surface, likely to be present within nucleoprotein complexes to prevent RNA degradation, has not been confirmed and such mechanisms are inconsistent with the traditional model of RNA transfer via EV fusion [502, 503]. Taken together, this evidence suggests that the *Drosophila* Cre-LoxP system may be successfully reporting the functional transfer of *cre* mRNA via EVs.

Beyond the transfer of *cre* mRNA via EVs, however, it is important to assess the impact of other mechanisms, such as Tunneling nanotubes (TNTs) or free media transfer of Cre protein, which may influence the *Drosophila* Cre-LoxP reporter system. TNTs are well known for their role in transferring cytoplasmic organelles; however, they have been shown to play a role in miRNA and mRNA transfer [518–521]. Further experiments described in **Section 4.6.1** will be carried out to investigate the impact of these alternative long-range cell communication mechanisms on the *Drosophila* Cre-LoxP system.

Cre transfer observed within and between tissues *in vivo*

Following the detection of EV-loaded *cre* mRNA and the successful functionality of the Cre-LoxP system in cell culture, this system was transferred to the *Drosophila in vivo* setting. EV-mediated communication can be harnessed by organisms to aid the proper coordination of function and development; however, their significance, especially *in vivo*, remains poorly understood. To date, much work has utilised EVs isolated from cell culture media; however, varying isolation methods, EV heterogeneity and cell-type specific functions have made understanding the

fundamental properties of EV-loading and transfer a challenge. Model organisms such as Zebrafish, mice, *C. elegans* and *Drosophila* have allowed a number of these constraints to be overcome, providing invaluable insights into the roles EVs play in development and disease progression [139, 271, 332].

The Cre-LoxP reporter system was combined with the powerful genetic tractability of *Drosophila* to take advantage of their readily accessible tissues, which are well established models of cell communication. Investigations first focused on the imaginal wing disc; a model tissue introduced in **Section 1.3.1**. Using the UAS-Gal4 system, UAS-Cre expression was limited to a specific subset of cells. The Actin-LoxP-myrGFP reporter was ubiquitously expressed, making it possible to observe distant cells receiving Cre by observing a green fluorescent signal beyond the Cre expression domain. When Cre was expressed in the anterior side of the IWD, under the control of the ptcGal4 driver, Reporter^{GFP+} signal was observed in both the anterior and posterior side. Lineage tracing experiments, and experiments carried out in the absence of a Gal4 driver, confirmed this was not due to 'leaky' Cre expression outside the ptc domain, nor due to spontaneous excision of the loxP flanked stop cassette from the reporter construct. This result suggested Cre transfer was occurring across the anterior/posterior boundary and that the *Drosophila* Cre-LoxP system was sensitive to detect long-range communication within a complex tissue.

Communication events were also observed between different cell types in close proximity. When UAS-Cre expression was driven in hemocytes, the immune surveillance cells of *Drosophila*, using hmlGal4, Reporter^{GFP+} signal was observed in imaginal disc epithelial cells. Bi-directional communication between epithelial cells and hemocytes has previously been reported, with hemocytes 'intimately associated' with sites of cell death to remove cell debris [522]. Hemocyte and epithelial communication occurs to promote apoptosis induced proliferation (AiP), where dying cells activate neighbouring surviving cells to divide [523]. In an immortalised model of AiP in imaginal discs, extracellular reactive oxygen species (eROS) was released to recruit hemocytes, which in turn promoted epithelial tissue repair and regeneration [355].

Interestingly, Hervera et al. (2018) reported results in a mammalian model which reflected those observed in this AiP *Drosophila* system, additionally implicating EVs in these processes [524, 525]. It is possible that similar EV-mediated communication is taking place at the interface between immune cells and *Drosophila* imaginal disc tissues, with the Cre-LoxP reporter system appearing sensitive to detect it. In addition, as results obtained using the Cre-LoxP reporter system were in the absence of AiP, epithelial cell and hemocyte communication may facilitate natural cell turnover which is required for growth regulation and tissue homeostasis [346, 526].

The *Drosophila* Cre-LoxP system highlighted communication events occurring during particular periods of development

Many of the long-range cell communication mechanisms studied in the imaginal wing disc have been related to morphogen release and transfer; vital processes for tissue patterning and development [527]. Cell fate and the orientation of tissue growth is largely determined during larval development by specific gene expression patterns and signalling pathways [346, 528]. Many signalling events originate at the boundaries between cellular compartments, such as the anterior/posterior boundary, which is established early in embryogenesis [529–531]. The temporal dynamics of these signalling and gene expression patterns has proved difficult to study and whilst advancements in *in vivo* imaging techniques and the long-term cultivation of imaginal wing discs have improved this field of study [532, 533], improved tools are required to understand the regulation and functional importance of these processes.

Using the Cre-LoxP reporter system, UAS-Cre expression could be temporally controlled using the Gal4 suppressor, Gal^{80}. This highlighted that communication from anterior to posterior was limited to early development stages only (up to 3 days after egg laying). This time point includes embryonic, first instar and second instar larval stages (**Figure 4.19**). It is possible that the communication events detected by the Cre-LoxP system were related to early signalling processes which establish fundamental properties of tissue organisation and growth. Further analysis will focus on narrowing down this 'active window' of Cre transfer, potentially highlighting stages in *Drosophila* wing disc development where long-range communication in the

anterior to posterior direction may play vital roles in the growth and development of imaginal wing discs.

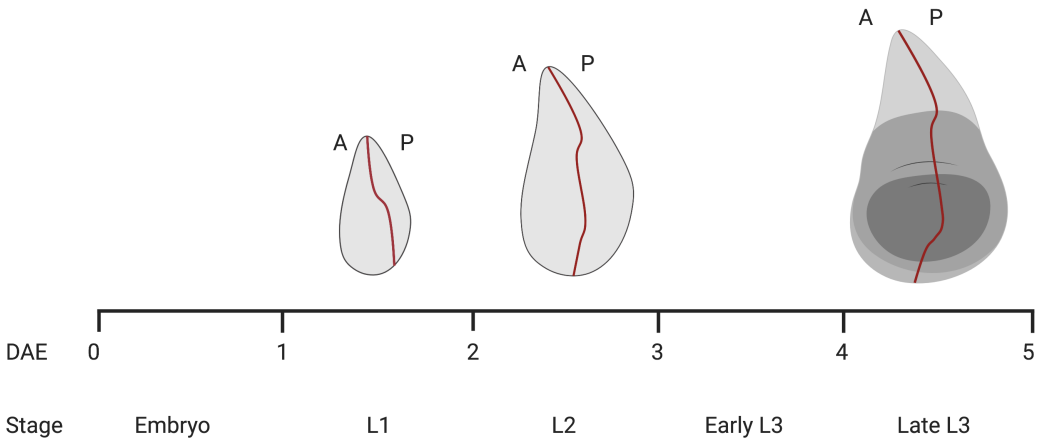


Figure 4.19: Schematic diagram depicting the anterior/posterior boundary (red line) during stages of *Drosophila* imaginal wing disc development. Developmental stage and days after egg laying (DAE) are also shown. L1 = First instar larvae. L2 = Second instar larvae. L3 = Third instar larvae. A = Anterior. P = Posterior.

EV-mediated cell communication may be restricted to specific areas of the imaginal wing disc

Within the IWD, when Cre expression was driven in the anterior compartment, the pattern of Reporter^{GFP+} signalling observed in the posterior compartment was highly reproducible, suggesting that Cre transfer was directed, as opposed to random Cre protein diffusion. Furthermore, posterior clones were frequently observed at a distance from anterior/posterior boundary, creating a 'gap region,' suggesting the involvement of long-range cell communication mechanisms and the potential involvement of cellular extensions. Cytonemes are cellular extensions which extend from the posterior to the anterior compartment and have been shown to be involved in the formation of morphogen gradients and EV release [28, 36, 362]. Extensions have also been observed in the posterior to anterior direction to facilitate morphogen reception [29]. It is possible that cytonemes play a role in the transfer of Cre across the anterior/posterior boundary; a hypothesis not incompatible with mechanisms of EV transfer.

An alternative interpretation of the 'gap region' is the growth of the Reporter^{GFP+} signal away from the anterior/posterior boundary during wing disc growth. This is unlikely, however, due to the heterogeneous dynamics of IWD growth. Cell proliferation first occurs along the proximo-distal axis and only later occurs in the anterior-posterior direction (**Figure 4.20**) [351, 534, 535]. Even though signalling events occur early in development, posterior signal is still likely to have occurred at a distance from the anterior/posterior boundary for the 'gap region' to be so prevalent in the third instar larvae stage. Future work will focus on imaging imaginal wings discs at earlier points in development to determine the location of posterior Reporter^{GFP+} clones. This is experimentally difficult, however, due to the small and delicate nature of these tissues at earlier points in development.

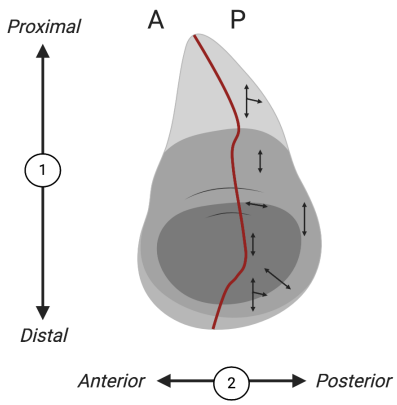


Figure 4.20: Schematic diagram depicting direction of growth during *Drosophila* imaginal wing disc development. Cell proliferation in the imaginal wing disc occurs primarily in the proximal-distal axis, with anterior/posterior (A/P) growth observed later in development (information extracted from Resino et al. (2002) and Tozluoglu et al. (2020) [351, 535]).

Additionally, the presence of the 'gap region' in the posterior could suggest that the cells adjacent to the anterior/posterior boundary are somewhat refractory to the signals originating from the anterior compartment. Experiments described in **Chapter 5**, however, show that when additional factors are disrupted, cells within this 'gap region' do commonly become Reporter^{GFP+}.

These results suggest that the Cre-LoxP reporter system may be reporting targeted transfer or uptake to cells with unique characteristics. Indeed, the presence of distinct differentiated cell populations in the IWD has been proposed [535].

In the future, gene expression patterns in posterior Reporter^{GFP+} cells could be compared to those observed in posterior Reporter^{GFP-} cells by fluorescence-activated cell sorting (FACS) dissociated cells prior to -omic analysis. This may identify potential factors involved in EV uptake and highlight the ability of the Cre-LoxP reporter system to identify novel sub-compartments of the imaginal wing disc.

4.5.2 Comparison of the *Drosophila* Cre-LoxP reporter system with other reporters of EV-mediated transfer

Various methods to study EV-loading and transfer have been developed for use in cell culture and *in vivo* models (introduced in **Section 1.2.8**). A major step forward in the EV field was the establishment of the *in vivo* Cre-LoxP mouse models, which were reporters of EV-mediated communication [309–312]. These models allowed the functional transfer of EV cargo to be identified and allowed cells which have taken up EVs to be differentiated from cells which have not, facilitating studies into the physiological effects of EV uptake. Transferring this Cre-LoxP system to *Drosophila* provided additional benefits, overcoming the genetic constraints of mouse models and harnessing the short generation time, as well as the extensively studied model tissues, to further advance the EV field.

It is possible to compare the sensitivity of the *Drosophila* Cre-LoxP reporter system with the mouse models used in a range of cellular scenarios. Ridder et al. (2014) described that recombination events were rare in Purkinje neurons (present within the brain) of healthy animals when the cerebellum was injected with Cre containing EVs. Upon the induction of chronic inflammation, however, the prevalence of reporter positive Purkinje neurons increased to an average of 5.8% per animal [310]. Das et al. (2020) reported similar recombination frequencies in cardiomyocytes ($2.41 \pm 0.37\%$), highlighting increased efficiencies upon the induction of stress via myocardial infarction ($4.29 \pm 0.79\%$) [308]. For Ridder et al. (2015), the percentage of recombined positive leukocytes present within a Cre producing tumour ranged from 0.1% to 2.2%, with a tendency for more reporter positive cells to be present in larger tumours. In addition, no distant recombination events were

observed [309]. Similar observations were made by Zomer at al. (2015), who reported approximately 0.6% positive reporter cells when studying local communication and 0.3% when studying distant transfer, adding that the EV transfer of *cre* was more efficient *in vivo* than in cell culture [312]. Using the *Drosophila* Cre-LoxP system, however, Reporter^{GFP+} posterior cells were identified in approximately 25 to 35% of wing discs analysed and 9 to 25% of eye/antennal discs analysed. Taken together, these studies show that the efficiency of reporter systems to detect functional transfer of EV cargo is low; however, the *Drosophila* Cre-LoxP system appears to show a higher efficiency than any other system previously described.

An additional strength of the *Drosophila* Cre-LoxP reporter system, when compared to the Cre-LoxP mouse models, is that it does not require the injection of purified EVs or the implantation of transgenic cells. Within the same organism, and even within the same tissue, Cre expression can be driven in a particular subset of cells whilst all other cells present a reporter background. The *Drosophila* Cre-LoxP system does not require the manipulation of EV production and does not require the induction of cancer, injury or inflammation to increase signalling events. This allows experiments to be carried out without the influence of EV isolation methods or disrupted molecular pathways (induced experimentally or as a result of tumour transplantation). It is hoped this will improve the reproducibility of results and allow EV-biology to be studied under more homeostatic conditions.

4.5.3 The mammalian 25-nt EV-loading motif did not influence *cre* mRNA loading in *Drosophila*

The mammalian 25-nucleotide motif, shown to be enriched in the microvesicle-loaded mRNAs cells and effective in enhancing mRNA EV-loading [248], did not alter *cre* mRNA EV-loading in the *Drosophila* Cre-LoxP system; reflected in cell culture, qPCR and *in vivo* analyses. Indeed, even in mammalian systems, the effectiveness of this motif has varied [536].

In mammalian cell culture, mRNA loading was enhanced when constructs containing the 25-nt motif were co-transfected with pre-miR-1289 [248]. Bioinformatic

analysis confirmed that no miRNA in the *Drosophila* miRNA database (r6.26) contained the entire pre-miR-1289 sequence; with dme-miR-2491 identified as the closest relative (**Supplementary Figure 7.19**). It is possible that this miRNA is not present in *Drosophila*, thus explaining the lack of biological influence of the 25-nt motif in particular cellular scenarios. It would be interesting to determine if *cre* mRNA EV-loading could be enhanced in *Drosophila* cell culture by co-transfecting cells with the Cre constructs containing the 25-nt motif (Cre-V5-25) and with a pre-miR-1289 construct.

An additional factor which may contribute to the influence of the 25-nt motif is the tissue-specific expression of RNA-binding proteins (RBPs). RBPs can alter RNA localisation, alternative splicing, stability, degradation or gene expression. Furthermore, they can be expressed in cell-specific patterns [537], indeed RNA EV-loading can also be highly tissue specific [246]. The RBP profile may differ between cell types and organisms, meaning *cre* mRNA localisation or stability may not be comparable.

4.6 Future work

4.6.1 Further validation of *cre* mRNA EV-loading and transfer

Investigations so far have successfully shown that a proportion of *cre* mRNA was protected from the RNase treatment of EVs, suggesting that this *cre* mRNA was likely present within the EV lumen. To fully confirm this, an additional Proteinase K treatment is required. In the presence of Proteinase K and RNase together, extracellular RNAs not contained within EVs but protected from degradation via extracellular proteins, can de degraded [443]. This would provide further confirmation that the *Drosophila* Cre-LoxP system is applicable to study RNA EV-loading mechanisms and long-range transfer.

An additional Cre RNAi experiment has been designed to determine the significance of *cre* mRNA versus Cre protein transfer (**Figure 4.21**). Using the LexA-LexAop binary expression system, manipulations of gene expression can be performed in target cells, whilst simultaneously using the UAS-Gal4 system to drive the Cre-LoxP reporter expression [538, 539]. Using a Hh LexA driver, LexAop Cre RNAi expression can be driven exclusively in the IWD posterior, in combination with the ptcGal4 driver to facilitate UAS Cre expression in the IWD anterior. If only *cre* mRNA transfer occurs, the presence of Cre RNAi in recipient cells will prevent Cre protein from being produced and Cre-mediated recombination to induce Reporter^{GFP+} signalling will not occur. If Cre protein is transferred, however, the Cre RNAi should not disrupt the appearance of posterior Reporter^{GFP+} signal. Although studies in the Cre-LoxP system developed in mice did show EV *cre* mRNA transfer and confirmed that Cre protein was undetectable within EVs via Western Blotting or ELISA [309, 311, 312], this did not rule out EV-mediated Cre protein transfer at very low levels. This experiment has the potential to answer this fundamental question in a complex *in vivo* context.

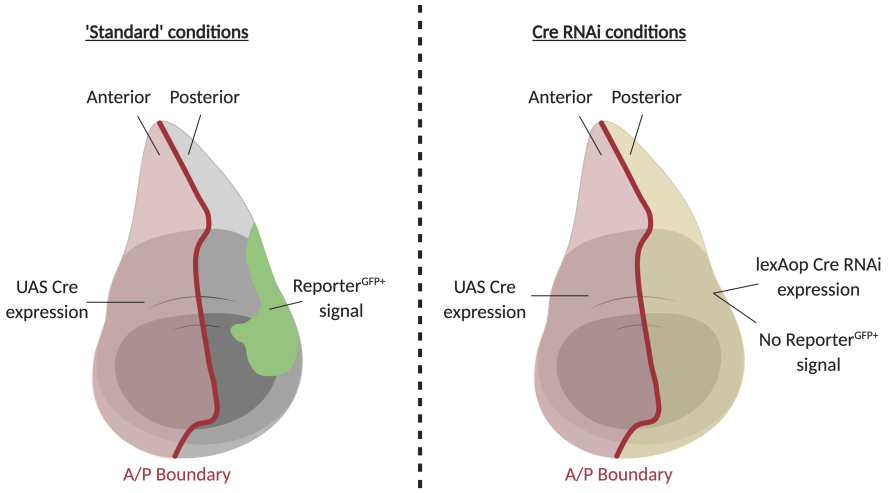


Figure 4.21: Cre RNAi *in vivo* experimental design. Under 'standard' experimental conditions, UAS Cre expression is limited to the anterior side of the IWD (using the ptc Gal4 driver), primarily in a strip of cells along the anterior/posterior (A/P) boundary. Under these conditions, Reporter^{GFP+} cells can be observed in the posterior side. Under the Cre RNAi conditions, the simultaneous expression of LexAop Cre RNAi on the posterior side can be achieved using the Hh LexA driver. If *cre* mRNA transfer is solely responsible for the posterior Reporter^{GFP+} signal, the presence of Cre RNAi in the posterior will prevent the translation of this *cre* mRNA into Cre protein and therefore, no posterior Reporter^{GFP+} signal will be observed. However, if Cre protein transfer occurs directly, posterior Reporter^{GFP+} signal will be observed.

Further work is also required to rule out the influence of other long-range communication mechanisms which may, at least partially, be response for distant Reporter^{GFP+} signal. In cell culture, free Cre transfer in cell media must be ruled out by EV-depleting the conditioned media from Cre producing cells via ultracentrifugation. The EV-depleted media can then be transferred to reporter cells and if Reporter^{GFP+} cells are no longer observed, free media Cre transfer is unlikely to be responsible for the Reporter^{GFP+} signal.

To gain insight into the influence of EV-mediated cell communication on this Cre-LoxP reporter system, attempts were made *in vivo* to knock-down factors known to influence EV release (such as Hrs and Vps4), using RNAi. These attempts we unsuccessful, however, as alterations in the prevalence of reporter signalling were also observed during control RNAi experiments. This was perhaps due to the potential overlap between RNAi and EV-loading machinery, for example, Argonaute and Dicer have been identified within EVs in certain contexts [334].

Further experiments are also required to determine the impact of cellular projections such as cytonemes on the Cre-LoxP reporter system, especially *in vivo*. Inhibiting and stabilising cytonemes is possible by manipulating Interference hedgehog (Ihog) expression [29, 540] or by manipulating gene expression of diaphanous (dia), shibire (shi), neuroglian (nrg), and capricious (caps) [35]. If these manipulations have a detrimental effect on Cre-LoxP reporter signalling, cytonemes may influence the cell communication detected by this system.

4.7 Chapter Summary

This chapter described the successful implementation of a Cre-LoxP reporter system of long-range cell communication in the *Drosophila* model. This *Drosophila* system builds upon previous models established in mice, removing the need for EV injection or cell transplantation to study EV-transfer in complex tissues. Cell culture qPCR experiments suggested that *cre* mRNA transfer occurs within EVs; however, further experiments are required to rule out the transfer of Cre protein or the influence of cellular projections on this reporter system. This tool was successfully used to show long-range communication within and between tissues in complex cellular environments *in vivo*. The function of a mammalian 25-nucleotide motif to encourage mRNA EV-loading may not be effective in *Drosophila* and this motif may function in a context specific manor.

Harnessing this *Drosophila* Cre-LoxP system to investigate the biological influence of EV-enriched *Drosophila* mRNA features (identified in **Chapter 3**) may highlight their effect on modulating mRNA EV incorporation. Such investigations may advance the current understanding of EV-biogenesis, loading and release mechanisms and identify novel targets to improve the EV incorporation and targeting of customised therapeutic cargo. This will be the focus of the research described in **Chapter 5**.

5

Identifying novel factors involved in modulating long-range communication

Contents

5.1 Introduction 160
5.2 Adapting the Cre-LoxP system to validate novel *Drosophila* EV motifs 160
5.2.1 Determining the effect of the *Drosophila* EV-enriched motifs in a cell culture based Cre-LoxP system 162
5.2.2 Determining the effect of the *Drosophila* EV-enriched motifs in the *in vivo* based Cre-LoxP system 164
5.3 Investigating the impact of caspase activity on the Cre-LoxP reporter system *in vivo* 167
5.3.1 High levels of Cre expression increased caspase activity and enhanced reporter signalling *in vivo* 168
5.3.2 Enhanced reporter signalling *in vivo* is likely not the result of cell debris release 169
5.3.3 Inhibiting effector caspases modulated the activity of the Cre-LoxP reporter system *in vivo* 170
5.4 Discussion 173
5.4.1 Cell culture and *in vivo* analysis of the *Drosophila* EV-enriched motif combination 173
5.4.2 Using the *Drosophila* Cre-LoxP system to identify novel factors involved in EV-loading and secretion 175
5.5 Future work 178
5.5.1 Using the *Drosophila* Cre-LoxP system to identify novel motif interacting factors 178
5.5.2 Determining the effect of mRNA stability on EV-loading 178
5.5.3 Using the *Drosophila* Cre-LoxP system to identify novel proteins involved in EV uptake 179
5.5.4 Using the *Drosophila* Cre-LoxP system to study EV-mediated communication in a disease context 180
5.6 Chapter Summary 181

5.1 Introduction

The bioinformatic analysis described in **Chapter 3** identified features present in *Drosophila* EV-enriched mRNAs which may enhance their EV-loading. Utilising the *Drosophila* Cre-LoxP reporter system described in **Chapter 4**, the biological significance of these features to enhance mRNA EV incorporation and transfer *in vivo* could be assessed. Beyond this, the *Drosophila* Cre-LoxP system also provided a novel tool to begin to investigate the fundamental mechanisms of cell communication. Combining this tool with inhibitors of important cellular pathways may highlight new molecular links and identify novel factors important for EV-biogenesis.

5.2 Adapting the Cre-LoxP system to validate novel *Drosophila* EV motifs

Four novel motifs (Motif A, B, C and D), enriched in EV-loaded mRNAs, were identified via the bioinformatic analysis of *Drosophila* S2R+ cellular and EV RNA content (described in **Chapter 3**). To determine if these motifs played a biological role in EV cargo loading, an mRNA 3'UTR which contained all four novel motifs was identified. This mRNA was *paralytic* (*para*, FBgn0285944, FBtr0344944), whose 3'UTR length (185 base pairs) was appropriate for easy manipulation and gene synthesis. Ubiquitin and UAS constructs containing Cre-V5 and the wild-type *paralytic* 3'UTR were generated (Cre-V5-paraWT). Also generated were Cre-V5-para constructs where various combinations of the A, C and D motifs were shuffled: A shuffled, C shuffled, A C shuffled and A C D shuffled (**Figure 5.1**). Shuffling consisted of the random re-ordering of bases within each motif to disrupt the primary, and subsequently the secondary, structure of the motif. Attempts were made to generate B shuffled versions of this construct, however, the molecular biology processes were not successful.

The predicted secondary structure of the wild-type and shuffled versions of the *paralytic* 3'UTRs was analysed using RNAfold. This analysis showed that the secondary structure of the *paralytic* 3'UTR was altered when specific motif regions

were shuffled (**Figure 5.2**). The effects of shuffling a specific motif were not fully autonomous, as structural changes throughout the entire 3'UTR were observed. For example, shuffling Motif C influenced not only its own structure, but also the structure of motif A, B and D (**Figure 5.2v**). By comparing the functionality of the Cre-LoxP reporter system when using these adapted Cre constructs, it may be possible to identify the biological influence of each motif, highlighting key mRNA sequence and structural features vital for modulating EV-mediated communication.

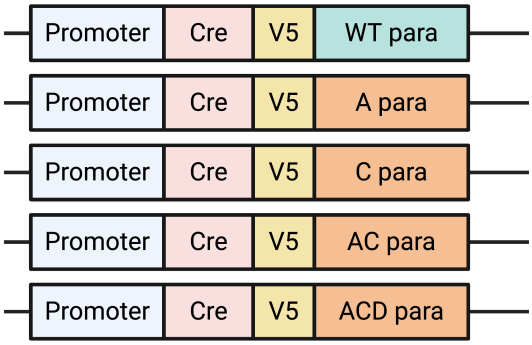


Figure 5.1: Cre-V5 constructs with the additional wild-type (WT), A shuffled, C shuffled, A C shuffled or A C D shuffled *paralytic* 3'UTR. Versions of these constructs containing either the Ubiquitin or UAS promoter were generated.

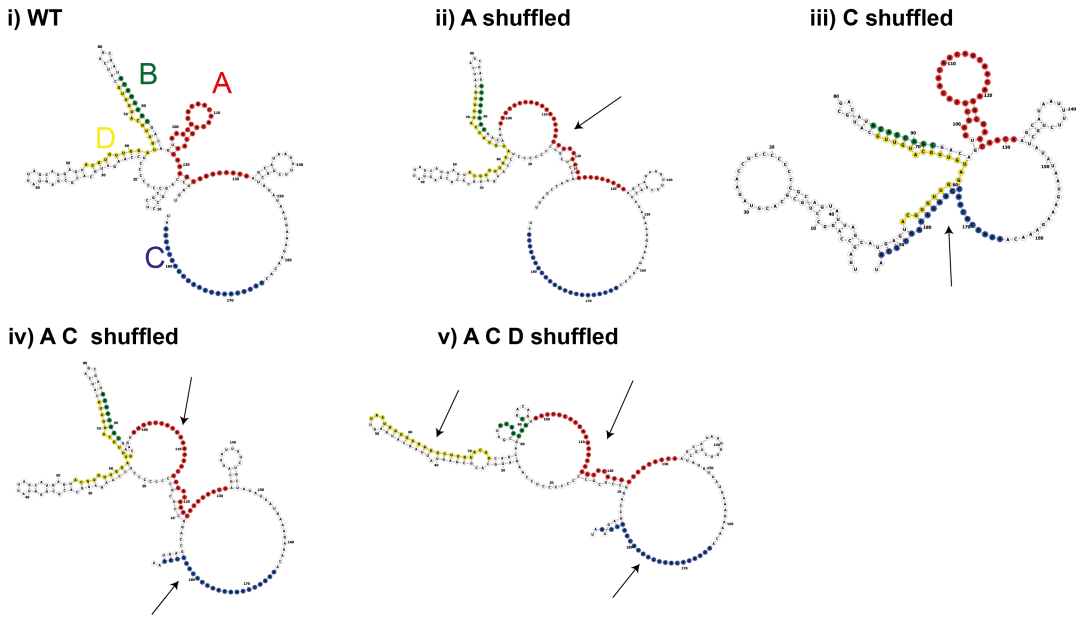


Figure 5.2: Predicted secondary structure of the wild-type (WT) and shuffled versions of the *paralytic* 3'UTR. Arrows highlight the shuffled motifs within each construct.

5.2.1 Determining the effect of the *Drosophila* EV-enriched motifs in a cell culture based Cre-LoxP system

To determine the biological significance of the novel *Drosophila* EV-enriched ABCD motif combination, adapted Cre-V5 constructs containing wild-type or shuffled versions of the *paralytic* 3'UTR were utilised within the Cre-LoxP reporter system. *Drosophila* S2R+ cells were transfected with these various Cre constructs and qPCR experiments were carried out to assess if the addition of the *para* WT 3'UTR to *cre* mRNA could enhance mRNA EV incorporation.[36] For this analysis, *rpl32* was used as the housekeeping gene due to the high and relatively consistent cellular and EV abundance levels observed in the RNA-seq analysis described in **Chapter 3** (Cell FPKM = 5972.5 ± 647.1, EV FPKM = 4429.6 ± 326.3, Mean ± standard deviation).

Encouragingly, a 5.3 fold increase in EV *cre* mRNA levels was observed in the Cre-V5-paraWT condition relative to the parental cell, whereas a 2.0 and 1.6 fold change was observed in the Cre-V5 and Cre-V5-25 conditions, respectively (**Figure 5.3**). This enhanced EV *cre* mRNA level was partially diminished upon shuffling the A, C and D motifs, as only a 3.99 fold increase was observed in the Cre-V5-paraACD condition. Whilst cellular *cre* mRNA levels remained consistent for Cre-V5 and Cre-V5-25 conditions, a 2.3 and 3.6 fold increase in was observed in the Cre-V5-paraWT and Cre-V5-paraACD conditions (relative to Cre-V5), respectively (**Supplementary Figure 7.20**). Due to the high level of variability observed between replicates of this cell culture based experiment, however, no significant differences were detected when comparing the various Ubi Cre constructs. Further replicates are now required to confirm the trends observed.

[36] by Elizabeth Dellar

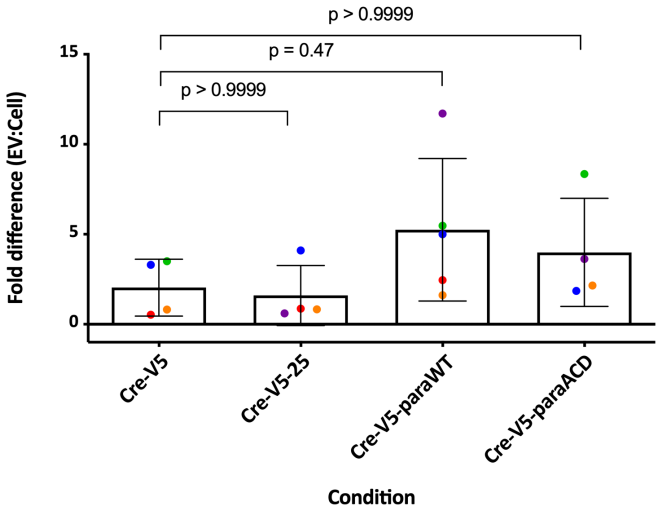


Figure 5.3: Fold differences observed in *cre* mRNA cellular and EV abundances for Cre-V5, Cre-V5-25, Cre-V5-paraWT and Cre-V5-paraACD conditions, determined via qPCR. S2R+ cells were transfected with Ubi Cre constructs and media was conditioned for 48 hours. Colours indicate individual biological replicates. Mean ± standard deviation is shown. Statistical analysis was carried out via a Dunn's multiple comparisons test. Delta Ct values were determined using *cre* as the gene of interest and *rpl32* as the housekeeping gene. Fold change (Delta delta Ct) values were then determined by normalising each EV Delta Ct value to the matched parental cell Delta Ct value.[37]

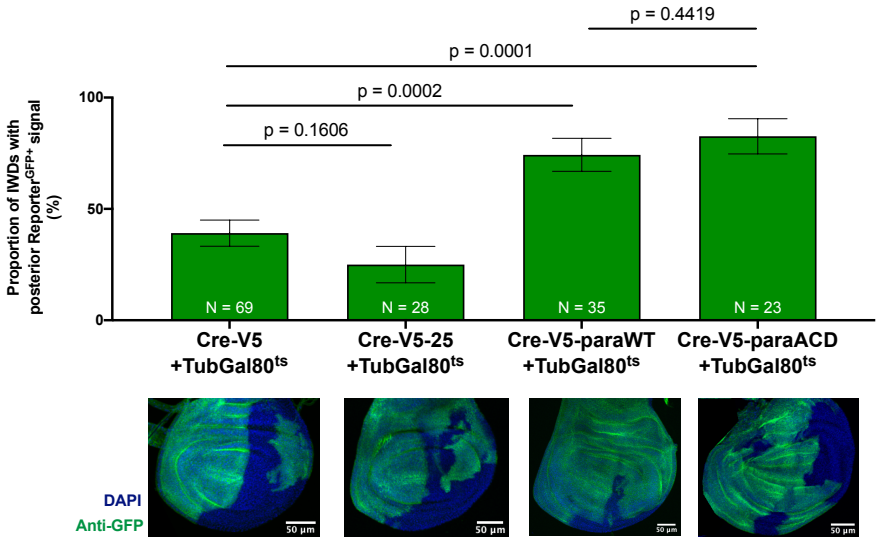
The Cre-LoxP system was then harnessed to assess if effects on reporter signalling could be observed when using various versions of this Cre construct. Initially a co-culture assay was harnessed, where Cre transfected cells were suspended in a Transwell insert over reporter transfected cells, allowing EV and free media transfer without allowing direct cell fusion. Levels of Reporter^{GFP+} signalling, however, were not significantly enhanced relative to the control transfection condition (**Supplementary Figure 7.21**), suggesting this assay was not sensitive to detect changes as a result of adapted mRNA sequences. This lack of sensitivity may be due to the dependency on two cell transfections, reducing the total cell number of Cre and Reporter cells available for the successful detection of Cre transfer. To overcome these constraints, future work harnessed the *Drosophila in vivo* Cre-LoxP system.

[37]Data generated by Elizabeth Dellar

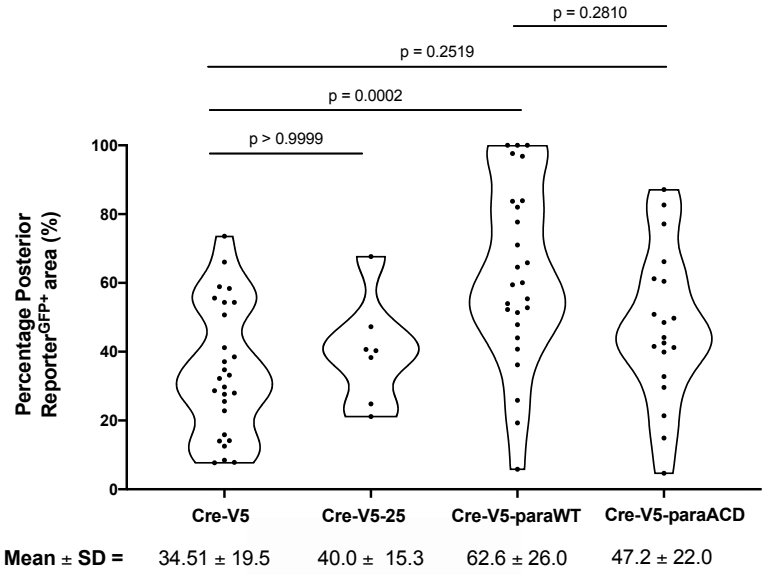
5.2.2 Determining the effect of the *Drosophila* EV-enriched motifs in the *in vivo* based Cre-LoxP system

Encouraged by the trends observed when assessing mRNA EV-loading in the cell culture system upon the incorporation of the *Drosophila* EV-enriched motifs, as well as to overcome constraints of the cell culture system, the *in vivo* Cre-LoxP system was adapted to study the biological effect of the *para* WT or shuffled 3'UTR sequences. Using the Cre-V5-paraWT construct, the IWD experiments described in **Chapter 4** were repeated. A significantly higher proportion of IWDs presenting Reporter^{GFP+} posterior signal was observed in the Cre-V5-paraWT condition, when compared to the Cre-V5 condition (Cre-V5: 39.1 ± 5.9%, Cre-V5-paraWT: 74.3 ± 7.4%, p = 0.0002, **Figure 5.4i**). The mean percentage area of this posterior Reporter^{GFP+} signal also significantly increased (Cre-V5: 34.5 ± 19.5%, Cre-V5-paraWT: 62.6 ± 26.0%, p = 0.0002, **Figure 5.4ii**), with a tendency for more posterior Reporter^{GFP+} signals to infiltrate the 'gap region' (the posterior area most adjacent to the anterior/posterior (A/P) boundary which often lacks Reporter^{GFP+} signal, **Figure 5.5** and **Supplementary Figure 7.17**).

Investigations to assess the effect of shuffling specific motif combinations were carried out using the Cre-V5-paraACD construct, where motifs A, C and D were shuffled. A significant increase in the frequency of posterior Reporter^{GFP+} signal, and an infiltration of the 'gap region', was still observed in the Cre-V5-paraACD condition compared to Cre-V5 (**Figure 5.4i** and **Supplementary Figure 7.17**). Intriguingly, however, the area of the posterior becoming Reporter^{GFP+} tended to decrease upon shuffling the ACD motifs (47.2 ± 22.0%) compared to the Cre-V5-paraWT condition (62.6 ± 26.0%, **Figure 5.4ii**); however, this decrease was not significant (p = 0.2810). These results mirrored those observed in cell culture, where mRNA EV incorporation tended to decrease upon ACD motif shuffling (**Figure 5.3**). These results suggested that the shuffling of the ACD motif combination may partially hinder the enhanced long-range mRNA transfer achieved using the *para* WT 3'UTR; however, as the effects observed were not significant, these motifs may not be the major contributing factor to the enhanced reporter signalling observed.



(i) Proportion of IWD posteriors displaying Reporter^{GFP+} signal. Also shown are projected immunofluorescence confocal images depicting Reporter^{GFP+} signal (Green) and DAPI (Blue) for a representative positive example. Statistical analysis was carried out as described in **Section 2.5.1**.



(ii) Percentage area of the posterior becoming Reporter^{GFP+}. Statistical analysis was carried out via a Kruskal-Wallis multiple comparisons test.

Figure 5.4: Frequency and percentage area of posterior Reporter^{GFP+} events under Cre-V5, Cre-V5-25, Cre-V5-paraWT and Cre-V5-paraACD conditions. Reporter^{GFP+} posterior signal significantly increased in frequency and area upon the incorporation of the *para* 3'UTR (Cre-V5-paraWT); however, this increase was not significantly altered upon shuffling the A, C and D EV motifs (Cre-V5-paraACD). Genotype: UAS Cre/+; ptcGal4/Actin-loxP-myrGFP; TubGal80^{ts}, UAS HisRFP/MKRS.

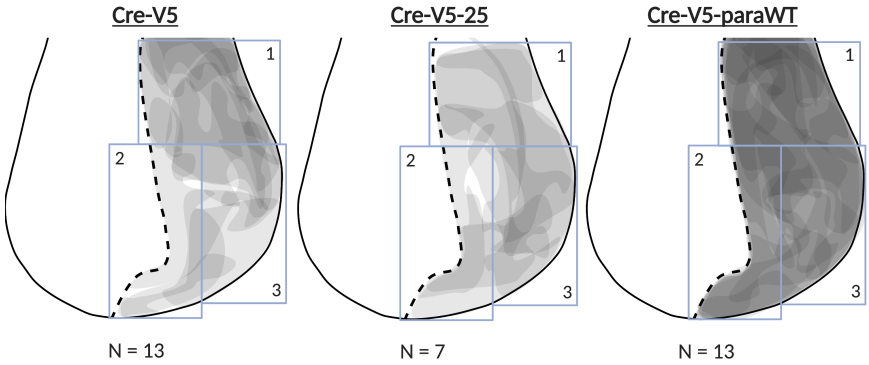


Figure 5.5: Schematic diagram depicting the posterior regions becoming Reporter^{GFP+} under Cre-V5, Cre-V5-25 and Cre-V5-paraWT conditions. Note that the area marked by box 2, which normally contains the 'gap region,' becomes more frequently infiltrated in the Cre-V5-paraWT condition. Genotype: UAS Cre/+; ptcGal4/Actin-loxP-myrGFP; TubGal80^{ts}, UAS HisRFP/MKRS.

Future work is needed to further analyse the effect of other shuffled motif combinations, to determine their influence on mRNA EV-loading. Future work is also needed to assess the potential contribution of additional factors potentially altered upon the addition of the *para* 3'UTR, such as mRNA stability, GC content and mRNA length.

5.3 Investigating the impact of caspase activity on the Cre-LoxP reporter system *in vivo*

Beyond utilising the *in vivo Drosophila* Cre-LoxP system to biologically validate the influence of mRNA features on EV incorporation, the potential of this tool to assess other mechanisms involved in EV biogenesis and loading was realised. In recent years, one such mechanism, the influence of non-apoptotic caspase activity on the loading, release and uptake of EVs, has been highlighted (introduced in **Section 1.2.3**). The true complexity of caspase activity and its influence on EV biology has not yet been fully realised. By combining the *Drosophila* Cre-LoxP system with tools to modulate the activity of specific components of the caspase pathway (**Figure 5.6**), the influence of these components on reporter signalling could be assessed. This had the potential to highlight possible functions for these proteins in modulating long-range communication in complex tissues.

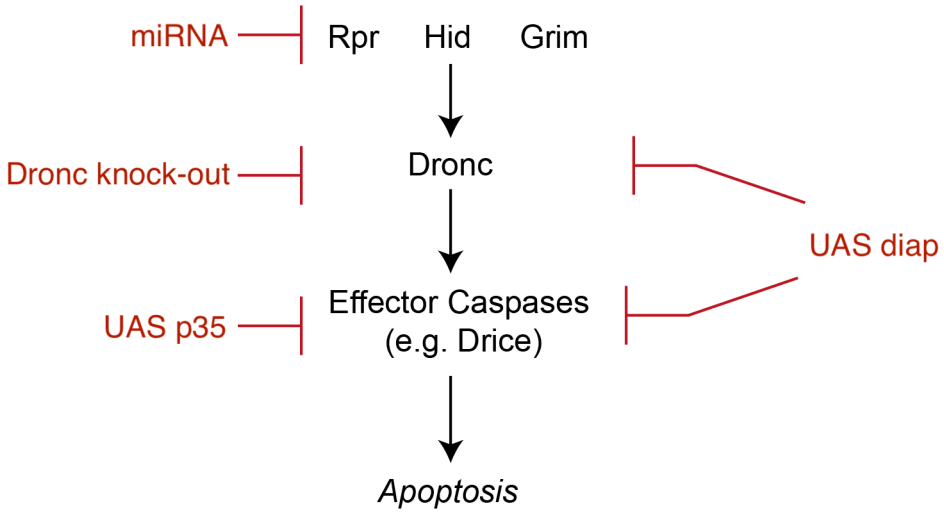


Figure 5.6: Caspase cascade and the point at which UAS p35, UAS diap, Dronc knock-out and miRNAs against reaper (Rpr), Hid and Grim act to disrupt the pathway.

5.3.1 High levels of Cre expression increased caspase activity and enhanced reporter signalling *in vivo*

In line with previous studies which have reported Cre toxicity in *Drosophila* [367, 510], high Cre expression levels (in the absence of the Gal4 suppresser, Gal^{80}) resulted in significantly increased cell death, shown via increased anti-DPC1 immunostaining ($p = 0.0002$, **Figure 5.7**). The inclusion of a thermosensitive $TubGal80^{ts}$ allowed Cre expression to be finely controlled, with Cre toxicity prevented when experiments were carried out at 25°C, bringing cell death levels within a normal range.

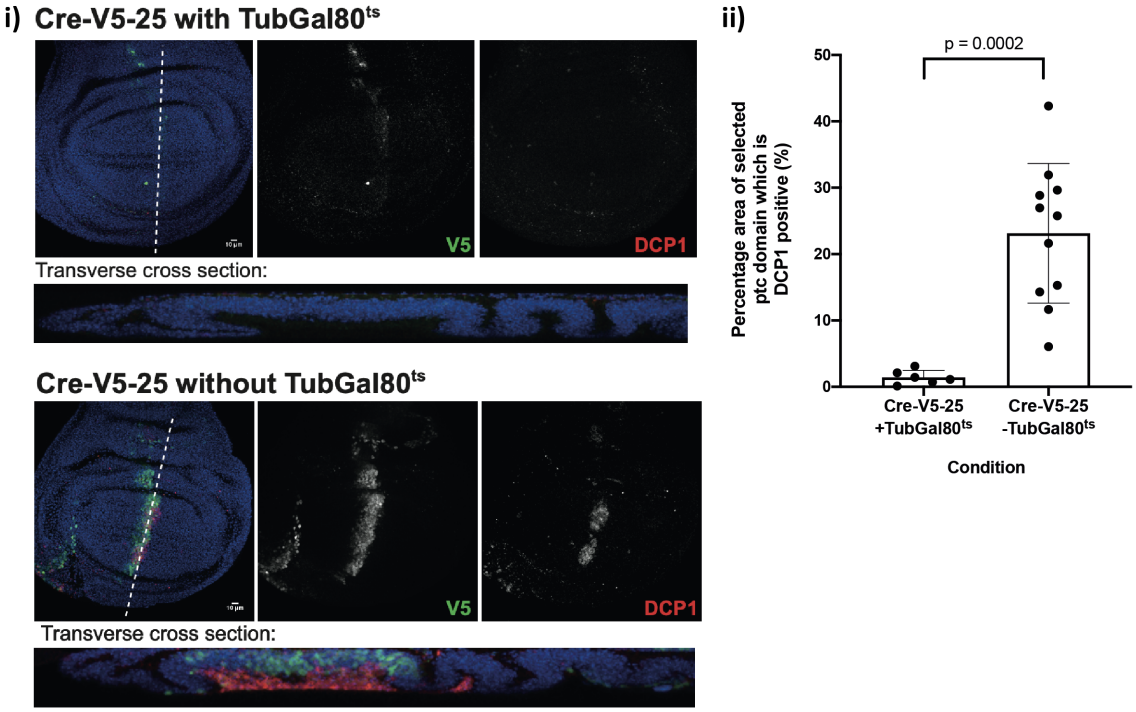


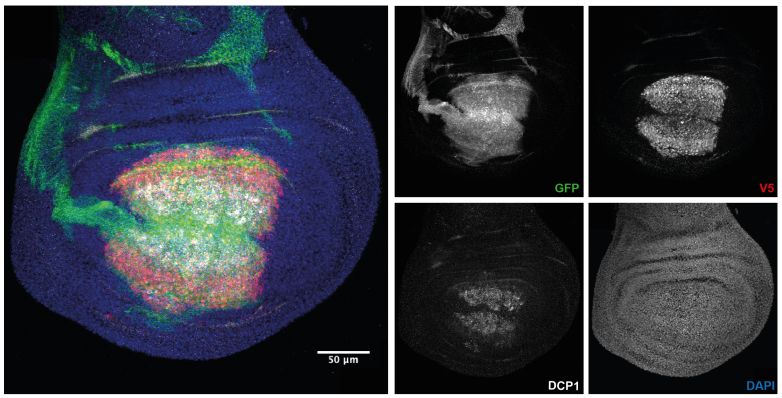
Figure 5.7: (i) Cre expression levels are low in the IWD using the ptcGal4 driver in the presence of the Gal4 suppressor, $TubGal80^{ts}$. Cre toxicity is observed in the absence of the $TubGal80^{ts}$, with transverse cross sections highlighting a basal accumulation of dead cells when Cre expression is high. Cell death was detected via anti-DCP1 immunostaining. Cre expression levels determined via anti-V5 immunostaining. The white dashed lines indicate the area used for the transverse cross section. (ii) Quantification of cell death levels (DCP1 staining) in the presence or absence of $TubGal80^{ts}$. Statistical analysis was carried out via a two-tailed Mann-Whitney test.

As previously shown in **Chapter 4**, a relatively low proportion of IWDs presented posterior Reporter^{GFP+} signal when TubGal80^{ts} was present (Cre-V5 = 39.1 ± 5.9%; Cre-V5-25 = 25 ± 8.2%, **Figure 4.16**). When TubGal80^{ts} was absent, the proportion of IWDs showing posterior signal significantly increased in each condition (Cre-V5 = 80 ± 12.7% (p = 0.0034); Cre-V5-25 = 71.4 ± 9.9% (p = 0.0003), **Figure 4.16**). No significant difference in the mean posterior Reporter^{GFP+} area was observed (**Figure 4.17**). Further experiments were required, however, to determine if the increased frequency of posterior Reporter^{GFP+} signal was a result of increased cell debris release (as a result of cell death), increased Cre levels or increased caspase activity enhancing EV release.

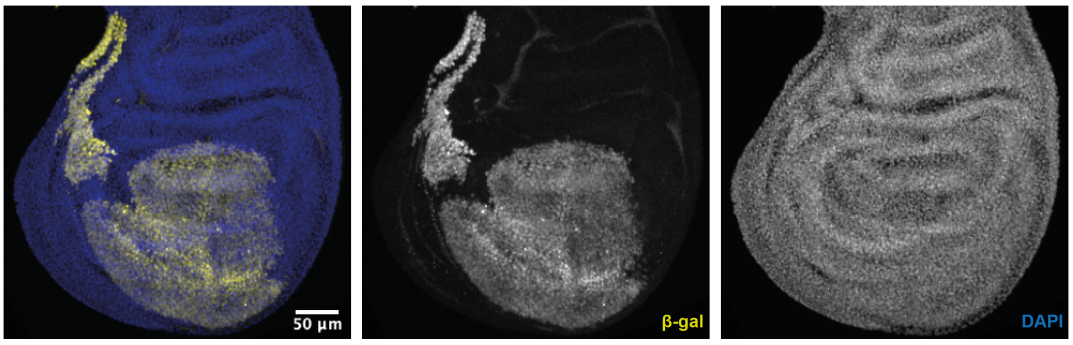
5.3.2 Enhanced reporter signalling *in vivo* is likely not the result of cell debris release

Using the spaltGal4 driver, Cre expression was driven within the IWD wing pouch, with results showing that no Reporter^{GFP+} signal was observed outside the spaltGal4 domain (**Figure 5.8**). SpaltGal4 drives expression from early third instar larval stages [541] and given that the Cre-LoxP system detected long-range signalling events occurring during early developmental stages only (**Section 4.4.2**), it is possible that by using spaltGal4, this 'active window' of Cre transfer was missed.

In the absence of TubGal^{80} whilst using spaltGal4, high levels Cre expression in the wing pouch resulted in high levels of DCP1 staining, indicative of increased cell death. Despite this enhanced cell death, however, long-range Reporter^{GFP+} signalling was not observed, suggesting that the enhanced Reporter^{GFP+} signalling previously observed when using the ptcGal4 driver in the absence of TubGal^{80} (**Figure 4.16**) was not the result of cell debris uptake. This result suggested that this enhanced Reporter^{GFP+} may be a result of increased *cre* expression levels or increased caspase activity modulating EV cargo loading and/or release.



(i) No Reporter^{GFP+} signal was observed beyond the spaltGal4 domain when Cre was expressed within the IWD wing pouch, despite the high levels of cell death, observed via DCP1 positive cells (N = 11). Genotype: UAS Cre-V5-25/+; Actin-loxP-myrGFP/spaltGal4; TM2/+).



(ii) LacZ lineage tracing to highlight the expression domain of the spaltGal4 driver. Genotype: UAS Flp/spaltGal4; Actin-FRT-stop-FRT-lacZ/TM2.

Figure 5.8: Long-range Reporter^{GFP+} signalling was not observed when Cre was expressed within the IWD wing pouch using the spaltGal4 driver.

5.3.3 Inhibiting effector caspases modulated the activity of the Cre-LoxP reporter system *in vivo*

To further assess the impact of caspase activity on *Drosophila* Cre-LoxP reporter signalling, initial investigations focused on inhibiting downstream effector caspases using UAS p35. Experiments described in **Section 4.4.1**, where Cre was expressed in the IWD anterior via the ptcGal4 driver, were repeated whilst simultaneously expressing UAS p35. Under these experimental conditions, effector caspase activity was inhibited solely within Cre-producing cells (**Figure 5.9**). When effector caspases were inhibited, the proportion of IWDs presenting posterior Reporter^{GFP+} signal significantly increased from 25 ± 8.2% in the absence of p35 to 72.7 ± 13.4% in the presence of p35 ($p = 0.0024$, **Figure 5.10**). No significant difference in the

mean posterior area becoming Reporter^{GFP+} was observed upon the addition of p35 (**Supplementary Figure 7.22**). These results suggested that effector caspases may play a role in modulating long-range cell communication mechanisms.

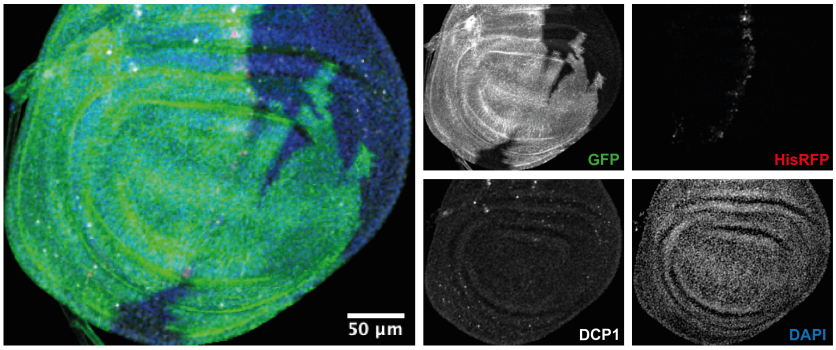


Figure 5.9: Posterior Reporter^{GFP+} signalling observed when UAS Cre-V5-25 and UAS p35 were co-expressed in the IWD anterior using the ptcGal4 driver. Genotype: UAS Cre-V5-25/+; Actin-loxP-myrGFP/ptcGal4; TubGal80^{ts}, UAS HisRFP/UAS p35. An immunofluorescence image of the control condition, in the absence of UAS p35, can be seen in **Figure 4.13**.

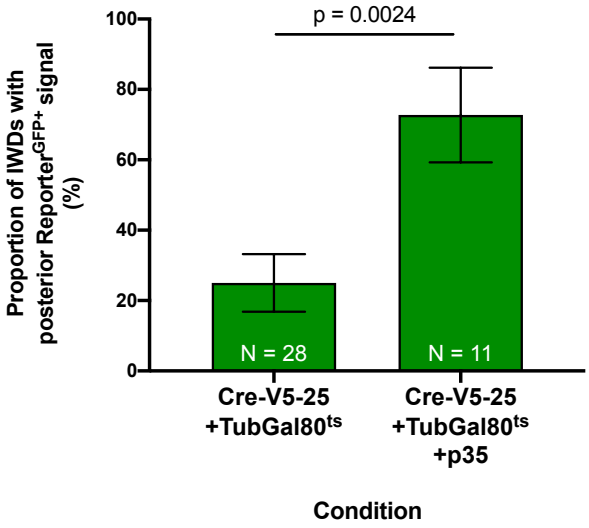


Figure 5.10: Proportion of IWD posteriors showing Reporter^{GFP+} signal in the absence or presence of p35. Presence of p35 in anterior Cre producing cells (causing effector caspase inhibition) resulted in enhanced Cre-LoxP reporter signalling in the posterior. Percentage proportion ± SE is shown. Statistical analysis was carried out as described in **Section 2.5.1**.

Given that inhibiting effector caspases appeared to enhance long-range cell communication, further experiments were carried out to determine if simultaneous Cre expression and caspase inhibition during late larval development, where normally no posterior signal was observed, could enhance long-range *cre* transfer. Timeline experiments described in **Section 4.4.2** were repeated with the addition of UAS p35. Even when UAS Cre-V5-25 and UAS p35 were co-expressed in the IWD anterior during late larval development stages only, no posterior signal was observed (**Figure 5.11**, N=8). This experiment provided further evidence to suggest the appearance of posterior IWD clones was dependent on Cre being expressed during early larval development only. This long-range communication occurred within a strict developmental time window which could not be altered by disrupting effector caspase activity. Future work is now required to assess the impact of other components of the caspase pathway on EV-loading and transfer. These experiments will be described in **Section 5.5**.

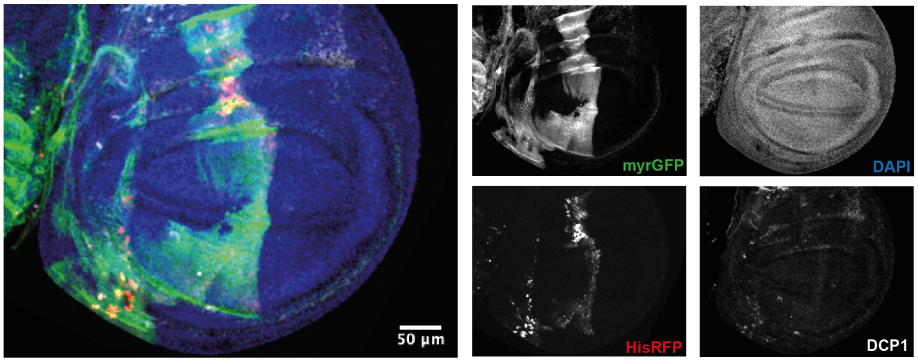


Figure 5.11: No posterior Reporter^{GFP+} signal was observed when UAS Cre-V5-25 and UAS p35 were co-expressed during late larval development stages only (N=8). Genotype: UAS Cre-V5-25/+; Actin-loxP-myrGFP/ptcGal4; TubGal80^{ts}, UAS HisRFP/UAS p35.

5.4 Discussion

5.4.1 Cell culture and *in vivo* analysis of the *Drosophila* EV-enriched motif combination

EV-enriched motifs have been identified in a number of mammalian systems [74, 222, 234, 245–249, 254], with a number of these motifs successfully shown to enhance RNA EV-loading in cell culture [74, 234, 246, 248, 249, 254]. The *Drosophila* Cre-LoxP system provided the unique opportunity to demonstrate the influence of potential EV-loading motifs in both cell culture and *in vivo* contexts. This chapter described the cell culture and *in vivo* assessment of the bioinformatically predicted EV-enriched 3'UTR mRNA motif combination identified in **Chapter 3**.

Influence of EV motifs on the EV incorporation of *cre* mRNA

As a starting point, qPCR based experiments were used to show that the addition of the *paralytic* 3'UTR, containing the *Drosophila* ABCD motif combination, tended to enhance *cre* mRNA EV incorporation. Whilst shuffling the A, C and D motifs within this 3'UTR tended to decrease *cre* EV mRNA levels, this decrease was not significant. This result suggested that the novel *Drosophila* EV motifs may contribute to mRNA EV incorporation, reflecting studies carried out in mammalian cell systems when assessing alternative EV-enriched 3'UTR motifs [246, 248]. These motifs, however, may only partially influence EV-loading, with further work needed to assess the influence of additional factors such as altered mRNA length, GC content or stability. Interestingly, no enhanced EV incorporation was observed upon the addition of the 25-nucleotide mammalian EV-loading motif [248] in *Drosophila*, suggesting this mammalian motif does not influence *Drosophila* mRNA EV-loading.

The *Drosophila* Cre-LoxP Transwell assay was not sensitive to detect changes in mRNA sequence

Following on from this qPCR analysis, a Transwell assay was utilised, where Cre and Reporter transfected cells were co-cultured, to determine if increased functional delivery of *cre* could be observed upon the addition of the *para* 3'UTR. It was

hypothesised that increased *cre* mRNA EV incorporation would result in increased long-range *cre* transfer and a higher proportion of Reporter^{GFP+} cells. Unfortunately, the *Drosophila* Cre-LoxP Transwell system was not sensitive to detect significant differences upon the addition of alternative 3'UTRs to the Cre constructs.

The low sensitivity of cell culture systems to detect EV transfer has recently been reported [318]. Limitations such as the small Transwell surface reducing the number of available donor cells or the blockage of some EV transfer due to the pore size, may reduce effectiveness [318]. Zomer at al. (2015) do report higher signalling levels using a mammalian Cre-LoxP Transwell system (with stable cell lines and a 7-day co-culture), ranging from approximately 0.4 to 5% and increasing to 10% if a 100:1 ratio of Cre to Reporter cells was used [312]. The low sensitivity of the *Drosophila* Cre-LoxP Transwell assay may be the result of using transient transfections, yielding a reduced number of Cre and Reporter cells; indeed, this assay may benefit from the development of stable cell lines. The use of a longer co-culture period may also enhance reporter signalling; however, cell viability is compromised, which may have detrimental effects on reporter assay function.

Modulating long-range mRNA transfer within complex tissues *in vivo* upon 3'UTR *cre* mRNA modifications

Using the *in vivo Drosophila* Cre-LoxP system, it was possible to assess the function of EV motifs in complex tissues, overcoming constraints of the cell culture system. A significant increase in the frequency and percentage area of the IWD posterior Reporter^{GFP+} signals was observed upon the addition of the *para* WT 3'UTR, containing the ABCD motif combination, to *cre* mRNA. Interestingly, simultaneously shuffling the A, C and D motifs within the *para* 3'UTR did not decrease the frequency of posterior Reporter^{GFP+} signal, however, the percentage area of the posterior becoming Reporter^{GFP+} did tend to decrease. These results provided the first *in vivo* evidence to suggest that mRNA 3'UTR modifications may functionally influence long-range mRNA transfer. The lack of significant changes upon shuffling the EV motifs within the 3'UTR, however, suggested that these motifs may only partially contribute to the EV incorporation of *cre* mRNA. Perhaps the addition

of the *para* 3'UTR also altered factors such as mRNA stability, GC content or length, which can also influence mRNA EV-loading.

An increase in mRNA stability upon the addition of the *para* 3'UTR is perhaps evidenced by the increased parental cell *cre* mRNA levels in the Cre-V5-paraWT and Cre-V5-paraACD conditions, observed via cell culture qPCR analysis. Improved stability may be the result the increased 3'UTR length or altered GC content upon the addition of the *para* 3'UTR. Future work is needed to determine the significance of these 3'UTR modifications on Cre-LoxP reporter signalling. This future work is highly relevant since alternative mRNA 3'UTRs have been shown to play roles in altering RNA stability [439, 454, 542, 543]. Improved stability may increase the potential for *cre* mRNA EV-loading and may influence its functional delivery by improving the likelihood for successful endosomal escape.

Future work described in **Section 3.5** aims to build an understanding of the molecular mechanisms involved the increased EV incorporation. The functional relevance of potential *trans*-acting factors, bioinformatically predicted in **Chapter 3**, can be validated *in vivo* by simultaneously inducing CRISPR-Cas9 mediated knock-out or mutations of the RBPs or miRNAs of interest in Cre producing cells. Excitingly, these analyses may identify novel targets to manipulate mRNA EV-loading in other organisms for therapeutic or research purposes.

5.4.2 Using the *Drosophila* Cre-LoxP system to identify novel factors involved in EV-loading and secretion

An exciting aspect of developing an *in vivo* Cre-LoxP reporter of cell communication in *Drosophila* was the possibility to investigate the influence of specific pathways on EV-loading and transfer. Whilst many studies have begun to elucidate key mechanisms of EV biogenesis and cargo loading (**Section 1.2.2** and **1.2.4**), the true complexity of these processes remains to be uncovered. By altering the expression of genes of interest in Cre producing cells, their involvement in long-range communication can be assessed. Moreover, gene expression can be specifically modulated in recipient cells, potentially allowing genes or proteins vital for EV

uptake and cargo processing to be identified. These are fundamental aspects of EV biology which are not yet fully characterised.

The caspase pathway in particular has garnered attention recently for its involvement in EV biogenesis, loading and uptake processes, not only in apoptotic conditions [150–153, 544], but also in non-apoptotic scenarios [157, 158, 165] (**Sections 1.2.2** and **1.2.3**). Initially the main function of these enzymes was believed to be solely related to cell death and apoptosis; however, recent research has implicated caspases in processes such as cell competition, stem cell function, cell communication, migration [166, 375, 545–547] and even in the cleavage of RBPs associated with mRNA EV-loading [255]. The involvement of caspase activity on EV cargo loading, release and uptake has not been fully explored, providing an interesting focus for downstream investigations using the *Drosophila* Cre-LoxP system.

Enhanced Cre-LoxP reporter signalling observed in the presence of increased caspase activity

Early experiments in this analysis highlighted that increased Cre expression levels resulted in a high degree of toxicity and caspase activity. Whilst Cre toxicity has been reported in other animal and plant models [367, 510], Cre toxicity in this system was circumvented by incorporating the Gal4 suppressor, Gal^{80}, to reduce Cre expression to non-toxic levels. Interestingly, high Cre expression levels in the anterior of the IWD, which caused increased caspase activity and cell death, resulted in a significant increase in the prevalence of posterior $Reporter^{GFP+}$ signal. When using the spaltGal4 driver, which drives Cre expression from early third instar larval stages [541], likely missing the 'active window' of transfer during early development, no remote IWD $Reporter^{GFP+}$ signal was observed, even when high Cre expression levels caused cell death in the wing pouch. This suggested that the enhanced $Reporter^{GFP+}$ signal was not the result of cell debris or apoptotic body release. Further analysis was required to determine if the increased $Reporter^{GFP+}$ signal was due to higher Cre expression levels or the result of increased caspase activity modulating EV-loading or secretion.

Inhibiting effector caspase activity resulted in enhanced Cre-LoxP reporter signalling

To further study the influence of caspase activity on Cre-LoxP reporter signalling, caspase activity was modulated in Cre producing cells, whilst maintaining Cre expression at non-toxic levels. Results showed that inhibiting downstream effector caspases (via p35 over-expression) significantly increased the prevalence of posterior Reporter^{GFP+} signals, perhaps by enhancing the level of *cre* cargo loading or increasing the level of EV-secretion. Effector caspases may play inhibitory roles and by removing their activity, EV release or mRNA loading was enhanced. This appears to contradict the results observed in toxic Cre conditions, however, where increased caspase activity increased signalling events. One alternative hypothesis is that upon effector caspase inhibition, upstream initiator caspases, such as Dronc (caspase-9), are more readily available to carry out EV-related functions. Indeed, it was an initiator caspase, caspase-8, which modulated EV-loading in the HTC116 colon cancer cell line investigated by Kim et al. (2017) [158].

The potential role for Dronc in promoting EV secretion may align with its role in compensatory proliferation, during which Dronc stimulates compensatory proliferation in neighbourings cells by activating morphogen or ROS release, for example, via p53 or JNK pathway activation [355, 361, 548, 549]. Indeed, EVs and caspase activity were together implicated in compensatory proliferation mechanisms identified by Gupta et al. (2017) [550]. Future work will focus on inhibiting the caspase pathway at various levels (Hid, Reaper and Grim or Dronc) to determine the level at which caspase activity influences EV-mediated cell communication. Additional experiments described within **Section 5.5** also aim to investigate the potential crosstalk between the caspase pathway, more particularly in a disease context, when both caspase activation and EV-mediated communication can become disrupted.

5.5 Future work

5.5.1 Using the *Drosophila* Cre-LoxP system to identify novel motif interacting factors

Complementary to the cell culture techniques described in **Section 3.5**, the *in vivo Drosophila* model could also be harnessed to identify factors interacting with *Drosophila* EV-enriched motifs. The RNA-protein interaction detection (RaPID) method uses a modified promiscuous biotin ligase derived from *Escherichia coli*, BirA*, to identify proteins within a 10 nm range of an RNA of interest [551, 552]. Transgenic *Drosophila* lines could be generated, each with the capacity to express RNAs with permutations of the EV-enriched motifs (described previously in this chapter) together with the 5' and 3' BoxB motifs to recruit BirA*. BirA* and the construct of interest can be driven in a tissue specific manner using the UAS/Gal4 system, leading to the biotinylation of RNA proximal proteins, aiding downstream analysis via mass spectrometry or western blot. Motif interacting protein profiles could be compared between cell types, such as the IWD epithelial cells or hemocytes.

5.5.2 Determining the effect of mRNA stability on EV-loading

3'UTR sequence features can recruit RBPs to aid RNA processing, such as polyadenylation and polyuridylation, which can alter mRNA stability and turnover [453–455, 553], potentially altering the likelihood that mRNAs become EV-loaded. 3' polyadenylation has been shown to be negatively correlated with EV secretion [245], with 3' polyuridylation appearing overrepresented in exosomes [233]. The mechanisms by which the enrichment of polyuridylation within EVs is achieved is not fully understood. It may be due to reduced cellular stability of polyuridylated RNAs, or an active secretion mechanism [233].

Harnessing 3' RNA-sequencing to assess cellular and EV RNAs, the relationship between the occurrence of specific EV-motifs and particular polyadenylation or polyuridylation patterns can be assessed [554]. This analysis could be complemented by biological validation of mRNA stability during time-course experiments. By

conditionally expressing the UAS-Cre constructs containing specific EV-motif combinations (Cre-V5, Cre-paraWT, Cre-paraACD etc.) in S2R+ cells, changes in cellular and EV mRNA levels can be assessed over time via qPCR. This may determine if the presence of specific EV-motif combinations significantly effect mRNA stability, potentially modulating mRNA EV incorporation.

Furthermore, additional *cre* constructs could be developed to assess the influence of altered 3'UTR GC content or length on mRNA stability and EV loading. These features may have been altered in the Cre-para constructs compared to the Cre-V5 condition, potentially influencing mRNA stability and enhancing EV incorporation, independently of the motif sequences studied.

5.5.3 Using the *Drosophila* Cre-LoxP system to identify novel proteins involved in EV uptake

In addition to using the Cre-LoxP system to study the mechanisms of EV biogenesis, loading and release, the *Drosophila* Cre-LoxP system also provides a unique opportunity to study EV uptake and functional delivery in complex *in vivo* environments. Gene expression and protein activity can be modulated (via CRISPR-Cas9 knock-out or mutation) in recipient cells by harnessing the alternative LexA-LexAop driver system [538, 539]. Factors previously shown to be involved in EV-uptake could be further validated *in vivo*, for example, tetraspanins, integrins or proteoglycans [289–292, 328]. In addition, novel factors required for EV internalisation or cargo endosomal escape may be identified, shedding light on processes which are not yet fully understood. This is an incredibly important aspect of EV biology because for EVs to be effectively harnessed for the delivery of customised cargos, a clear understanding of how cargos are targeted, internalised and processed in recipient cells is needed. This has relevance for the development of EV-related technologies or therapeutics.

5.5.4 Using the *Drosophila* Cre-LoxP system to study EV-mediated communication in a disease context

Investigations using the *Drosophila* Cre-LoxP system have so far focused on long-range communication under physiological conditions; however, in the context of diseases such as cancer, neurodegeneration or infection, EV biology can be modulated to contribute to the progression of the pathology [45, 49, 51]. This *Drosophila* Cre-LoxP model can now be combined with previously established disease models to study long-range communication in these contexts. One such disease model has been established by Xu et al. (2020) to study the interaction between caspase activity, hemocyte recruitment and cancerous epithelial tissues [555]. As previously highlighted in **Chapter 4**, the Cre-LoxP system was sensitive to detect communication between hemocytes and imaginal disc tissues. It would be interesting to determine the extent of this communication in a disease context and determine if this communication is bi-directional. Furthermore, since caspase activity has been related to various aspects of EV biology and appeared to influence reporter signalling in the Cre-LoxP system (**Section 5.4.2**), this tool may be harnessed to determine the extent of long-range communication upon the ablation of caspase activity in cancerous tissues. This may highlight additional roles of non-apoptotic caspase activity in immune cell recruitment and tumour microenvironment interaction, potentially identify novel targets for drug treatments to impede cancer progression.

5.6 Chapter Summary

This chapter described the assessment of the novel *Drosophila* EV-enriched 3'UTR mRNA motif combination, identified via the bioinformatic analysis described in **Chapter 3**, in cell culture and *in vivo* contexts. Results suggested that mRNA EV-loading may be enhanced upon the inclusion of the *para* 3'UTR, which in turn may successfully enhance the long-range mRNA transfer in complex tissues. The EV-enriched motifs within this 3'UTR, however, were shown to only partially contribute to the enhanced reporter signalling, since upon motif shuffling, no significant decrease in the prevalence of reporter signalling was observed. Additional factors such as mRNA stability, GC content or length may have been altered upon the addition of the *para* 3'UTR. Future work is now needed to determine the molecular mechanisms facilitating this enhancement, determining the influence of these additional features on EV-loading and long-range mRNA transfer.

In addition, the Cre-LoxP system was utilised to identify the potential influence of caspase activity on the processes of long-range cell communication, highlighting prospective non-apoptotic roles for these proteins which warrant further investigation.

In the future, the *Drosophila* Cre-LoxP system will also be combined with well-established disease models, such as those used to understand cancer, to determine how the mechanisms of long-range cell communication are disrupted in pathological scenarios. In addition, the Cre-LoxP system can be used in combination with the LexA-LexAop expression system to allow the modulation of specific factors in recipient cells, to study the mechanisms of EV-uptake and cargo processing. This provides an exciting opportunity to explore many aspects of EV biology which remain unexplored, potentially uncovering fundamental knowledge of this important cell communication mechanism with physiological and pathological relevance.

6 Discussion

Contents

6.1 Thesis summary 183
6.2 Key findings 185
6.2.1 Bioinformatic analysis of *Drosophila* cellular and EV RNA profiles identified EV-enriched mRNAs 185
6.2.2 Combinations of novel mRNA features may interact to modulate EV-loading versus cell-retention 186
6.2.3 EV-enriched motifs are prevalent within longer transcripts 186
6.2.4 EV-enriched motifs show significant sequence similarity with RBP binding sites 187
6.2.5 Atypical *Drosophila* motif identified which may be involved in alternative regulatory processes 188
6.2.6 Successful establishment of a *Drosophila* Cre-LoxP system to study cellular communication 189
6.2.7 Advantages of the *Drosophila* Cre-LoxP reporter system 190
6.2.8 Function of the novel *Drosophila* EV-loading motifs *in vivo* 191
6.2.9 Modulating the caspase pathway altered the activity of the *Drosophila* Cre-LoxP reporter system 191
6.3 Concluding remarks 194

6.1 Thesis summary

Whilst EV transfer has been acknowledged as a key mechanism of long-range cell communication, much remains unknown about the molecular mechanisms which facilitate the loading of specific cargo into these membrane bound structures. In particular, many aspects of the biological processes which facilitate the sorting of mRNAs for EV incorporation remain unexplored. Filling this knowledge gap is especially important as the functional delivery of mRNA to recipient cells holds the potential to alter basic biological processes and change cell behaviour.

To better define the molecular features promoting mRNA loading into EVs, an RNA-sequencing analysis of *Drosophila* cellular and EV RNAs was conducted, followed by a bioinformatic analysis of mRNA sequence features. This analysis identified numerous 3'UTR motif sequences abundant within EV-enriched mRNAs. Intriguingly, these motifs commonly appeared in combinations, showing a preferential arrangement within the 3'UTR and tending to occur within particular RNA secondary structures. Furthermore, one motif sequence showed highly significant sequence similarity with the consensus binding site for two RNA binding proteins, suggesting its potential role in recruiting trans-acting factors to modulate RNA localisation, which may alter the incorporation of specific RNAs into EVs. Interestingly, this analysis identified a number of EV mRNA features which have been separately identified in a number of mammalian studies, suggesting potential functional conservation of these features between species.

The development of a *Drosophila* Cre-LoxP reporter system provided the opportunity to assess the biological function of these EV mRNA features. qPCR analysis highlighted that the addition of the *para* 3'UTR (containing the EV-enriched motif combination) to *cre* mRNA in *Drosophila* tended to enhance mRNA EV incorporation, with further *in vivo* analysis suggesting that this resulted in a significant increase in long-range reporter signalling. Shuffling three of the EV-enriched motifs present in this 3'UTR, however, resulted in no significant decrease in mRNA EV-loading and long-range signalling, suggesting additional factors such as mRNA length, stability or GC content may also impact mRNA EV-loading. Future

work is needed to fully explore the contribution of individual EV motifs, as well as these additional factors, to the process of mRNA EV-loading and transfer. The specificity of these results was corroborated, however, by the fact that a previously identified 25-nucleotide mammalian EV-loading motif [248] failed to enhance *cre* mRNA EV-loading in this *Drosophila* system.

The *Drosophila* tool developed here may be used to uncover additional factors involved in the long-range transfer of mRNAs, for example, disrupted effector caspase activity resulted in altered Cre-LoxP reporter signalling within the IWD. The incorporation of the alternative LexA-LexAop binary system in combination with the UAS-Gal4 based Cre-LoxP system also provides the unique opportunity to study molecular mechanisms important for EV uptake in recipient cells. These experiments may identify novel targets which could be relevant to manipulate long-range mRNA transfer in mammalian systems, relevant for therapeutic or scientific purposes.

6.2 Key findings

6.2.1 Bioinformatic analysis of *Drosophila* cellular and EV RNA profiles identified EV-enriched mRNAs

Whilst the presence of diverse EV RNA profiles has been indented from many cell types, how these RNA cargos get selectively incorporated into EVs is less well understood. Mechanisms determining RNA incorporation into EVs have been more widely studied for microRNAs, [74, 216, 217, 233, 234, 240, 249, 250, 254, 256, 556–558]; however, less is known about mRNA EV incorporation [177, 245, 247, 248]. To gain insight into specific mRNA loading and to probe for mRNA features which may facilitate EV incorporation, a bioinformatic analysis of *Drosophila* cellular and EV RNAs was carried out.

Comparing the RNA profile between *Drosophila* S2R+ cells and EVs revealed that whilst cellular RNA remained consistent between biological replicates, EV RNA cargo was more variable; a characteristic also observed in mammalian EV RNA-sequencing datasets [216, 224]. EV RNA loading is highly dynamic, varying depending on environmental factors and cell stimulation status [167].

Aside from rRNA, the most common RNA type in both the cellular and EV fractions was mRNA. These mRNAs may be loaded by chance and therefore, their EV abundance would likely reflect their cellular abundance. However, by comparing cellular and EV mRNA content, EV-specific and EV-enriched mRNAs ('EV+' mRNAs) were identified, providing an interesting focus for downstream analyses. These mRNAs may become enriched due to active loading processes, to facilitate the transfer of specific transcripts to recipient cells. Interestingly, gene ontology analysis highlighted that these 'EV+' mRNAs presented predicted functions related to development, pattern specification and regulation. This suggested that RNAs with particular functions may be selectively incorporated into EVs. By probing the 'EV+' mRNA cargo, it was anticipated that common features may be present which encourage these transcripts to become incorporated into EVs.

6.2.2 Combinations of novel mRNA features may interact to modulate EV-loading versus cell-retention

The UTRs of mRNAs are well known for their roles in RNA localisation, stability or translational regulation [463]. EV-loading is likely modulated via similar processes [237], with EV-enriched features previously identified within mRNA UTRs in mammalian models [245, 247–249, 462]. Interestingly, motif sequences were found to be enriched in the *Drosophila* 'EV+' 3'UTR sequences, compared to the same region in cell specific mRNAs. These motifs frequently occurred in combination with one another, with the same motif also observed multiple times within the same UTR sequence. Similar observations have been made in mammalian models, where it was proposed that single motifs alone may have poor regulatory potential, but akin to DNA regulatory elements which recruit transcription factors, motifs may act together to increase their regulatory capacity [245, 464].

Some EV-enriched motif combinations were entirely absent from the cell specific transcripts, suggesting they may have EV-loading functions (e.g. ABCD). Conversely, motif B alone was EV-depleted, suggesting it may influence cell-retention, a phenomenon previously highlighted in a mammalian model [234], suggesting a potential evolutionary conservation of this function. An interesting avenue of future work will be to investigate mRNA alternative splicing, as splice variants of the same transcript may contain different motif combinations. Isoform level transcriptomics would potentially shed light on how motifs work in combination with one another to modulate mRNA EV-loading or cell-retention.

6.2.3 EV-enriched motifs are prevalent within longer transcripts

In this study, median transcript length profiles for the top 1000 most abundant genes in cells and EVs were highly similar, reflecting observations made by Wei et al. (2017) [215]. Whilst EV-loaded RNAs are typically thought to be 200 to 1000 nucleotides long [224, 480], longer EV transcripts (up to approximately 6900 nts) have been detected [246], with qPCR analysis highlighting full length transcripts

up to 3000 nts long within EVs [215]. These results suggested that longer RNAs may not be excluded from becoming EV-loaded.

Indeed, an analysis of transcript length highlighted that motif occurrences may increase the EV abundance of longer transcripts. Novel *Drosophila* motifs were present in the 3'UTR of transcripts which were significantly longer (either full length or 3'UTR length) than those transcripts which lacked any 3'UTR motif occurrences. Transcripts with long 3'UTRs may simply be more likely to carry motifs, with this in itself a potential loading feature. Tushev et al. (2018) describe that differential 3'UTR lengths provide transcripts with a varying capacity to carry 3'UTR elements, altering their 'regulatory competitiveness' [439]. In the future, a read coverage analysis at the transcript isoform level will determine if the long motif carrying transcripts become EV-loaded in their entirety, or if fragments of these mRNAs are instead present. EV-loaded mRNA fragments can still carry out biological functions, for example, they can provide competitive binding sites for *trans*-acting factors [210] and therefore, RNA fragment loading and the influence of motif occurrences in this process may also be informative.

6.2.4 EV-enriched motifs show significant sequence similarity with RBP binding sites

Similarly to DNA regulatory elements recruiting transcription factors, these RNA motifs may be involved in the recruitment of *trans*-acting factors to modulate EV-loading. RBP binding to EV-enriched RNA motifs has been highlighted in mammalian studies [175, 217, 234, 245, 247, 249, 462]. *Drosophila* motifs occurred predominantly within the UTR, with a number of motifs entirely absent from coding regions. UTRs are accessible for making interactions important for regulatory purposes, but coding sequences are compacted to protect the open reading frame and allow sub-cellular RNA localisation to be controlled [463].

Trans-acting factor interactions may be facilitated by motif sequences forming particular RNA secondary structures. Indeed, the context of the motif occurrence, their location and structure, are important factors which may influence their function

[245]. A quantitative analysis of the secondary structure of the 3'UTRs of transcripts containing the ABCD motif combination revealed that each motif had an individual preference to occur in looped, unpaired secondary structures. Interestingly, previous studies have also highlighted the presence of motif secondary structure, with a visual analysis of motifs regions revealing their prevalence within stem-loop regions [245, 246, 248]. Secondary structure specificities may influence RNA binding protein recruitment or alter membrane association, ultimately modulating RNA incorporation in EVs.

Comparing the novel *Drosophila* motifs with a database of known *Drosophila* RNA binding proteins revealed that motif C showed highly significant sequence similarity with binding sites for CNOT4 and Smooth (a hnRNPL protein ortholog). These RBPs are included in the ExoCarta database as they have been identified within EVs in either protein or mRNA form. The functions of the RBP hits identified relate to processes such as Jak/Stat signalling, mRNA degradation, translational regulation, alternative splicing and even EV-loading [465, 466, 468, 469, 472, 476]. These preliminary RBP hits can now be validated via biochemical analysis.

6.2.5 Atypical *Drosophila* motif identified which may be involved in alternative regulatory processes

One recurring phenomenon when analysing the novel *Drosophila* motifs, was the apparent differences in the nature of motif B. Motif B was never found alone within the 'EV+' transcripts; it was always found in combination with another motif. In addition, motif B (together with motif A) showed a higher abundance in coding regions rather than in the UTRs. It also tended to appear predominantly in paired regions rather than loops. Similarly to the EV motifs identified by Batagov et al. (2011) [245], motif B did not show sequence similarity with any of the known RBP binding sites, suggesting that this motif may bind with unconventional RBPs or have roles in alternative regulatory processes, such as miRNA binding. It has been shown that the mammalian 25-nucleotide motif associates with pre-miR-1289 to enhance RNA loading [248]. It is possible that motif B is involved in similar

processes within *Drosophila* to regulate RNA cell retention versus EV-loading and in the future, it would be interesting to carry out further analysis to identify potential interactions between motif B and *Drosophila* miRNAs.

6.2.6 Successful establishment of a *Drosophila* Cre-LoxP system to study cellular communication

In order to biologically validate the bioinformatically predicted EV-enriched motifs and determine their effect on mRNA EV incorporation, a *Drosophila* Cre-LoxP system was established. This system was based upon a tool developed in mice to visualise the transfer of EVs containing *cre* mRNA *in vivo* [309–312]. Initial cell culture analysis using *Drosophila* S2R+ cells confirmed the functionality of the reporter system and confirmed EV-loading of *cre* mRNA, with a proportion of this *cre* mRNA protected from RNase degradation suggesting its incorporation into the EV lumen. This result was in line with previous mammalian systems, where *cre* mRNA and not Cre protein was observed within EVs [309–312]. Future Western Blot analysis of EV cargo and experiments harnessing Cre RNAi expression in recipient cells are required to confirm the absence of Cre protein within EVs when using the *Drosophila* Cre-LoxP system.

Following the success of the cell culture analysis, the *Drosophila* Cre-LoxP system was transferred *in vivo*. Using the UAS-Gal4 system [559], Cre expression was spatially controlled, allowing cell communication between specific cell and tissue types to be investigated. Temporal Cre expression was achieved using the Gal4 suppressor, Gal^{80} [560], which allowed Cre expression to be switched-on during precise durations of development. Control experiments were carried out to ensure any $Reporter^{GFP+}$ signal observed was not due to spontaneous excision of the reporter stop-cassette, nor the result of 'leaky' Cre expression outside the expected expression domain.

Using the *in vivo Drosophila* Cre-LoxP system, cell communication was observed in the anterior to posterior direction of the IWD. Posterior $Reporter^{GFP+}$ signal was largely reproducible, with a 'gap region' frequently observed in the cells most

adjacent to the anterior/posterior boundary. This suggested more long-range and directed Cre transfer mechanisms, perhaps mediated by EVs, rather than random Cre protein diffusion. Using the temporal expression of Cre, the 'active window' of this transfer was defined as up to three days after egg laying, suggesting this communication may be involved in early development or tissue patterning; with EVs previously implicated in these processes in the posterior to anterior direction [36, 47, 362, 490]. These results suggested that the Cre-LoxP system may have highlighted novel cell communication events within the IWD, with potential functional importance in early IWD development.

Communication was not only identified within tissues, but also between different cell types. When Cre expression was driven in *Drosophila* immune cells, the hemocytes, Reporter^{GFP+} cells were observed in imaginal disc tissues. Hemocytes and imaginal disc tissues are closely associated, with hemocytes not only removing dead cells and debris, but also inducing apoptosis induced proliferation (AiP) to stimulate the surviving cells to divide [355, 522], with EVs implicated in similar processes in mammalian models [524, 525]. It is possible that the *Drosophila* Cre-LoxP system detected EV-mediated cell communication in this context. In the future, it would be interesting to determine the functional relevance of this signalling during diseases such as cancer, when immune cell regulation is disturbed.

6.2.7 Advantages of the *Drosophila* Cre-LoxP reporter system

The *Drosophila* Cre-LoxP system presented an increased efficiency compared to those observed in previous mouse models. IWD anterior to posterior communication events were detected in approximately 40% of wing discs analysed under standard Cre-V5 conditions, compared to alternative mouse models which showed an efficiency of approximately 0.3 to 5.8%, depending on the cellular scenario [308–310, 312]. In the *Drosophila* system, Cre and reporter expression is controlled within a single organism, without the need for stress induction, EV injection or tumour grafts, as required in mouse models [309–312], reducing experimental variation and providing

more biologically relevant conditions. An additional benefit of the *Drosophila* Cre-LoxP system is the ability to harness the *Drosophila* toolbox to carry out genetic manipulations in EV-producing or recipient cells more simply.

6.2.8 Function of the novel *Drosophila* EV-loading motifs *in vivo*

The *Drosophila* Cre-LoxP system was used to determine the effects of the bioinformatically predicted EV-enriched ABCD motif combination on mRNA EV-loading. Encouragingly, qPCR experiments highlighted that upon the addition of the *para* 3'UTR (containing the EV-enriched motifs) to *cre* mRNA, EV *cre* mRNA incorporation tended to increase. *In vivo*, a significant increase in the frequency and area of IWD posterior Reporter^{GFP+} signal was observed upon the addition of the *para* 3'UTR to *cre*. Upon the simultaneous shuffling of motifs A, C and D within this 3'UTR, however, only a partial and non-significant difference in *cre* mRNA EV-loading was observed via qPCR. Furthermore, whilst the posterior area becoming Reporter^{GFP+} tended to decrease upon motif shuffling, the prevalence of posterior Reporter^{GFP+} signal remained unchanged. Future work will focus on determining if shuffling alternative motif combinations hinders the long-range transfer of *cre* mRNA in an *in vivo* context. Additionally, factors such as mRNA length, stability and GC content may have been altered upon the addition of the *para* 3'UTR, and future work is needed to assess the influence of these factors on mRNA EV-loading. These experiments have the potential to demonstrate, for the first time, the successful manipulation of mRNA EV incorporation and transfer *in vivo* via 3'UTR sequence manipulation, with future work needed to determine the mechanism of action facilitating this increase.

6.2.9 Modulating the caspase pathway altered the activity of the *Drosophila* Cre-LoxP reporter system

The *Drosophila* Cre-LoxP system also provided a tool to identify novel pathways involved in modulating long-range communication. Caspases have recently been implicated in the process of cargo cleavage for EV-loading and uptake [158, 165],

separate from their roles in Apoptotic EV (ApoEV) formation [150–152]. This adds to an ever-growing appreciation of non-apoptotic caspase functions [166, 545, 546]. Intriguingly, caspase-3 has been shown to cleave a number of RBPs (YBOX-1, SYNCRIP, hnRNPA2B1 and hnRNPK) known to associate with RNA EV-motifs to influence RNA EV incorporation [255]. Caspases may be key regulators of RNA and RBP interactions to influence RNA stability and localisation, which may ultimately modulate mRNA EV-loading. Given that RNA motifs, and their RBP associations, have been shown to have either cell retention or EV incorporation roles [234, 263], caspases may enhance or diminish EV-loading depending on the context (**Figure 6.1**). In this study, the absence of effector caspase activity resulted in increased reporter signalling, suggesting that enhanced *cre* mRNA EV-loading and/or secretion was taking place. In this context, caspase activity may function as a break, to reduce secretion.

Conversely, however, reduced effector caspase activity may allow initiator caspases, such as Dronc, to be more readily available to promote EV-loading and/or secretion. Indeed, it was the initiator caspase, caspase-8, involved in cargo loading in mammalian cells [158]. Future work will focus on modulating the caspase pathway at the upstream levels (Hid, Grim, Reaper and Dronc) to determine the influence of these factors on EV-loading and secretion. It will also be possible to harness the additional LexA-LexAop expression system to modulate gene expression in recipient cells to identify roles played by caspases, and other molecular mechanisms, in EV uptake and cargo processing. This tool may identify novel pathways harnessed by cells to control EV-mediated communication and these may be relevant in mammalian systems.

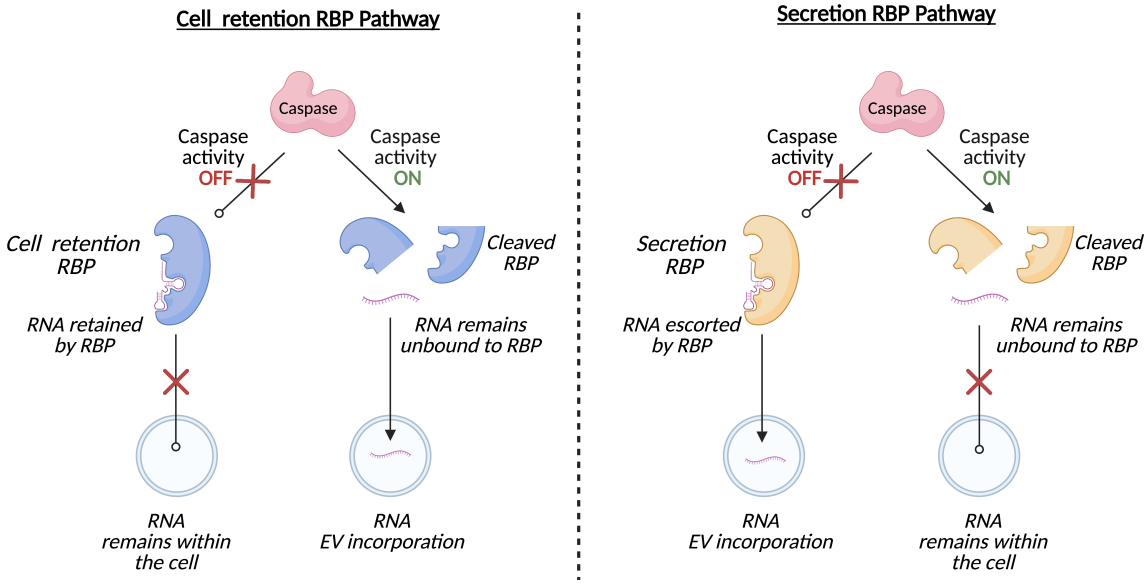


Figure 6.1: Possible mechanism of action of caspase activity for modulating RNA and RBP associations to influence RNA EV incorporation.

6.3 Concluding remarks

The EV-mediated transfer of RNA is an important mechanism facilitating the regulation of gene expression at a distance, playing roles in development, tissue patterning, survival and immune regulation. The mechanisms by which cells release particular RNA cargo via EVs, particularly mRNA, are not fully understood. The bioinformatic analysis of *Drosophila* cellular and EV mRNAs revealed 3'UTR features of EV-enriched mRNAs which may facilitate their EV incorporation. Interestingly, similar features have been separately identified in a number of mammalian systems [234, 245–249, 462], highlighting their potential functional conservation.

This project saw the successful establishment of a *Drosophila* Cre-loxP based reporter system to study long-range cell communication in complex tissues. This tool was based on a previously established system in mice [309–312], however, use of the *Drosophila* model negated the need for EV injections or tumour grafts. Cell culture experiments confirmed the presence of EV-loaded *cre* mRNA, with future work needed to rule out the presence of EV-loaded Cre protein.

Cell communication events were detected between cells of the imaginal wing disc, with further investigations highlighting that these communication events occurred during the early stages of development only. Communication was also detected between immune cells and imaginal disc tissues. Communication between these cells and tissues has been established [47, 355, 362, 490, 522, 523], suggesting that the *Drosophila* Cre-LoxP system may be used to gain insights into the regulation of biologically relevant cellular communication.

The *Drosophila* Cre-LoxP system was harnessed to assess the function of the mRNA 3'UTR features identified in EV-enriched mRNAs. The addition of the *paralytic* 3'UTR, containing EV-enriched motifs, to *cre* mRNA tended to increase *cre* mRNA EV incorporation, also reflected *in vivo* where significantly enhanced $\text{Reporter}^{\text{GFP+}}$ signal was observed. A non-significant reduction in this enhancement was observed, both in cell culture and *in vivo*, upon the simultaneous shuffling of motifs A, C and D, highlighting that they may only play a partial role in altering EV-loading. Future work is needed to determine the influence of the individual

motif sequences within the ABCD motif combination, as well as to assess additional factors such as mRNA length, stability and GC content, on mRNA EV-loading. This research has the potential to provide *in vivo* evidence that mRNA sequences may be manipulated to enhance long-range mRNA transfer in complex tissues. This tool may be used to identify novel mechanisms to manipulate EV-loading and delivery of mRNA cargo for scientific or therapeutic purposes in mammalian settings.

This study has also demonstrated the potential of the *Drosophila* Cre-LoxP system to study molecular pathways involved in regulating EV-mediated communication. Upon the inhibition of effector caspases, proteins primarily known for their role in cell death but also shown to be involved in modulating EV-loading and uptake [157, 158, 165], enhanced reporter signalling was observed in the IWD. The molecular mechanisms responsible for this enhancement are not yet understood, but recent evidence suggests that caspase-mediated cleavage of RBPs may alter mRNA EV-loading [255]. Future work will focus on modulating caspase activity at various points along the pathway to determine their influence on EV-mediated cell communication.

The *Drosophila* Cre-LoxP system has the potential to provide a deeper understanding of the fundamental mechanisms regulating EV-biogenesis, cargo loading and transfer. This tool has the capacity to accelerate the field of cell communication research, allowing complex genetic experiments to be carried out more efficiently. This tool has already yielded insights into the cell communication occurring within and between various cell types in *Drosophila*; however, the study of alternative tissues, such as at the neuromuscular junction or the brain, as well as in disease contexts such as cancer, remains unexplored. Information gained using this tool may be utilised to enhance our ability to load and target customised cargos via EVs, potentially for therapeutic purposes. It is hoped the *Drosophila* Cre-LoxP system will have far reaching applications and be used to study wide a range of biological contexts.

7 Supplementary

Contents

7.1 Supplementary Information 197
7.1.1 Generation of *cre* fragments with the *paralytic* 3'UTR . 197
7.1.2 Optimisation of conditions for *Drosophila* SR2+ cell culture, EV extraction and EV characterisation 200
7.2 Supplementary Figures 204
7.3 Supplementary Tables 213
7.4 Supplementary Listings 237

7.1 Supplementary Information

7.1.1 Generation of *cre* fragments with the *paralytic* 3-prime untranslated regions

```
   1 GCGGCCGCAT GTCCAATTTA CTGACCGTAC ACCAAAATTT GCCTGCATTA
  51 CCGGTCGATG CAACGAGTGA TGAGGTTCGC AAGAACCTGA TGGACATGTT
 101 CAGGGATCGC CAGGCGTTTT CTGAGCATAC CTGGAAAATG CTTCTGTCCG
 151 TTTGCCGGTC GTGGGCGGCA TGGTGCAAGT TGAATAACCG GAAATGGTTT
 201 CCCGCAGAAC CTGAAGATGT TCGCGATTAT CTTCTATATC TTCAGGCGCG
 251 CGGTCTGGCA GTAAAAACTA TCCAGCAACA TTTGGGCCAG CTAAACATGC
 301 TTCATCGTCG GTCCGGGCTG CCACGACCAA GTGACAGCAA TGCTGTTTCA
 351 CTGGTTATGC GGCGGATCCG AAAAGAAAAC GTTGATGCCG GTGAACGTGC
 401 AAAACAGGCT CTAGCGTTCG AACGCACTGA TTTCGACCAG GTTCGTTCAC
 451 TCATGGAAAA TAGCGATCGC TGCCAGGATA TACGTAATCT GGCATTTCTG
 501 GGGATTGCTT ATAACACCCT GTTACGTATA GCCGAAATTG CCAGGATCAG
 551 GGTTAAAGAT ATCTCACGTA CTGACGGTGG GAGAATGTTA ATCCATATTG
 601 GCAGAACGAA AACGCTGGTT AGCACCGCAG GTGTAGAGAA GGCACTTAGC
 651 CTGGGGGTAA CTAAACTGGT CGAGCGATGG ATTTCCGTCT CTGGTGTAGC
 701 TGATGATCCG AATAACTACC TGTTTTGCCG GGTCAGAAAA AATGGTGTTG
 751 CCGCGCCATC TGCCACCAGC CAGCTATCAA CTCGCGCCCT GGAAGGGATT
 801 TTTGAAGCAA CTCATCGATT GATTTACGGC GCTAAGGATG ACTCTGGTCA
 851 GAGATACCTG GCCTGGTCTG GACACAGTGC CCGTGTCGGA GCCGCGCGAG
 901 ATATGGCCCG CGCTGGAGTT TCAATACCGG AGATCATGCA AGCTGGTGGC
 951 TGGACCAATG TAAATATTGT CATGAACTAT ATCCGTAACC TGGATAGTGA
1001 AACAGGGGCA ATGGTGCGCC TGCTGGAAGA TGGCGATGGT AAGCCTATCC
1051 CTAACCCTCT CCTCGGTCTC GATTCTACGT AGGACGTCTG AGCCAGGCCT
1101 CGCCCCCCCC TCCAAGATGC ACGCGAGTAT TAGCATGAGT ACGTGTTGGA
1151 TGTTGCATGT TGCATCAGCG ATGCAGCAGC GACAGTGAGC TAACAACAAC
1201 AGCAACAAGG CGAACAACAA GCATAATTTT CTGCTATATA ATGAAAGAAC
1251 ACCATACAAA CAAATACAAC AACAATTCTA GACTCGAGGA CGTCTGAGCC
1301 AGGCCTCGCC CCCCCCTCCA AGATGCACGC GAGTATTAGC ATGAGTACGT
1351 GTTGGATGTT GCATGTTGCA TCAGCGATGC AGCAGCGACA GTAATCCCAA
1401 GAAAAAACGG CACAGGCGAA CAACAAGCAT AATTTTCTGC TATATAATGA
1451 AAGAACACCA TACAAACAAA TACAACAACA ATTCTAGACT CGAGGGTACC
```

Figure 7.1: GENEWIZ construct containing *cre* along with the *paralytic* 3'UTR in wild-type (WT) form or with the A motif mutated. Red = Cre, Blue = V5, Yellow = WT, Green = A shuffled.

```
   1 GCGGCCGCAT GTCCAATTTA CTGACCGTAC ACCAAAATTT GCCTGCATTA
  51 CCGGTCGATG CAACGAGTGA TGAGGTTCGC AAGAACCTGA TGGACATGTT
 101 CAGGGATCGC CAGGCGTTTT CTGAGCATAC CTGGAAAATG CTTCTGTCCG
 151 TTTGCCGGTC GTGGGCGGCA TGGTGCAAGT TGAATAACCG GAAATGGTTT
 201 CCCGCAGAAC CTGAAGATGT TCGCGATTAT CTTCTATATC TTCAGGCGCG
 251 CGGTCTGGCA GTAAAAACTA TCCAGCAACA TTTGGGCCAG CTAAACATGC
 301 TTCATCGTCG GTCCGGGCTG CCACGACCAA GTGACAGCAA TGCTGTTTCA
 351 CTGGTTATGC GGCGGATCCG AAAAGAAAAC GTTGATGCCG GTGAACGTGC
 401 AAAACAGGCT CTAGCGTTCG AACGCACTGA TTTCGACCAG GTTCGTTCAC
 451 TCATGGAAAA TAGCGATCGC TGCCAGGATA TACGTAATCT GGCATTTCTG
 501 GGGATTGCTT ATAACACCCT GTTACGTATA GCCGAAATTG CCAGGATCAG
 551 GGTTAAAGAT ATCTCACGTA CTGACGGTGG GAGAATGTTA ATCCATATTG
 601 GCAGAACGAA AACGCTGGTT AGCACCGCAG GTGTAGAGAA GGCACTTAGC
 651 CTGGGGGTAA CTAAACTGGT CGAGCGATGG ATTTCCGTCT CTGGTGTAGC
 701 TGATGATCCG AATAACTACC TGTTTTGCCG GGTCAGAAAA AATGGTGTTG
 751 CCGCGCCATC TGCCACCAGC CAGCTATCAA CTCGCGCCCT GGAAGGGATT
 801 TTTGAAGCAA CTCATCGATT GATTTACGGC GCTAAGGATG ACTCTGGTCA
 851 GAGATACCTG GCCTGGTCTG GACACAGTGC CCGTGTCGGA GCCGCGCGAG
 901 ATATGGCCCG CGCTGGAGTT TCAATACCGG AGATCATGCA AGCTGGTGGC
 951 TGGACCAATG TAAATATTGT CATGAACTAT ATCCGTAACC TGGATAGTGA
1001 AACAGGGGCA ATGGTGCGCC TGCTGGAAGA TGGCGATGGT AAGCCTATCC
1051 CTAACCCTCT CCTCGGTCTC GATTCTACGT AGGACGTCTG AGCCAGGCCT
1101 CGCCCCCCCC TCCAAGATGC ACGCGAGTAT TAGCATGAGT TAATTGTGTT
1151 GGGCTTGTAC GGCATCAGCG ATGCAGCAGC GACAGTAATC CCAAGAAAAA
1201 ACGGCACAGG CGAACAACAA GCATAATTTT CTGCTATATA ATGAAAGAAA
1251 CAAAAACACA ACCAATACAA CTCAATTCTA GACTCGAGGA CGTCTGAGCC
1301 AGGCCTCGCC CCCCCCTCCA AGATGCACGC GAGTATTAGC ATGAGTACGT
1351 GTTGGATGTT GCATGTTGCA TCAGCGATGC AGCAGCGACA GTAATCCCAA
1401 GAAAAAACGG CACAGGCGAA CAACAAGCAT AATTTTCTGC TATATAATGA
1451 AAGAAACAAA AACACAACCA ATACAACTCA ATTCTAGACT CGAGGGTACC
```

Figure 7.2: GENEWIZ construct containing *cre* along with the *paralytic* 3'UTR with the ACD motifs mutated or the AC motifs mutated. Red = Cre, Blue = V5, Yellow = ACD shuffled, Green = AC shuffled.

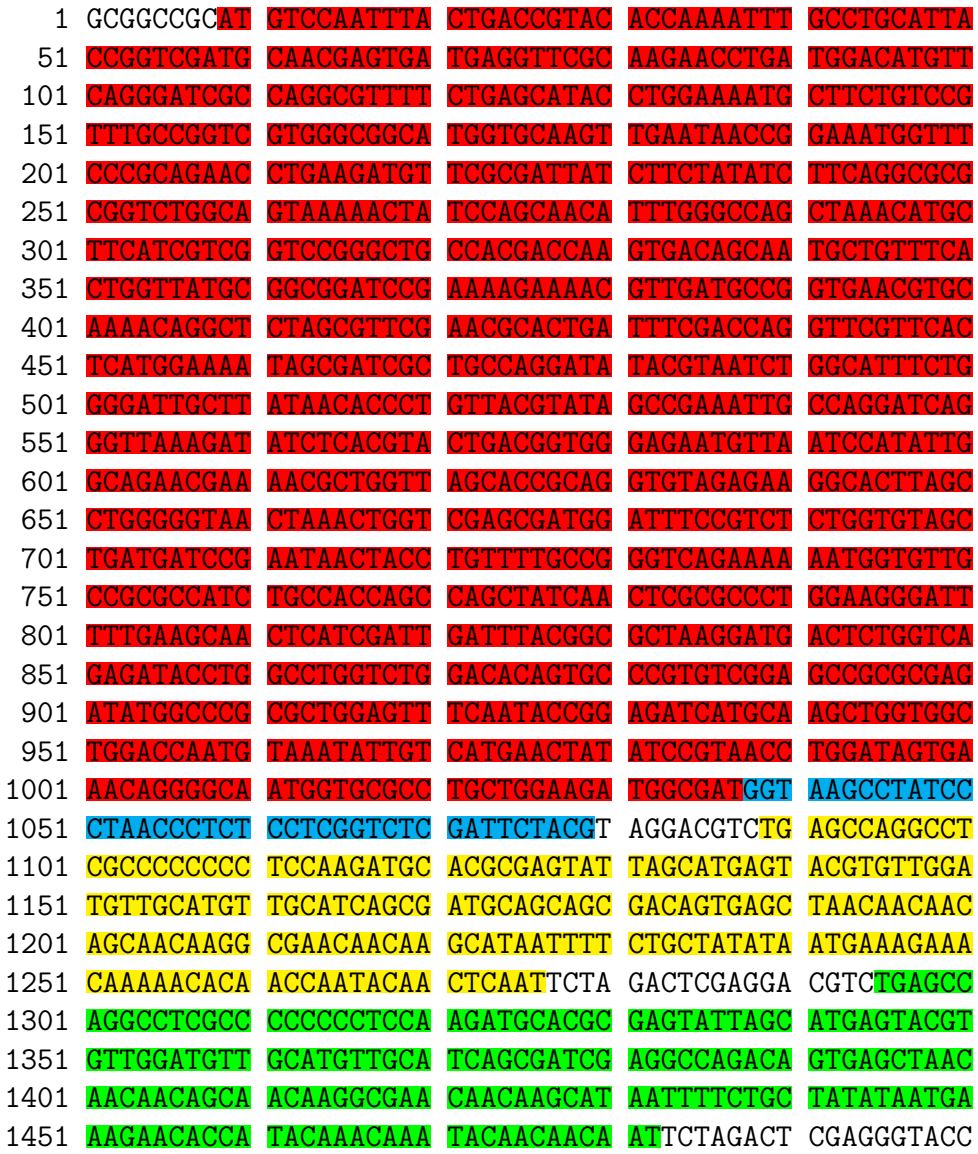


Figure 7.3: GENEWIZ construct containing *cre* along with the *paralytic* 3'UTR with the C motif mutated or the B motif mutated. Red = Cre, Blue = V5, Yellow = C shuffled, Green = B shuffled.

7.1.2 Optimisation of conditions for *Drosophila* SR2+ cell culture, EV extraction and EV characterisation

In order to confirm the ability of the Cre-LoxP reporter system to detect EV-mediated transfer of *cre* mRNA, qPCR experiments of EV-loaded RNA from Cre transfected cells were carried out. This process first required the optimisation of the cell culture and EV extraction process for *Drosophila* S2R+ cells.

Optimisation of cell culture conditions for EV extraction

According to the MISEV2018 guidelines, EVs are ideally obtained from culture medium conditioned by cells in the absence of fetal bovine serum (FBS) [561]. This is due to EV contamination from bovine sources. It is also possible to deplete EVs from the FBS before use, via ultracentrifugation [562]. It has been shown, however, that the use of serum free media or media supplemented with EV-depleted FBS can affect cellular behaviour or the nature and composition of EVs released, due to the removal of important growth factors along with the EVs [563, 564].

Different cell culture conditions were assessed to determine the effect of serum free or EV-depleted media on cell proliferation and cell death levels (as described in **Section 2.2.5**). Results showed that cell number significantly decreased when EV-depleted or serum free culture conditions were used, when compared to normal media conditions (EV-depleted p-adj = 0.0011; serum free p-adj = 0.0426) (**Figure 7.4i**). No significant difference in the percentage of DCP1 positive cells (a cell death marker) was observed in any of the cell culture conditions tested (**Figure 7.4ii**), confirming that increased cell death was not the cause of lower cell numbers. The decrease in cell number may be accounted for by a reduction in proliferation, shown by a decrease in the percentage of PH3 positive cells. Whilst this decrease was highly significant in the serum free condition (p-adj = 0.0008), it was not significant in the EV-depleted condition (p-adj = 0.1542) (**Figure 7.4iii**). These experiments provided initial evidence to suggest that EV-depleted and serum free conditions over a short culture period (3 days) did not significantly increase cell death levels. It was likely that EV preparations from these cell culture conditions

would not be contaminated with increased levels of apoptotic bodies. This was a single experimental replicate, however, with more experimental replicates required to confirm this relationship.

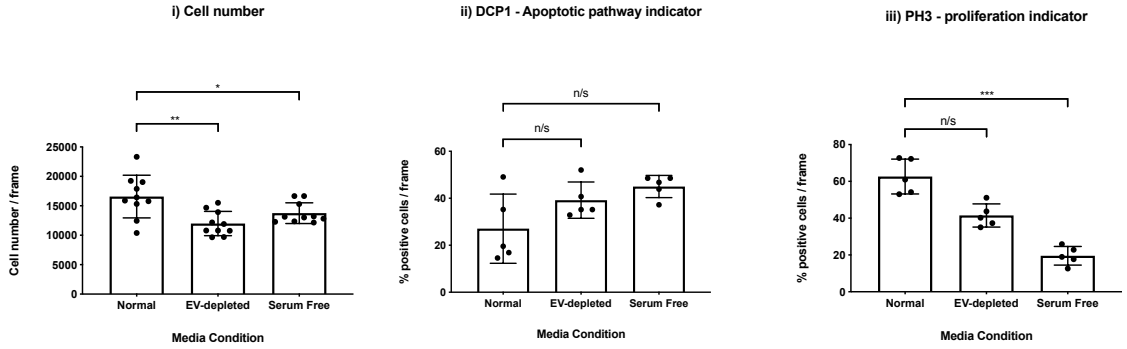


Figure 7.4: Cell culture conditioning trial to determine (i) cell number, (ii) apoptosis (via DCP1 immunostaining) and (iii) proliferation (via PH3 immunostaining) levels in normal, EV-depleted and serum free cell culture conditions for *Drosophila* S2R+ cells. Mean ± standard deviation is displayed. DCP1 and PH3 statistical significance determined via a Kruskal-Wasllis multiple comparisons test. Cell number statistical significance determined via an ordinary one-way ANOVA as the data passed the D'Agostino and Pearson normality test. For (i), each point relates to quantifications from a single image (five per technical replicate). For (ii) and (iii), each point relates to quantifications from a single image (five from a single technical replicate of each condition).

Optimisation of EV extraction protocols for use with *Drosophila* S2R+ cell culture and characterisation of EVs

The size-exclusion chromatography (SEC) method for EV extraction has previously been optimised within the lab for use with mammalian cell culture. In mammalian cell culture, EVs (with minimal protein contamination) are isolated from fractions 6 to 10 after SEC (data not shown). The fractions appropriate for use to isolate *Drosophila* EVs were not yet determined. EV extractions were carried out according to **Section 2.2.6**. The protein concentration within each fraction was determined via a BCA assay and particle concentration was determined via Nanosight Nanoparticle Tracking Analysis (NTA). Results confirmed that *Drosophila* S2R+ EVs were isolated in fractions 6 to 10, with particle concentration reaching its peak in fraction 8 (mean = 2.01 x 10^{-9} particles per ml) (**Figure 7.5**). Protein concentration did

not increase until after fraction 10. After this point, particle concentrations became variable between replicates, perhaps due to the presence of protein aggregates misidentified as particles via NTA. These results confirmed that fractions 6 to 10 should be retained for the analysis of *Drosophila* S2R+ EVs.

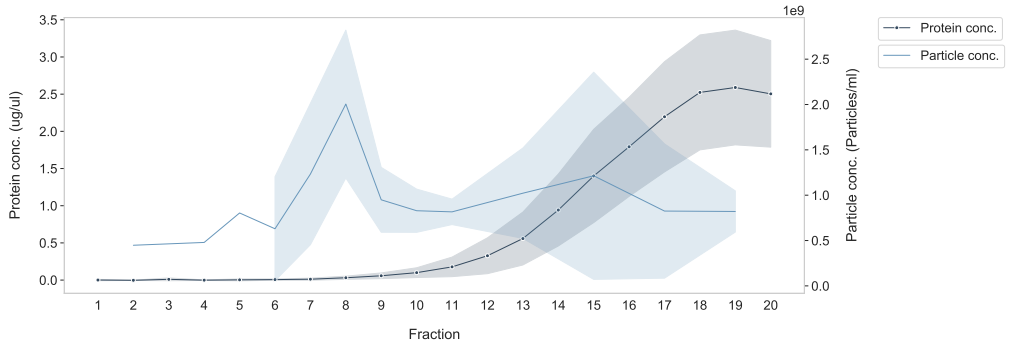


Figure 7.5: Nanosight Nanoparticle Tracking Analysis and protein concentration analysis of fractions 1 to 20 obtained after SEC of *Drosophila* S2R+ cell conditioned media. Mean values for protein concentrations determined from four technical replicates. Mean values for particle concentrations values determined from four technical replicates (Fractions 6-10, 13, 15 and 17), two technical replicates (Fraction 19) or one technical replicate (Fraction 2, 4 and 5). Shaded areas define standard deviations.

Transmission electron microscopy (TEM) analysis and Western Blotting was carried out[38] to confirm the presence of bona fide EVs within the SEC-isolated EV fractions. TEM analysis highlighted that particles isolated via SEC presented a cup shaped morphology, in the size range of 40 to 150 nm, typical for EVs (**Figure 7.6**). In addition, Latebloomer (the only *Drosophila* tetraspanin for which an antibody was available) was clearly visible in *Drosophila* EVs (**Figure 7.7**). Beta-actin, a pan-EV marker [565], was also clearly observed in *Drosophila* EVs. TSG101 and Hsp70 are common mammalian EV markers; however, they were absent in *Drosophila* EVs. The golgi marker, Golgin-84, presented either a weak or undetected signal, with the nuclear marker, Histone H3, also not visible. This highlighted that EV fractions were not contaminated with cellular material. A non-conditioned media control confirmed that the Western Blot signals observed were due to the

[38]by Elizabeth Dellar

presence of particles isolated from cell conditioned media and not the result of contamination from the media or supplemented FBS.

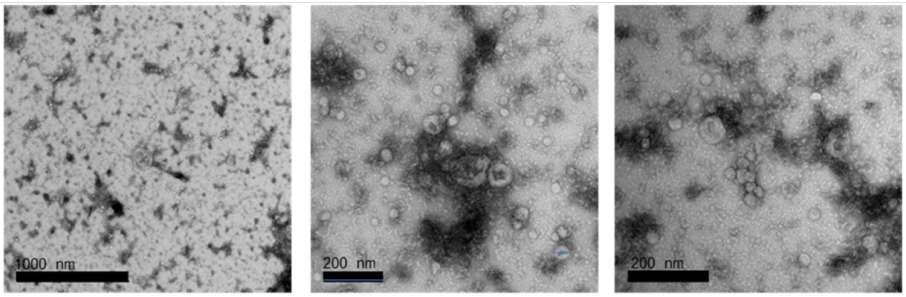


Figure 7.6: Representative transmission electron microscopy (TEM) of negatively staining *Drosophila* S2R+ EVs, purified from conditioned media containing 10% FBS. The typical cup shaped morphology can be observed, in the size range of 40 to 150 nm, typical for EVs.[39]

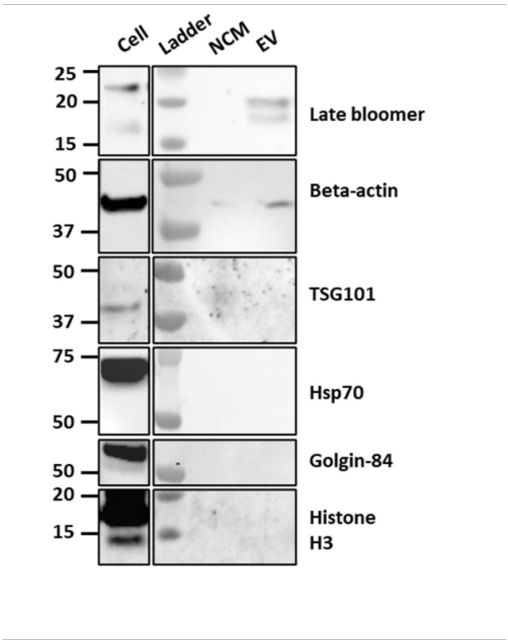
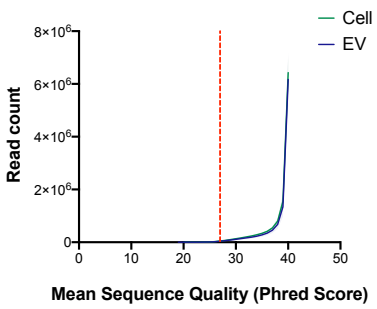


Figure 7.7: Representative Western Blots of EV marker proteins and contaminants for *Drosophila* S2R+ EV preparations (pooled SEC fractions 6-10). Cell culture media was supplemented with 10% FBS. 20 µg of protein was loaded (8 µg for Golgin-84 and actin blots). NCM = non-conditioned media control.[40]

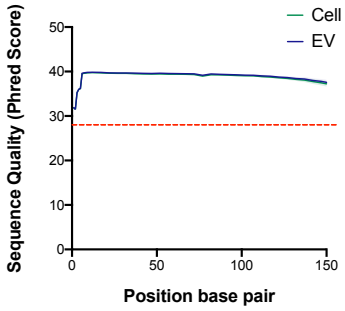
[39]Data generated by Elizabeth Dellar and Errin Johnson (Electron Microscopy Facility, Sir William Dunn School of Pathology)

[40]Data generated by Elizabeth Dellar

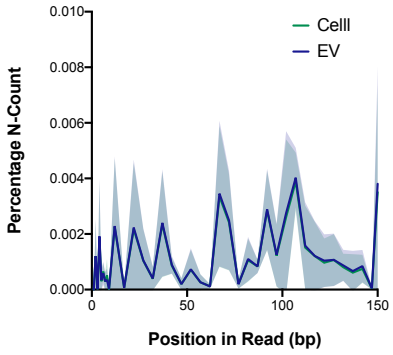
7.2 Supplementary Figures



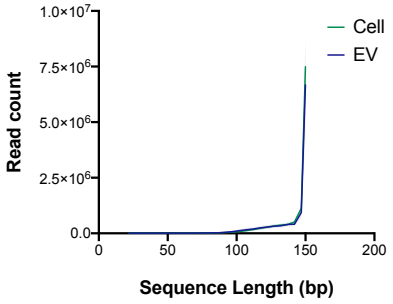
(i) The number of reads with appropriate sequence quality (Phred) scores



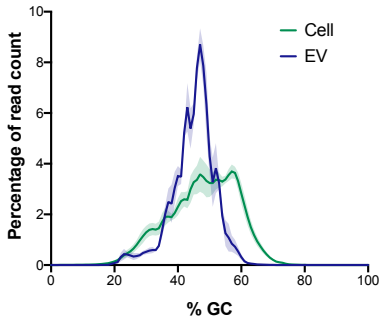
(ii) Sequence quality across each base position in the cell and EV reads.



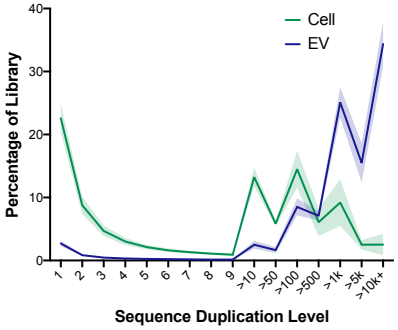
(iii) Per base N-Count.



(iv) Sequence length distribution.

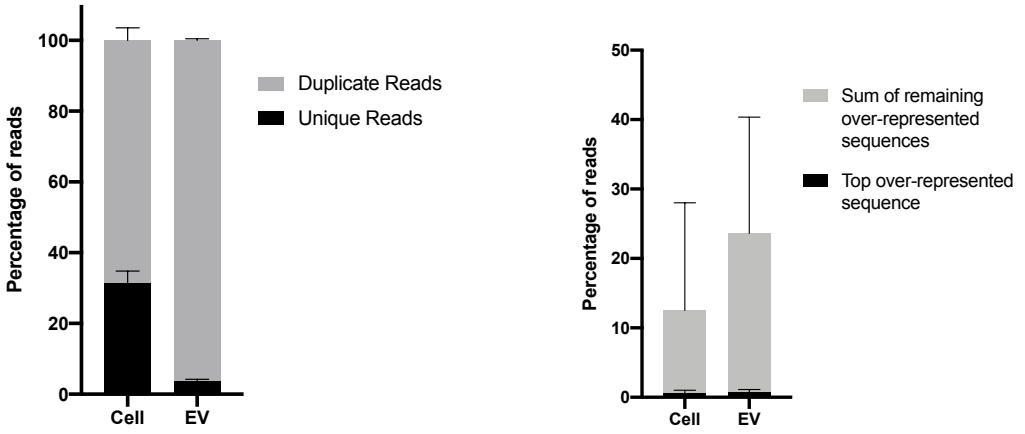


(v) Percentage GC content.



(vi) Sequence duplication levels.

Figure 7.8: MultiQC data for cellular and EV RNA-sequencing reads. Lines represent the average values across three biological replicates (paired end reads) and shaded areas represent standard deviation. (i and ii) All samples passed the 28 Phred score threshold. (iii) All samples passed the 5% Percentage N-Count threshold (N = when a sequencer is unable to confidently make a base call). (iv) All samples showed similar sequence length distributions. (v and vi) Varying percentage GC content and percentage duplication levels observed for cell and EV samples, likely due to the high abundance of rRNA in EV samples.



(i) Prevalence of duplicated reads.

(ii) Prevalence of overrepresented sequences.

Figure 7.9: MultiQC data for duplicated and overrepresented sequences. Shown are mean and standard deviation values across three biological replicates (paired end reads). High levels of duplication and overrepresentation are likely due to the high abundance of rRNA with EV samples.

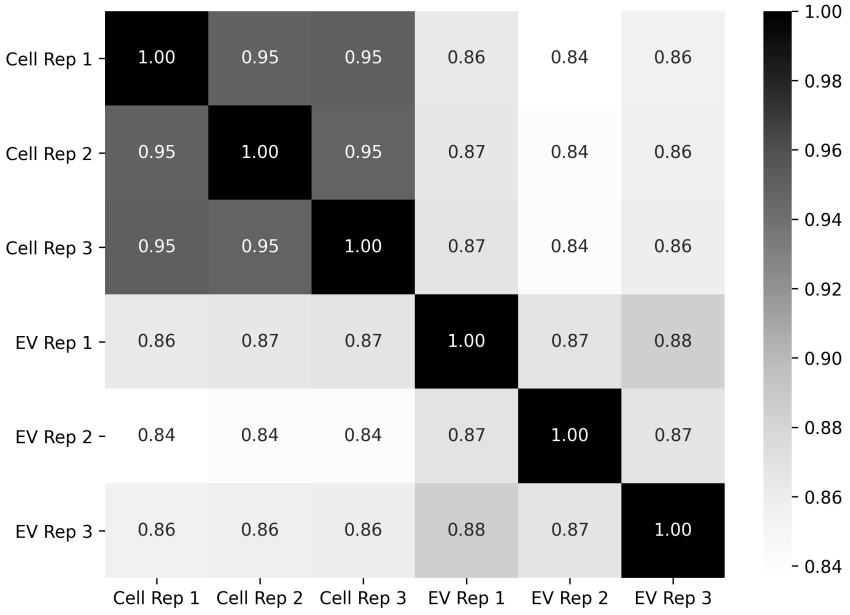


Figure 7.10: Pairwise Spearman correlation coefficients (ρ) obtained when comparing normalised read count values obtained for cell and EV RNA replicates.

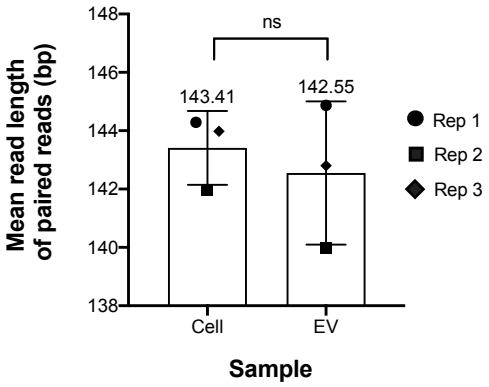


Figure 7.11: Mean read length of paired cell and EV RNA-sequencing reads. Shown are the values for three biological replicates (Rep.) and standard deviations. Highlighted above each bar are the mean values for each sample. Data generated via the Collect Alignment Summary Metrics Galaxy Tool.

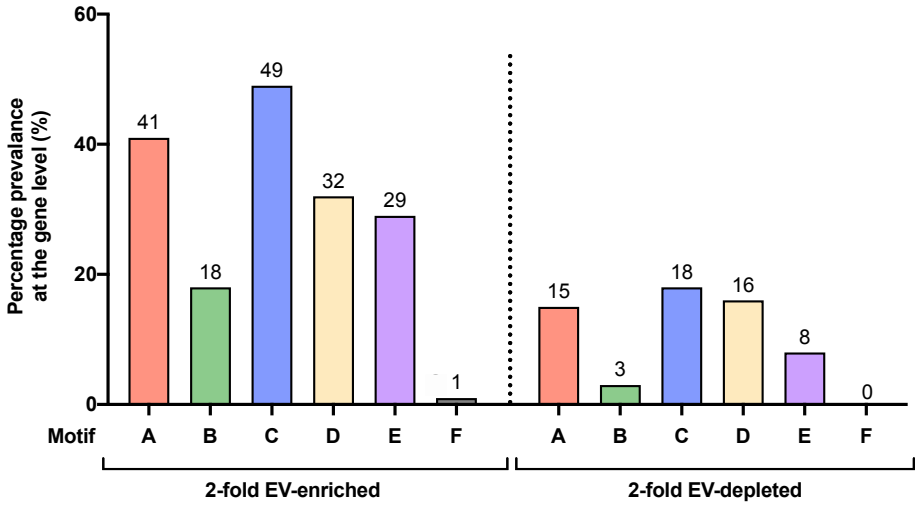


Figure 7.12: Percentage prevalence of motifs within the 2-fold significantly enriched or depleted genes in EVs relative to cell.[41]

[41] Data generated by Elizabeth Dellar

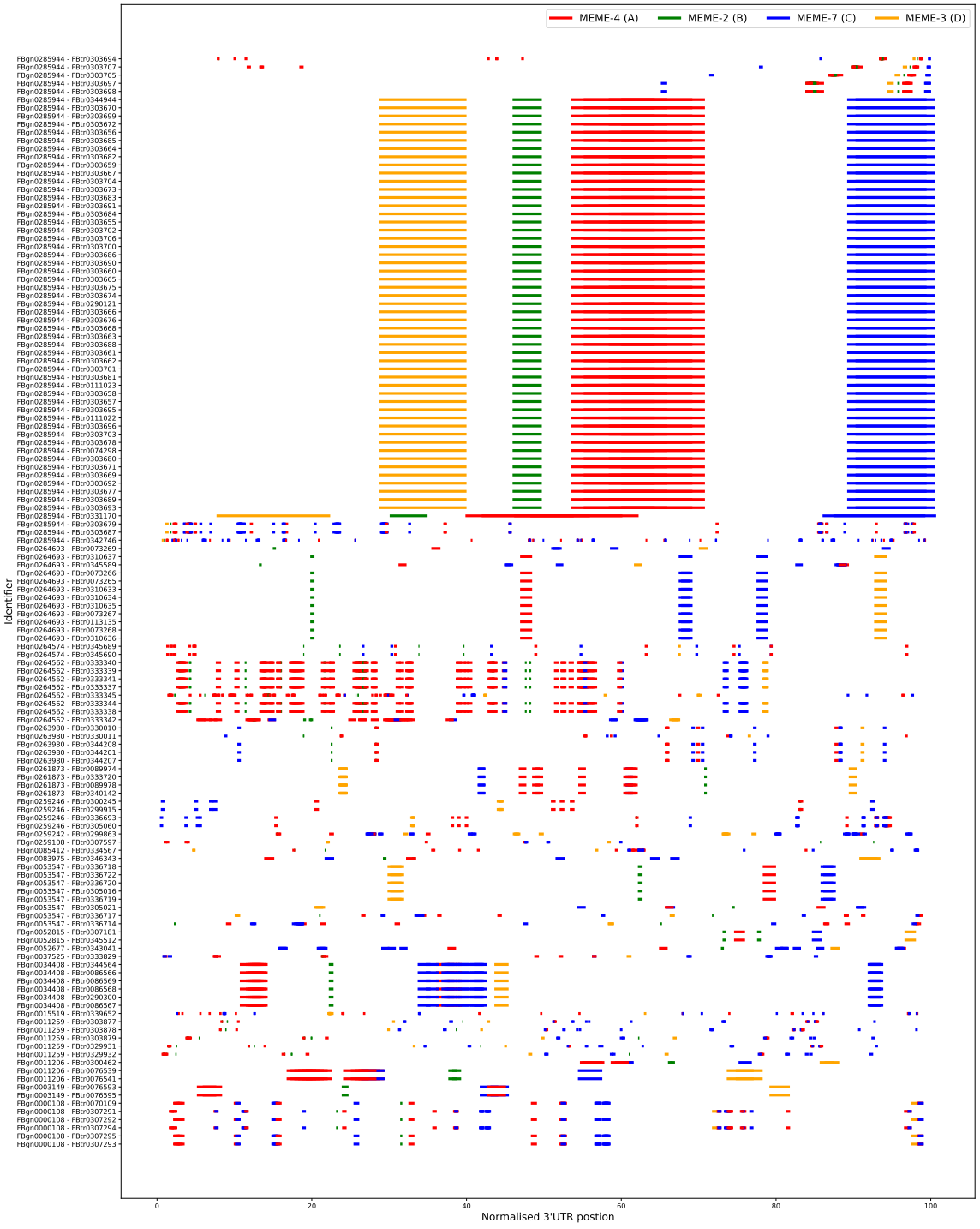


Figure 7.13: Motif start and end positions, normalised to total 3'UTR length displayed for all transcripts containing the A, B, C and D motifs. Motif A (Red), Motif B (Green), Motif C (Blue), Motif D (Yellow). Note, overlapping motif regions may be hidden.

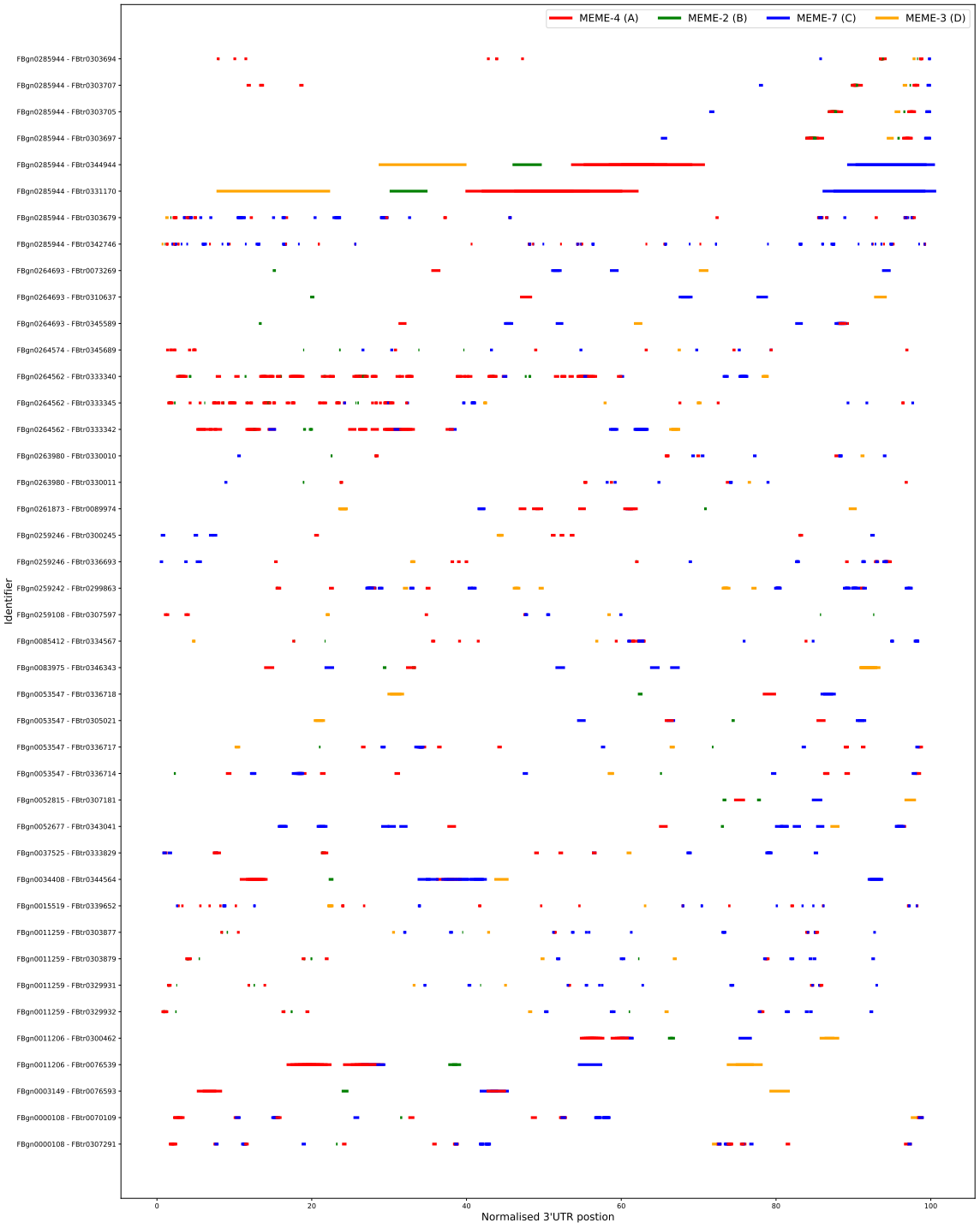


Figure 7.14: Motif start and end positions, normalised to total 3'UTR length displayed for shortlisted transcripts containing the A, B, C and D motifs. Motif A (Red), Motif B (Green), Motif C (Blue), Motif D (Yellow). Note, overlapping motif regions may be hidden. The shortlist was generated by selecting one transcript per gene presenting a single 3'UTR motif pattern.

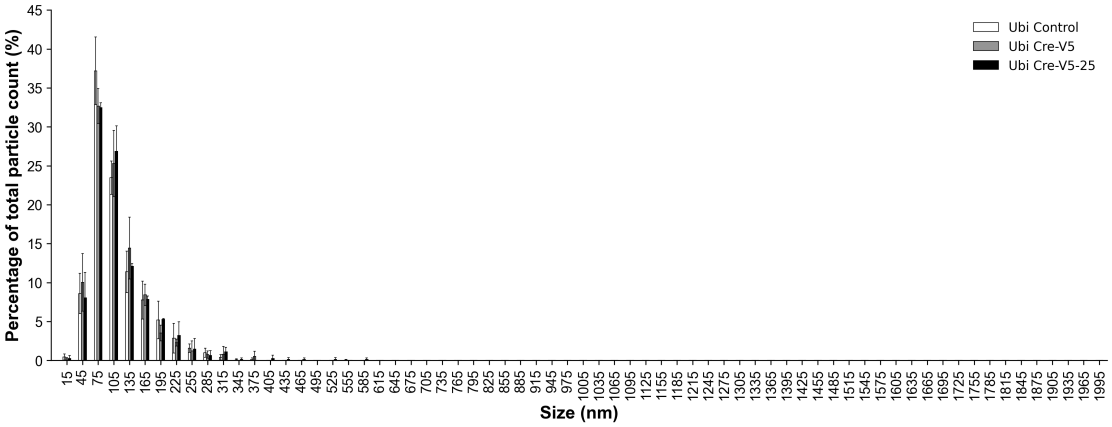


Figure 7.15: Size distribution profiles determined via ParticleMetrix NTA for EVs extracted from Ubi control, Ubi Cre-V5 and Ubi Cre-V5-25 transfection conditions. Data shown for pooled fractions 6 to 10 after SEC extraction. Mean ± standard deviation of three biological replicates is shown. Particle counts in each size window were normalised to the total number of particles per sample to give a percentage of total particles value (%).

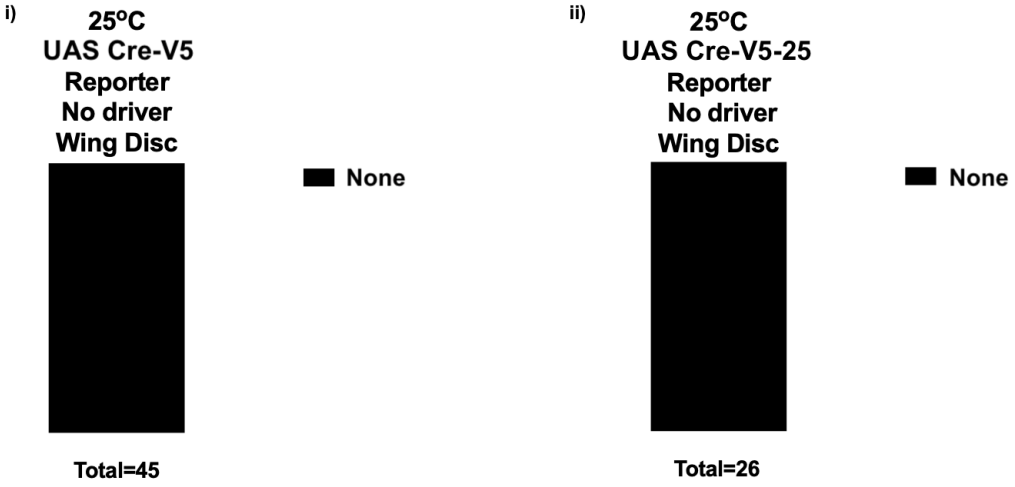


Figure 7.16: Proportion of imaginal wing discs presenting Reporter^{GFP+} signal in the IWD posterior (or none) in the absence of a Gal4 driver when using (i) UAS Cre-V5 and (ii) UAS Cre-V5-25. Genotype: UAS Cre/+; Actin-loxP-myrGFP/+; MKRS/+.

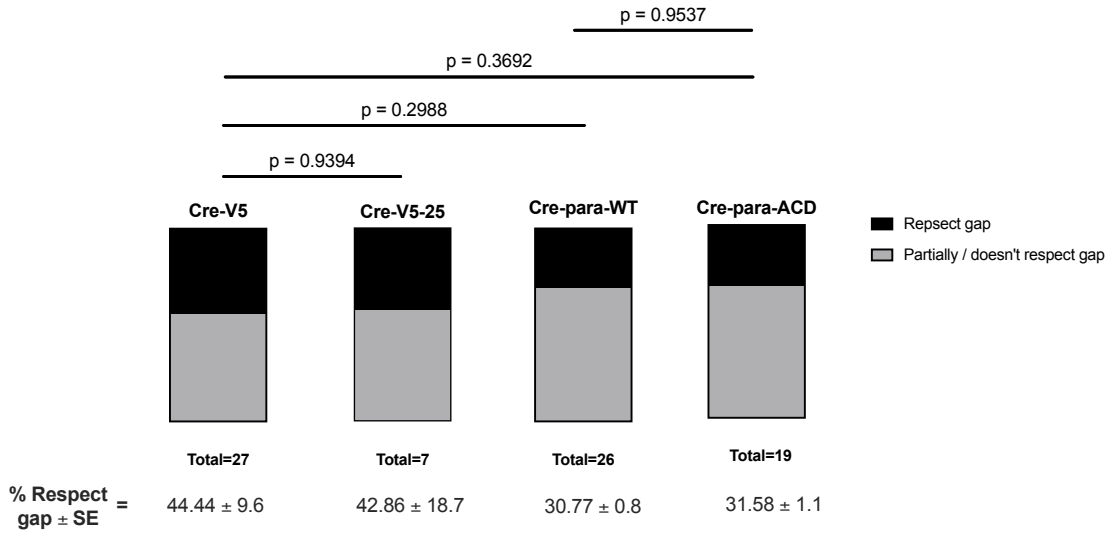


Figure 7.17: Proportion of Reporter^{GFP+} signal in the IWD posterior respecting the 'gap region' under Cre-V5, Cre-V5-25, Cre-V5-paraWT and Cre-V5-paraACD conditions. Genotype: UAS Cre/+; ptcGal4/Actin-loxP-myrGFP; TubGal80^{ts}, UAS HisRFP/MKRS. Statistical analysis was carried out as described in **Section 2.5.1**.

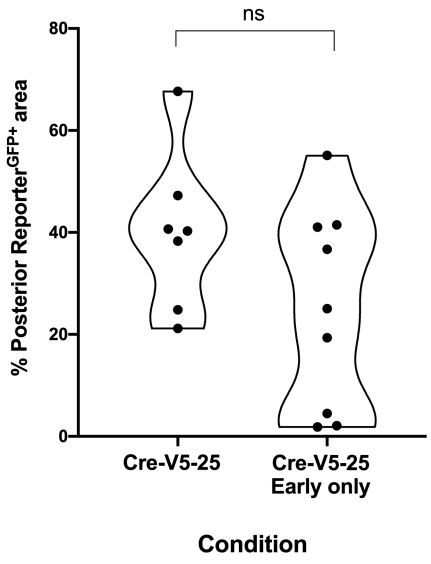


Figure 7.18: Percentage area of IWD posterior Reporter^{GFP+} signal when Cre-V5-25 is expressed throughout larval development or during early development stages only. Statistical significance values were determined via a two-tailed Mann-Whitney test. No significant difference was observed (p = 0.2523). Genotype: UAS Cre-V5-25/+; ptcGal4/Actin-loxP-myrGFP; TubGal80^{ts}, UAS HisRFP/MKRS.

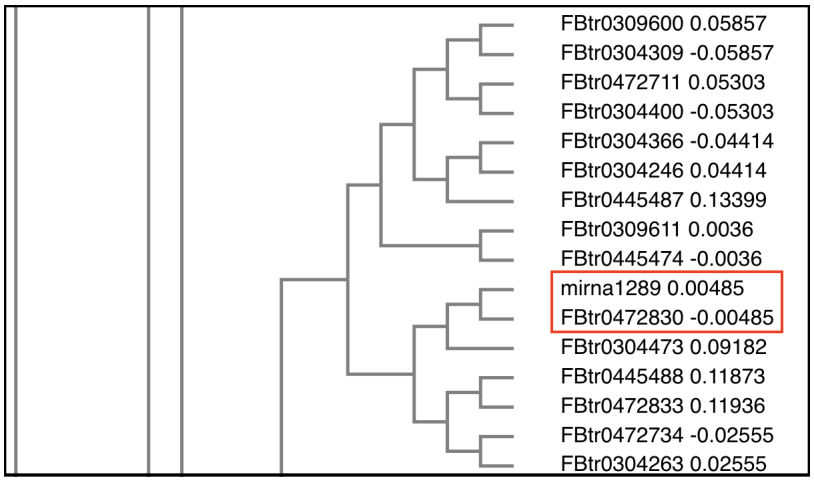


Figure 7.19: Snapshot from the neighbour-joining phylogenetic tree generated from a ClustalW alignment of human miR-1289 and all *Drosophila* miRNAs (from genome release r6.26).

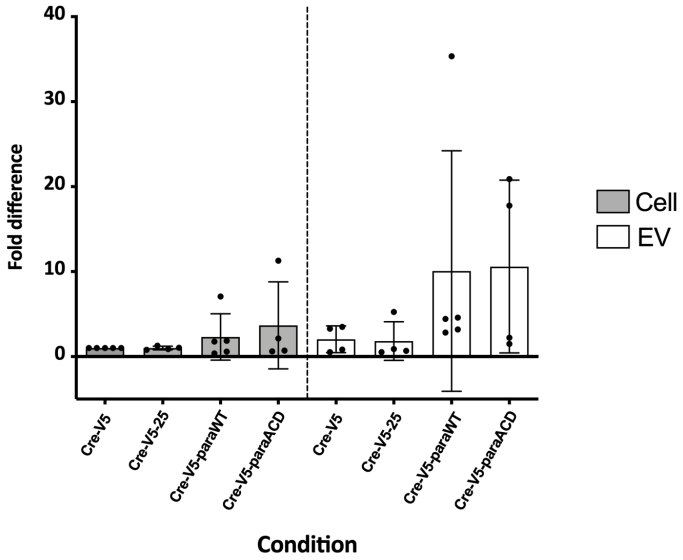


Figure 7.20: Fold differences observed in *cre* mRNA cellular and EV abundances for Cre-V5, Cre-V5-25, Cre-V5-paraWT and Cre-V5-paraACD conditions, determined via qPCR. Delta Ct values were determined using *cre* as the gene of interest and *rpl32* as the housekeeping gene. Fold difference values (Delta delta Ct) were determined by normalising all Delta Ct values to the matched cellular Cre-V5 replicate Delta Ct value.[42]

[42]Data generated by Elizabeth Dellar

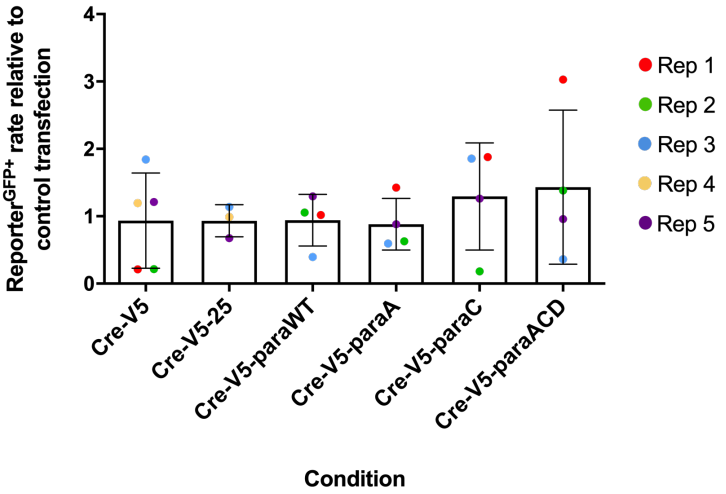


Figure 7.21: Rate of Reporter^{GFP+} signalling observed when Reporter transfected cells were co-cultured in a Transwell assay with cells transfected with various Ubi Cre constructs. Values were normalised to the background level of Reporter^{GFP+} signalling observed when reporter cells were co-cultured with a control Ubiquitin plasmid not producing Cre. Mean ± standard deviation is shown.[43]

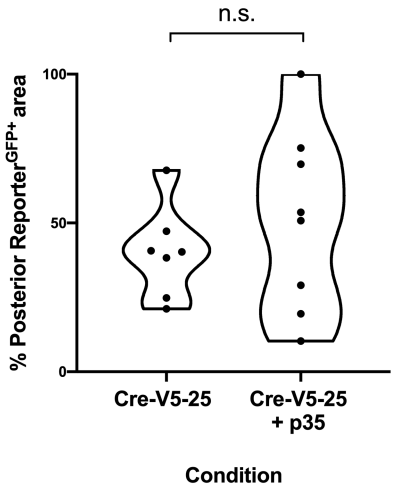


Figure 7.22: Percentage area of IWD posterior Reporter^{GFP+} signal when Cre-V5-25 is expressed in the absence or presence of p35. Statistical significance values determined via a two-tailed Mann-Whitney test. No significant difference was observed ($p = 0.4634$). Genotype: UAS Cre-V5-25/+; ptcGal4/Actin-loxP-myrGFP; TubGal80^{ts}, UAS HisRFP/UAS p35.

[43]Data generated in collaboration with Elizabeth Dellar

7.3 Supplementary Tables

Junction	Definition
Gap junctions (Connexins)	Tube-forming connections between the cytoplasm of two adjacent cells which allow molecules to be transferred along concentration gradients.
Ion channels	Gated pores which allow the transfer of specific ions through membranes.
Pannexins	Transmembrane channels connecting the intra- and extracellular space.
Tight junctions	Anchor neighbouring cells together to form a sheet and allow selective permeability.
Adherens	Cell junctions linked via cadherins to the actin cytoskeleton and microtubules to develop transcellular networks.
Desmosomes	Desmosomal cadherins, desmogleins (Dsgs) and desmocollins (Dscs), interact in the extracellular space to connect cells together and regulate signalling pathways.
Focal adhesions and hemidesmosomes	Utilise integrins to connect the actin cytoskeleton on the inside of cells to the ECM, aiding signal transduction.

Table 7.1: Summary of cell junctions utilised for cell communication via the transfer of molecules or transduction of signals [2]. ECM = Extracellular matrix.

Mechanism	Description	Examples
Clathrin-mediated endocytosis (CME)	Involves the formation of clathrin-coated pits (CCPs) which hold the cargo to be internalised, followed by the formation of clathrin-coated vesicles (CCVs). These CCVs deform the membrane, folding it inward before budding off [206].	Treatment of ovarian cancer cells or macrophages with chlorpromazine to inhibit clathrin-mediated endocytosis [566] decreased EV uptake [295, 327]. Dynamin 2 (a GTPase) plays a key role [567, 568], with its inhibition preventing almost all EV uptake in a number of studies [283, 327, 569].
Caveolin-dependent endocytosis (CDE)	CDE also involves the formation of invaginations termed caveolae [206]. Caveolae are a special form of lipid-raft, enriched in cholesterol, sphingolipids and caveolins [570–573]. Although the process of caveolae internalisation is clathrin independent, it depends on dynamin, with an additional sensitivity to cholesterol depletion [570].	Caveolin-1 (CAV1) is a protein required for caveolae formation [302], with the small interfering RNA (siRNA) mediated knockdown of CAV1 in CNE1 cancer cells shown to significantly reduce exosome internalisation [574]. However, endocytosis has been shown to be negatively regulated by caveolin-1 using CAV1 null mice [280], highlighting possible interplay and redundancy between uptake pathways.
Phagocytosis	Harnessed by cells to uptake larger cargoes, with particles presenting a diameter of 85 nm shown to be internalised via this mechanism [575].	Phosphoinositide 3-kinases (PI3Ks) inhibition, enzymes with vital roles in phagocytosis [576], resulted in a dose-dependent decrease in EV uptake. Co-localisation was also observed between internalised EVs and fluorescent phagosome markers [327]. Phosphatidylserine (PS), a phospholipid enriched on the outer leaflet of EVs, may also facilitate cellular uptake, as treatment of cells with TIM-4 antibody (blocking the TIM-4 receptor for PS) reduced EV uptake via phagocytosis [327]. Dynamin 2 has also been implicated, with a Dynamin 2 binding protein, Amphiphysin II, recruited to phagocytic invaginations in a PI3K dependent manner [576, 577].

Table 7.2: Table continued on the next page.

Table continued from the previous page.

Mechanism	Description	Examples
Macropinocytosis	Involves membrane ruffling and cellular extensions which envelop part of the extracellular environment, subsequently internalising it [578]. Utilises the actin cytoskeleton to form a cup (a macropinosome) which is pinched off for internalisation. Actin-independent macropinocytosis-like processes have also been described [579].	Small GTPases such as Rac1 and Cdc42, as well as Rab5, ARF6, cholesterol and PI3K are involved in macropinosome formation [580–583]. Exosome uptake by microglia was shown to be actin and Rac1-GTPase dependent, with Na+/H+- exchange activity (involved in membrane ruffling), dynamin and a low pH also required [283].
Lipid rafts	Protein, cholesteroland sphingolipid rich regions within the EV surface are involved in cellular uptake [206, 570]. Flotillins are a family of proteins which associate with lipid-rafts [584, 585] and have previously been associated with endocytosis and EV uptake [586].	EV uptake was significantly reduced when EV-producing dendritic cells were treated with fumonisin B1 and N-butyldeoxynojirimycin hydrochloride, inhibitors which reduce glycosphingolipid levels within the plasma membrane [207]. Recipient cell lipid-raft regions are also be important for EV uptake, with disrupted EV uptake observed when target cells were pretreated with cholesterol reducing molecules [280, 295, 298, 327]. Anti-flotillin-1 antibody treatment of EVs significantly inhibited EV uptake by C26 or B16BL6 target cells by 26.7% and 56.8%, respectively [587].
Receptor binding	Proteins on the EV surface can act as ligands to receptors on recipient cells. This is particularly relevant in the context of immunity, with reports as early as 1996 highlighting the presence of antigen-presenting EVs [61, 588–590].	Pro-inflammatory (Th1) activated neural stem cell derived EVs contained surface interferon gamma receptor 1 (Ifngr1) which carried interferon gamma (IFN-γ) to to activate Stat1 signalling in recipient cells [591].
	EV RNAs have also been shown to act as ligands to recipient cell receptors to induce a signalling change.	In addition to the immunomodulatory role of internalised EV miRNAs [319, 592], EV surface miRNAs can also bind Toll-like receptors in immune cells to trigger an inflammatory response which could lead to tumour growth and metastasis [593].

Table 7.2: Mechanisms of EV uptake or signal transduction in recipient cells.

Component	Subunits	Function	Summary	References
ESCRT-0	Hrs (Vps27) and Stam (HseI) heterodimer	Ubiquitin-interacting domains allow these components to cluster cargo. Hrs binds the endosomal lipid phosphatidylinositol 3-phosphate (PtdIns(3)P) through a FYVE zinc-finger domain. Hrs is indirectly involved in ILV formation as it binds the ESCRT-I subunit, TSG101, recruiting it to the endosomal membrane. This is the least conserved ESCRT and it is not found in plants; other proteins are likely to function in parallel or in place of these proteins.	Bind and concentrate ubiquitylated cargo. Recruit ESCRT-I for ILV formation.	[90–94, 594–602]
ESCRT-I	TSG101 (Vps23), Vps28, Vps37 and Mvb12 heterote-tramer	This was the first ESCRT to be described. It acts as an assembly hub. It contains 3 main regions: a headpiece (binding ESCRT-II), a stalk and an endpiece (binding ubiquitin and ESCRT-0).	Bind ubiquitylated cargo. Recruiting ESCRT-II. Potential role in membrane deformation.	[95, 603–605]
ESCRT-II	Vps36, Vps22 and two Vps25	The ESCRT-II subunit Vps36 contains a GLUE domain (GRAM-like ubiquitin-binding in Eap45) which binds ESCRT-I (Vps28), ubiquitin and the endosomal lipid PtdIns(3)P. This component links ESCRT-I with membranes.	Bind ubiquitylated cargo. Linking ESCRT-I with membranes. Recruiting ESCRT-III.	[101, 606–612]
ESCRT-III	Vps20, Vps32, Vps24 and Vps2	The ESCRT-III subunit Vps20 binds ESCRT-II (Vps25). Vps20 forms filamentous oligomers of Vps32 which is capped by Vps24. Vps2 associates with the Vps24 cap to recruit the ATPase Vps4.	Membrane deformation.	[99, 101, 102]

Table 7.3: Table continued on the next page.

Table continued from the previous page.

Component	Subunits	Function	Summary	References
Vps4	Vps4 and its cofactor VTA1 cofactor	Vps4 enters membrane invaginations generated by ESCRT-III to disassemble ESCRT-III filaments. This means the ESCRT-III subunits do not enter ILVs, allowing them to be reused.	Disassembly of ESCRT-III complex at membrane invaginations.	[97]
ALIX	N/A	ALIX functions as an alternative branch to ESCRT-I and -II, but feeds into the common ESCRT-III and Vps4 machinery. The syndecan heparan sulphate proteoglycans and their cytoplasmic adaptor syntenin control the formation of ILVs / exosomes via syntenin directly interacting with ALIX.	ILV formation.	[66, 87, 103, 104]
SNX3	N/A	Binds PtdIns(3)P and is involved in ILV formation via membrane deformation but not receptor loading.	Membrane deformation.	[613]

Table 7.3: ESCRT components and accessory factors and their involvement in ILV formation and cargo loading [87, 89, 112, 113, 614].

	Mean read length (paired reads)	Reads mapped in proper pairs
EV Rep 1	144.866	263086
EV Rep 2	139.976	151904
EV Rep 3	142.801	197344
EV Average	142.548	204111.333
EV SD	2.455	55899.078
Cell Rep 1	144.291	13708476
Cell Rep 2	141.957	13670232
Cell Rep 3	143.978	9265098
Cell Average	143.409	12214602
Cell SD	1.267	2554416.966

Table 7.4: Read alignment metrics for each cellular and EV RNA replicate, together with means and standard deviations.

Sidak's multiple comparisons test	Mean Diff.	95% CI of diff.	Significant?	Summary
CDS	5.233	1.794 to 8.673	Yes	**
3'UTR	-1.300	-4.740 to 2.140	No	ns
5'UTR	-1.133	-4.573 to 2.306	No	ns
Intron	-2.300	-5.740 to 1.140	No	ns

Table 7.5: Sidak's multiple comparisons test to compare the difference in relative abundance of genes aligning to particular gene regions in our cell and EV samples. Regions include coding regions (CDS), 3-prime untranslated regions (3'UTRs), 5-prime untranslated regions (5'UTR) or intronic regions (intron) (**Figure 3.5**).

	Percentage of transcripts (%)		
Transcript type	**Cell only**	**Cell and EV**	**EV only**
mRNA	77.31	96.33	84.00
ncRNA	18.41	1.80	10.18
snoRNA	2.39	1.40	2.55
pseudogene	1.29	0.28	1.82
snRNA	0.50	0.16	1.09
rRNA	0.10	0.03	0.00
tRNA	0.00	0.00	0.36

Table 7.6: Percentage of transcripts within each transcript type category for 'Cell only', 'Cell and EV' and 'EV only' active gene groupings.

GO Cluster	GO ID	Term	p-value
1	GO:0009888	tissue development	4.88E-06
1	GO:0060429	epithelium development	3.30E-05
1	GO:0048729	tissue morphogenesis	2.72E-05
1	GO:0001736	establishment of planar polarity	0.004610399
1	GO:0001738	morphogenesis of a polarized epithelium	0.000470998
1	GO:0002009	morphogenesis of an epithelium	1.47E-05
1	GO:0090175	regulation of establishment of planar polarity	0.00963801
1	GO:1905330	regulation of morphogenesis of an epithelium	0.003336196
1	GO:0042067	establishment of ommatidial planar polarity	0.006569344
1	GO:0090251	protein localization involved in establishment of planar polarity	0.009806784
2	GO:0007476	imaginal disc-derived wing morphogenesis	0.002756615
2	GO:0016335	morphogenesis of larval imaginal disc epithelium	0.009806784
2	GO:0007472	wing disc morphogenesis	0.00154248
2	GO:0007560	imaginal disc morphogenesis	0.002544199
2	GO:0007444	imaginal disc development	0.00373775
2	GO:0035220	wing disc development	0.000760214
2	GO:0035239	tube morphogenesis	0.001873507
2	GO:0060562	epithelial tube morphogenesis	0.003160697
3	GO:0071695	anatomical structure maturation	0.008049497
3	GO:0021700	developmental maturation	0.00815507
3	GO:0032502	developmental process	0.00017422
3	GO:0007525	somatic muscle development	4.37E-05
3	GO:0061061	muscle structure development	0.000306918
3	GO:0048856	anatomical structure development	3.41E-05
3	GO:0007164	establishment of tissue polarity	0.004610399
3	GO:0009653	anatomical structure morphogenesis	9.23E-07
3	GO:0048646	anatomical structure formation involved in morphogenesis	0.008185881
4	GO:0009887	animal organ morphogenesis	4.16E-05
4	GO:0048563	post-embryonic animal organ morphogenesis	0.002544199
4	GO:0048513	animal organ development	0.000119006
4	GO:0048569	post-embryonic animal organ development	0.003006101
4	GO:0007399	nervous system development	0.000831925
4	GO:0048731	system development	7.33E-05
5	GO:0048598	embryonic morphogenesis	0.00271281
5	GO:0009790	embryo development	3.42E-06
5	GO:0009792	embryo development ending in birth or egg hatching	0.006409799
5	GO:2000026	regulation of multicellular organismal development	0.007544127
5	GO:0002165	instar larval or pupal development	0.00299108
5	GO:0009791	post-embryonic development	0.001705712
5	GO:0007552	metamorphosis	0.007657807
5	GO:0007275	multicellular organism development	1.10E-05
5	GO:0035295	tube development	0.002399918
6	GO:0035120	post-embryonic appendage morphogenesis	0.001775157
6	GO:0009886	post-embryonic animal morphogenesis	0.002786145
6	GO:0048707	instar larval or pupal morphogenesis	0.004259099
6	GO:0048736	appendage development	0.001267554
6	GO:0048737	imaginal disc-derived appendage development	0.001147809
6	GO:0035107	appendage morphogenesis	0.002520699
6	GO:0035114	imaginal disc-derived appendage morphogenesis	0.002368526
7	GO:0003012	muscle system process	0.001508858
7	GO:0032501	multicellular organismal process	0.000399569
8	GO:0009880	embryonic pattern specification	0.000768879
8	GO:0007350	blastoderm segmentation	0.001208782
8	GO:0035290	trunk segmentation	0.004178459
8	GO:0009798	axis specification	0.008049497
8	GO:0035282	segmentation	0.00072577
8	GO:0009953	dorsal/ventral pattern formation	0.006462538
8	GO:0009952	anterior/posterior pattern specification	0.007193511
8	GO:0003002	regionalization	0.000135599
8	GO:0007389	pattern specification process	8.85E-06
9	GO:0031327	negative regulation of cellular biosynthetic process	0.003001433
9	GO:2000113	negative regulation of cellular macromolecule biosynthetic process	0.004440528
9	GO:0051172	negative regulation of nitrogen compound metabolic process	0.008533698
9	GO:0010629	negative regulation of gene expression	0.008071118
9	GO:0009890	negative regulation of biosynthetic process	0.003001433
9	GO:0010558	negative regulation of macromolecule biosynthetic process	0.004440528
10	GO:0046068	cGMP metabolic process	0.005058152
10	GO:0097502	mannosylation	0.00271517
10	GO:0035268	protein mannosylation	0.00271517
10	GO:0035269	protein O-linked mannosylation	0.00212363
11	GO:0051179	localization	2.34E-05
11	GO:0051234	establishment of localization	0.002511185
11	GO:0006812	cation transport	0.005302661
11	GO:0030001	metal ion transport	0.005558522
11	GO:0006810	transport	0.00531681
11	GO:0006811	ion transport	0.005365437
12	GO:0051674	localization of cell	0.003957499
12	GO:0016477	cell migration	0.004106896
12	GO:0048870	cell motility	0.003446895

Table 7.7: Table continued on the next page.

Table continued from the previous page.

GO Cluster	GO ID	Term	p-value
13	GO:0030029	actin filament-based process	5.25E-08
13	GO:0030048	actin filament-based movement	0.001699479
13	GO:0070252	actin-mediated cell contraction	0.00291027
13	GO:0007154	cell communication	2.77E-05
13	GO:0038001	paracrine signaling	0.009806784
13	GO:0007017	microtubule-based process	0.002590468
13	GO:0006928	movement of cell or subcellular component	0.000327155
13	GO:0007018	microtubule-based movement	0.003916003
13	GO:0007155	cell adhesion	0.009632912
13	GO:0007163	establishment or maintenance of cell polarity	0.00013562
13	GO:0030010	establishment of cell polarity	0.001848714
14	GO:0023051	regulation of signaling	0.000892607
14	GO:0023052	signaling	1.22E-05
14	GO:2000574	regulation of microtubule motor activity	0.009806784
14	GO:0032879	regulation of localization	0.003658862
14	GO:0065008	regulation of biological quality	0.002127054
14	GO:0050789	regulation of biological process	0.000873265
14	GO:0065007	biological regulation	0.000281544
15	GO:0050896	response to stimulus	0.004777381
15	GO:0051716	cellular response to stimulus	0.007782795
15	GO:0045743	positive regulation of fibroblast growth factor receptor signaling pathway	0.009806784
15	GO:0007168	receptor guanylyl cyclase signaling pathway	0.009806784
15	GO:0019934	cGMP-mediated signaling	0.004178459
15	GO:0007165	signal transduction	0.000103266
15	GO:0035556	intracellular signal transduction	0.006056568
16	GO:0050709	negative regulation of protein secretion	0.004810986
16	GO:0031333	negative regulation of protein-containing complex assembly	0.002523084
16	GO:0010646	regulation of cell communication	0.000892607
16	GO:0048523	negative regulation of cellular process	0.004879748
16	GO:0050794	regulation of cellular process	0.000675252
17	GO:0007314	oocyte anterior/posterior axis specification	0.009676551
17	GO:0007308	oocyte construction	0.006833561
17	GO:0007309	oocyte axis specification	0.005628669
18	GO:0003382	epithelial cell morphogenesis	0.00610815
18	GO:0003383	apical constriction	0.002172807
19	GO:0042692	muscle cell differentiation	0.00071587
19	GO:0051146	striated muscle cell differentiation	0.000235626
19	GO:0055001	muscle cell development	0.000413125
19	GO:0055002	striated muscle cell development	0.00037275
19	GO:0022008	neurogenesis	0.006234878
19	GO:0048699	generation of neurons	0.00527236
19	GO:0048468	cell development	0.000517392
19	GO:0048469	cell maturation	0.007193511
19	GO:0030855	epithelial cell differentiation	0.001581707
19	GO:0030154	cell differentiation	0.000142699
19	GO:0048869	cellular developmental process	0.000171427
20	GO:0051653	spindle localization	0.001030781
20	GO:0051293	establishment of spindle localization	0.001030781
20	GO:0051294	establishment of spindle orientation	0.003336196
20	GO:0000132	establishment of mitotic spindle orientation	0.001324647
20	GO:0040001	establishment of mitotic spindle localization	0.003336196
21	GO:0010927	cellular component assembly involved in morphogenesis	0.002051001
21	GO:0071689	muscle thin filament assembly	0.007109777
21	GO:0030239	myofibril assembly	8.94E-05
21	GO:0045214	sarcomere organization	0.001175374
22	GO:0097435	supramolecular fiber organization	1.05E-06
22	GO:1990709	presynaptic active zone organization	0.005058152
22	GO:0007015	actin filament organization	0.000477089
22	GO:0030036	actin cytoskeleton organization	3.68E-07
22	GO:0031032	actomyosin structure organization	8.77E-06
22	GO:0072393	microtubule anchoring at microtubule organizing center	0.007109777
22	GO:0007010	cytoskeleton organization	3.13E-08
22	GO:0048789	cytoskeletal matrix organization at active zone	0.009806784

Table 7.7: Clustered biological process gene ontology terms significantly enriched in 'EV+' genes.

GO Cluster	GO ID	Term	p-value
1	GO:0043292	contractile fiber	0.003962136
1	GO:0005874	microtubule	0.007454451
1	GO:0099513	polymeric cytoskeletal fiber	0.000726677
2	GO:0005871	kinesin complex	0.002397852
2	GO:0005875	microtubule associated complex	0.007087053
3	GO:0005856	cytoskeleton	6.54E-07
3	GO:0015629	actin cytoskeleton	0.000634915
4	GO:1990351	transporter complex	0.005750523
4	GO:0034702	ion channel complex	0.005193056
4	GO:1902495	transmembrane transporter complex	0.003039497
5	GO:0098797	plasma membrane protein complex	0.002913995
5	GO:0097060	synaptic membrane	0.0031919
5	GO:0016324	apical plasma membrane	0.00166179
5	GO:0005886	plasma membrane	1.11E-05
5	GO:0016327	apicolateral plasma membrane	0.007190686
5	GO:0098590	plasma membrane region	0.00165332
6	GO:0005887	integral component of plasma membrane	0.000932027
6	GO:0031226	intrinsic component of plasma membrane	0.000230212
7	GO:0016020	membrane	0.002402786
7	GO:0031224	intrinsic component of membrane	0.009930629
8	GO:0043197	dendritic spine	0.006408522
8	GO:0042995	cell projection	0.000307724
8	GO:0120025	plasma membrane bounded cell projection	0.00027914
8	GO:0044309	neuron spine	0.006408522
8	GO:0030424	axon	0.002188499
8	GO:0043005	neuron projection	0.000319774
9	GO:0005915	zonula adherens	0.006408522
9	GO:0043296	apical junction complex	0.004451664
9	GO:0070160	tight junction	0.006931768
9	GO:0005911	cell-cell junction	0.000983019
9	GO:0005912	adherens junction	0.002830673
10	GO:0099568	cytoplasmic region	0.001448469
10	GO:0005938	cell cortex	0.00013793
10	GO:0099738	cell cortex region	0.000975101
11	GO:0099080	supramolecular complex	0.000227206
11	GO:0099081	supramolecular polymer	6.82E-06
11	GO:0099512	supramolecular fiber	5.63E-06
12	GO:0071944	cell periphery	1.84E-07
12	GO:0036477	somatodendritic compartment	0.002746178
12	GO:0045177	apical part of cell	0.000886401
13	GO:0045202	synapse	0.000132134
13	GO:0098794	postsynapse	0.005193056
13	GO:0030054	cell junction	3.90E-07
13	GO:0070161	anchoring junction	0.00167869

Table 7.8: Clustered cellular component gene ontology terms significantly enriched in 'EV+' genes.

GO Cluster	GO ID	Term	p-value
1	GO:0005539	glycosaminoglycan binding	0.00194585
1	GO:0008201	heparin binding	0.009765588
2	GO:0008017	microtubule binding	1.25E-07
2	GO:0015631	tubulin binding	1.94E-06
3	GO:0005488	binding	0.008704284
3	GO:0005515	protein binding	0.000204699
4	GO:0008093	cytoskeletal anchor activity	0.006503711
4	GO:0051015	actin filament binding	0.000730237
4	GO:0003779	actin binding	0.000236472
4	GO:0008092	cytoskeletal protein binding	4.25E-08
5	GO:0099529	neurotransmitter receptor activity, regulation of postsynaptic membrane potential	0.00531805
5	GO:1904315	transmitter-gated ion channel activity, regulation of postsynaptic membrane potential	0.00531805
6	GO:0008273	calcium, potassium:sodium antiporter activity	0.004397909
6	GO:0005216	ion channel activity	0.00925751
6	GO:0005261	cation channel activity	0.000991379
7	GO:0003774	motor activity	0.000907533
7	GO:0003777	microtubule motor activity	0.009454491
8	GO:0048018	receptor ligand activity	0.00186797
8	GO:0030545	receptor regulator activity	0.002962574
8	GO:0030546	signaling receptor activator activity	0.002505378

Table 7.9: Clustered molecular function gene ontology terms significantly enriched in 'EV+' genes.

	Motif	MEME-4 (A)	MEME-2 (B)	MEME-7 (C)	MEME-3 (D)	MEME-1	MEME-21
	Bit score	13	10.2	15.8	15.7	11.4	16.8
EV only plus EV enriched	**Unique gene hits**	118	35	126	97	88	19
	Percentage of genes with motif (%)	37.58	11.15	40.13	30.89	28.03	6.05
	Unique transcript hits	476	185	508	390	352	91
	Percentage of transcripts with motif (%)	47.70	18.54	50.90	39.08	35.27	9.12
Cell only	**Unique gene hits**	37	13	38	44	37	0
	Percentage of genes with motif (%)	4.77	1.68	4.90	5.67	4.77	0.00
	Unique transcript hits	87	33	87	86	72	0
	Percentage of transcripts with motif (%)	6.18	2.35	6.18	6.11	5.12	0.00

Table 7.10: Percentage prevalence of motifs within the 'EV only plus EV enriched' and 'Cell only' active gene lists. Total number of cell only active genes = 776. Total number of EV only plus EV enriched active genes = 314. Total number of Cell only active transcripts = 1407. Total number of EV only plus EV enriched active transcripts = 998.

Dunn's multiple comparisons test	Mean rank diff.	Significant?	Symbol	Adjusted P Value	
MEME-4 (A) vs. MEME-2 (B)	64.82	Yes	*	0.0375	A-B
MEME-4 (A) vs. MEME-7 (C)	-89.23	Yes	****	<0.0001	A-C
MEME-4 (A) vs. MEME-3 (D)	-94.57	Yes	**	0.0081	A-D
MEME-2 (B) vs. MEME-7 (C)	-154.0	Yes	****	<0.0001	B-C
MEME-2 (B) vs. MEME-3 (D)	-159.4	Yes	****	<0.0001	B-D
MEME-7 (C) vs. MEME-3 (D)	-5.340	No	ns	>0.9999	C-D

Table 7.11: Dunn's multiple comparisons test to compare the distribution of normalised 3'UTR position hit for each motif (**Figure 3.14**).

Motif	Skewness	Standard Error of Skewness (SES)
A	0.4169	0.1431
B	0.9893	0.2426
C	-0.3411	0.1513
D	-0.3201	0.3163

Table 7.12: Skewness and standard error values determined for the distribution of normalised 3'UTR position hit for each motif (**Figure 3.14**).

Dunn's multiple comparisons test	Mean rank diff.	Significant?	Symbol	Adjusted P Value	
MEME-4 (A) vs. MEME-2 (B)	-166.6	Yes	****	<0.0001	A-B
MEME-4 (A) vs. MEME-7 (C)	49.67	Yes	*	0.0225	A-C
MEME-4 (A) vs. MEME-3 (D)	-248.9	Yes	****	<0.0001	A-D
MEME-2 (B) vs. MEME-7 (C)	216.3	Yes	****	<0.0001	B-C
MEME-2 (B) vs. MEME-3 (D)	-82.28	No	ns	0.0805	B-D
MEME-7 (C) vs. MEME-3 (D)	-298.6	Yes	****	<0.0001	C-D

Table 7.13: Dunn's multiple comparisons test to compare the percentage of paired bases per motif occurrence (**Figure 3.17**).

Drosophila RBP	Significance	Motif	Flybase ID	Human Ortholog (Flybase)	Ortholog score (/15)	ExoCarta mRNA EV loaded	EzoCarta Protein EV loaded
CNOT4	$p < 0.001$	C	FBgn0051716	CNOT4	9	[68, 615]	NA
SM	$p < 0.001$	C	FBgn0003435	HNRNPL	12	NA	[179, 470, 471]
				PABPC1	1	[615]	[179, 470, 471, 591, 616–621]
				PABPC3	1	[615]	[470]
				PABPC4	1	[68, 615]	[179, 470, 471, 618, 619]
				PTBP1	1	[615]	[40, 470, 471, 618–620, 622]
				PTBP3	1	NA	[470, 620]
FMR1	$p < 0.01$	C	FBgn0028734	FXR1	12	NA	[179, 470, 619, 620]
				FXR2	11	NA	[179]
SXL	$p < 0.01$	D	FBgn0264270	CIRBP	1	[615]	[618]
				EIF4B	1	[615, 623]	[471, 621, 624]
				HNRNPA0	1	[615, 623]	[179, 471, 619, 620]
				HNRNPA1	1	[68]	[471, 617–620, 622, 625, 626]
				HNRNPA2B1	1	[615]	[179, 470, 471, 617–620, 622, 624, 626–628]
				HNRNPA3	1	[68, 615]	[179, 471, 619, 620, 624, 626]
				HNRNPAB	1	[68, 615]	[179, 471, 617, 626]
				HNRNPC	1	[615]	[179, 470, 471, 616, 618, 620, 622, 629]
				HNRNPCL2	1	NA	[179, 470]
				HNRNPCL3	1	NA	[470]
				HNRNPCL4	1	NA	[470]
				HNRNPD	1	[615]	[179, 470, 471, 618–620]
				HNRNPDL	1	NA	[179, 470, 471, 618–620]
				RALY	1	NA	[471, 619, 620]
				RALYL	1	NA	[619]
				RBM3	1	NA	[40, 618]
				SRSF10	1	[68]	[179]
				TRA2A	1	NA	[619]
				TRA2B	1	[615]	[471]
RBP9	$p < 0.01$	D	FBgn0010263	ELAVL1	9	[615]	[470, 620, 626]
				CSK	1	[615]	[470, 471, 620, 630, 631]
				CSNK1A1	1	[68, 615]	[471]
SF1	$p < 0.05$	C	FBgn0025571	SF1	13	[68, 615]	NA
				KHDRBS1	1	NA	[471]
U2AF50	$p < 0.05$	A	FBgn0005411	U2AF2	15	[615]	[471, 618]
ELAV	$p < 0.05$	D	FBgn0260400	ELAVL1	6	As above	As above
				COPB1	1	NA	[179, 470, 471, 616, 619, 620, 632]
				CSK	1	As above	As above
				CSNK1A1	1	As above	As above
FNE	$p < 0.05$	D	FBgn0086675	ELAVL1	10	As above	As above
				CSK	1	As above	As above
				CSNK1A1	1	As above	As above
ARET	$p < 0.05$	D	FBgn0000114	CELF1	14	NA	[471, 620]
				CELF2	14	NA	[620]
SNRNP70K	$p < 0.05$	A	FBgn0016978	SNRNP70	13	NA	[68, 179, 471, 620]
				HNRNPM	1	NA	[470, 471, 617, 620, 626]
				RBM24	1	[615]	NA
				SRSF1	1	[615]	[470, 471, 617, 618, 620]
				SRSF4	1	NA	[471]
				SRSF6	1	[623]	[471, 619]
				SRSF9	1	NA	[471, 618, 620]
				TIAL1	1	NA	[620]
RBP1	$p < 0.05$	A	FBgn0260944	SRSF3	8	[615, 623]	[68, 179, 471, 617–620]
				SRSF7	6	NA	[179, 471, 618–620]
				SRSF10	1	As above	As above
RBP1-LIKE	$p < 0.05$	A	FBgn0030479	SRSF3	7	As above	As above
				SRSF7	7	NA	As above
				SRSF10	1	As above	As above

Table 7.14: Table continued on the next page.

Table continued from the previous page.

Drosophila RBP	Significance	Motif	Flybase ID	Human Ortholog (Flybase)	Ortholog score (/15)	ExoCarta mRNA EV loaded	EzoCarta Protein EV loaded
TRA2	$p < 0.05$	A	FBgn0003742	TRA2A	12	As above	As above
				TRA2B	11	As above	As above
				CIRBP	1	As above	As above
				EIF4B	1	As above	As above
				EIF4H	1	NA	[40, 470, 618]
				HNRNPA0	1	As above	As above
				HNRNPA1	1	As above	As above
				HNRNPA2B1	1	As above	As above
				HNRNPA3	1	As above	As above
				HNRNPAB	1	As above	As above
				HNRNPC	1	As above	As above
				HNRNPCL2	1	NA	As above
				HNRNPCL3	1	NA	As above
				HNRNPCL4	1	NA	As above
				HNRNPD	1	As above	As above
				HNRNPDL	1	NA	As above
				NCL	1	NA	[68, 471, 620–622, 626, 629]
				RALY	1	NA	As above
				RALYL	1	NA	As above
				RBM3	1	NA	As above
				SLTM	1	NA	[471]
				SRSF10	1	As above	As above

Table 7.14: ExoCarta database literature hits from human or mouse models [173–175] obtained when searching for human orthologs of *Drosophila* RNA binding proteins which have a binding motif significantly similar to the novel *Drosophila* EV-enriched motifs. Orthologs were determined using Flybase (version FB2020_03 [468]). NA = Not applicable.

GO Cluster	GO ID	Term	p-value
1	GO:0048598	embryonic morphogenesis	4.81E-27
1	GO:0009790	embryo development	6.50E-44
1	GO:0001700	embryonic development via the syncytial blastoderm	7.91E-25
1	GO:0009792	embryo development ending in birth or egg hatching	9.53E-24
1	GO:0060429	epithelium development	1.25E-66
1	GO:0009888	tissue development	6.00E-72
1	GO:0000902	cell morphogenesis	1.15E-66
1	GO:0009653	anatomical structure morphogenesis	9.60E-105
1	GO:0048646	anatomical structure formation involved in morphogenesis	5.08E-35
1	GO:0061061	muscle structure development	7.37E-28
1	GO:0032502	developmental process	1.91E-113
1	GO:0048856	anatomical structure development	2.15E-113
1	GO:0048729	tissue morphogenesis	2.73E-60
1	GO:0002009	morphogenesis of an epithelium	5.26E-59
1	GO:0060562	epithelial tube morphogenesis	3.21E-45
2	GO:0007472	wing disc morphogenesis	1.81E-35
2	GO:0007560	imaginal disc morphogenesis	3.99E-45
2	GO:0007476	imaginal disc-derived wing morphogenesis	9.37E-35
2	GO:0035220	wing disc development	1.49E-40
2	GO:0007444	imaginal disc development	6.15E-56
2	GO:0007447	imaginal disc pattern formation	3.18E-21
3	GO:0035282	segmentation	4.69E-21
3	GO:0003002	regionalization	7.52E-43
3	GO:0007389	pattern specification process	8.15E-47
4	GO:0035120	post-embryonic appendage morphogenesis	5.61E-38
4	GO:0009886	post-embryonic animal morphogenesis	4.01E-49
4	GO:0048707	instar larval or pupal morphogenesis	2.44E-47
4	GO:0048736	appendage development	9.37E-42
4	GO:0048737	imaginal disc-derived appendage development	2.11E-41
4	GO:0035107	appendage morphogenesis	4.82E-39
4	GO:0035114	imaginal disc-derived appendage morphogenesis	4.98E-39
4	GO:0009791	post-embryonic development	4.90E-50
4	GO:0002165	instar larval or pupal development	1.85E-46
4	GO:0007552	metamorphosis	2.00E-44
4	GO:0007275	multicellular organism development	7.06E-116
4	GO:0035295	tube development	1.53E-62
4	GO:0007424	open tracheal system development	1.34E-25
4	GO:0060541	respiratory system development	4.92E-26
4	GO:0035239	tube morphogenesis	1.01E-48
4	GO:0007399	nervous system development	6.90E-99
4	GO:0001654	eye development	3.86E-48
4	GO:0048749	compound eye development	2.32E-44
4	GO:0048880	sensory system development	3.86E-48
4	GO:0150063	visual system development	3.86E-48
4	GO:0048569	post-embryonic animal organ development	1.30E-47
4	GO:0007423	sensory organ development	8.06E-50
4	GO:0048513	animal organ development	2.53E-91
4	GO:0035272	exocrine system development	1.61E-20
4	GO:0048731	system development	2.03E-126
4	GO:0007431	salivary gland development	1.61E-20
4	GO:0048732	gland development	7.52E-22
4	GO:0048563	post-embryonic animal organ morphogenesis	3.99E-45
4	GO:0009887	animal organ morphogenesis	1.93E-81
4	GO:0090596	sensory organ morphogenesis	2.32E-41
5	GO:0044703	multi-organism reproductive process	2.12E-28
5	GO:0019953	sexual reproduction	2.12E-28
5	GO:0007276	gamete generation	3.85E-31
5	GO:0007292	female gamete generation	2.31E-33
5	GO:0022414	reproductive process	1.58E-28
5	GO:0003006	developmental process involved in reproduction	7.23E-33
5	GO:0022412	cellular process involved in reproduction in multicellular organism	4.39E-32
5	GO:0048609	multicellular organismal reproductive process	1.53E-29
5	GO:0007281	germ cell development	9.89E-30
5	GO:0048477	oogenesis	2.77E-34
6	GO:0001751	compound eye photoreceptor cell differentiation	7.75E-27
6	GO:0001754	eye photoreceptor cell differentiation	1.38E-28
6	GO:0048592	eye morphogenesis	2.32E-41
6	GO:0001745	compound eye morphogenesis	1.31E-38

Table 7.15: Table continued on the next page.

Table continued from the previous page.

GO Cluster	GO ID	Term	p-value
7	GO:0061564	axon development	6.02E-53
7	GO:0031175	neuron projection development	4.11E-58
7	GO:0008038	neuron recognition	1.53E-21
7	GO:0048666	neuron development	2.49E-67
7	GO:0048667	cell morphogenesis involved in neuron differentiation	2.65E-58
7	GO:0030182	neuron differentiation	6.80E-80
7	GO:0046530	photoreceptor cell differentiation	1.29E-30
7	GO:0022008	neurogenesis	1.01E-83
7	GO:0048699	generation of neurons	1.62E-82
7	GO:0048858	cell projection morphogenesis	1.15E-58
7	GO:0120039	plasma membrane bounded cell projection morphogenesis	1.15E-58
7	GO:0032989	cellular component morphogenesis	1.16E-56
7	GO:0032990	cell part morphogenesis	2.19E-59
7	GO:0000904	cell morphogenesis involved in differentiation	1.54E-60
7	GO:0002064	epithelial cell development	1.84E-21
7	GO:0030855	epithelial cell differentiation	1.90E-24
7	GO:0048468	cell development	1.62E-87
7	GO:0048869	cellular developmental process	1.71E-95
7	GO:0030154	cell differentiation	2.98E-95
7	GO:0045165	cell fate commitment	6.55E-32
8	GO:0040007	growth	4.60E-46
8	GO:0048589	developmental growth	3.80E-40
8	GO:0048638	regulation of developmental growth	2.84E-29
8	GO:0040008	regulation of growth	2.63E-36
8	GO:0051124	synaptic growth at neuromuscular junction	2.97E-27
8	GO:0007416	synapse assembly	2.58E-36
8	GO:0051963	regulation of synapse assembly	5.34E-23
8	GO:0008582	regulation of synaptic growth at neuromuscular junction	2.34E-20
8	GO:0051094	positive regulation of developmental process	8.74E-32
8	GO:0051093	negative regulation of developmental process	5.50E-23
8	GO:0050793	regulation of developmental process	6.00E-71
8	GO:0022603	regulation of anatomical structure morphogenesis	4.16E-35
8	GO:0051960	regulation of nervous system development	1.49E-40
8	GO:2000026	regulation of multicellular organismal development	2.99E-60
9	GO:0007409	axonogenesis	1.49E-53
9	GO:0048812	neuron projection morphogenesis	5.91E-58
9	GO:0097485	neuron projection guidance	1.77E-47
9	GO:0007411	axon guidance	1.04E-47
10	GO:0045595	regulation of cell differentiation	5.94E-43
10	GO:0060284	regulation of cell development	3.57E-35
10	GO:0045664	regulation of neuron differentiation	1.76E-23
10	GO:0050767	regulation of neurogenesis	4.19E-26
11	GO:0006357	regulation of transcription by RNA polymerase II	5.62E-23
11	GO:0006355	regulation of transcription, DNA-templated	1.33E-27
11	GO:0045892	negative regulation of transcription, DNA-templated	3.74E-20
11	GO:1903506	regulation of nucleic acid-templated transcription	1.33E-27
11	GO:2001141	regulation of RNA biosynthetic process	1.33E-27
11	GO:1902679	negative regulation of RNA biosynthetic process	3.74E-20
11	GO:1903507	negative regulation of nucleic acid-templated transcription	3.74E-20
11	GO:0031326	regulation of cellular biosynthetic process	7.39E-36
11	GO:0031327	negative regulation of cellular biosynthetic process	8.97E-25
11	GO:0010558	negative regulation of macromolecule biosynthetic process	1.54E-25
11	GO:0010556	regulation of macromolecule biosynthetic process	5.47E-35
11	GO:2000112	regulation of cellular macromolecule biosynthetic process	2.78E-35
11	GO:2000113	negative regulation of cellular macromolecule biosynthetic process	1.54E-25
11	GO:0051252	regulation of RNA metabolic process	1.21E-33
11	GO:0051253	negative regulation of RNA metabolic process	5.54E-22
11	GO:0019219	regulation of nucleobase-containing compound metabolic process	2.27E-34
11	GO:0045934	negative regulation of nucleobase-containing compound metabolic process	8.63E-23
11	GO:0006351	transcription, DNA-templated	1.57E-21
11	GO:0006366	transcription by RNA polymerase II	3.03E-20
11	GO:0032774	RNA biosynthetic process	1.68E-21
11	GO:0097659	nucleic acid-templated transcription	1.57E-21
12	GO:0006468	protein phosphorylation	2.12E-24
12	GO:0016310	phosphorylation	3.73E-21
12	GO:0044260	cellular macromolecule metabolic process	3.39E-20
13	GO:0031325	positive regulation of cellular metabolic process	6.69E-34
13	GO:0031324	negative regulation of cellular metabolic process	3.14E-28
13	GO:0031323	regulation of cellular metabolic process	6.62E-58
13	GO:0010629	negative regulation of gene expression	7.23E-28
13	GO:0010604	positive regulation of macromolecule metabolic process	1.88E-32
13	GO:0010468	regulation of gene expression	3.30E-40
13	GO:0010605	negative regulation of macromolecule metabolic process	4.67E-29

Table 7.15: Table continued on the next page.

Table continued from the previous page.

GO Cluster	GO ID	Term	p-value
13	GO:0060255	regulation of macromolecule metabolic process	6.33E-54
13	GO:0051173	positive regulation of nitrogen compound metabolic process	1.76E-31
13	GO:0051171	regulation of nitrogen compound metabolic process	8.75E-55
13	GO:0051172	negative regulation of nitrogen compound metabolic process	8.99E-28
13	GO:0080090	regulation of primary metabolic process	5.07E-56
13	GO:0009893	positive regulation of metabolic process	3.13E-35
13	GO:0009892	negative regulation of metabolic process	7.88E-30
13	GO:0019222	regulation of metabolic process	3.38E-57
13	GO:0009889	regulation of biosynthetic process	1.71E-35
13	GO:0009890	negative regulation of biosynthetic process	8.97E-25
14	GO:0051246	regulation of protein metabolic process	8.77E-24
14	GO:0032268	regulation of cellular protein metabolic process	1.81E-24
15	GO:0030030	cell projection organization	2.55E-43
15	GO:0120036	plasma membrane bounded cell projection organization	2.05E-43
15	GO:1901888	regulation of cell junction assembly	2.98E-23
15	GO:0044087	regulation of cellular component biogenesis	7.44E-29
15	GO:0050807	regulation of synapse organization	5.13E-26
15	GO:1904396	regulation of neuromuscular junction development	5.27E-21
15	GO:0051130	positive regulation of cellular component organization	3.19E-23
15	GO:0051128	regulation of cellular component organization	1.10E-49
16	GO:0034329	cell junction assembly	2.63E-38
16	GO:0007528	neuromuscular junction development	7.14E-30
16	GO:0034330	cell junction organization	3.38E-51
16	GO:0050808	synapse organization	4.16E-46
16	GO:0071840	cellular component organization or biogenesis	3.15E-45
16	GO:0097435	supramolecular fiber organization	7.03E-28
16	GO:0016043	cellular component organization	1.60E-54
17	GO:0007010	cytoskeleton organization	1.59E-33
17	GO:0030036	actin cytoskeleton organization	6.12E-38
18	GO:0042221	response to chemical	3.19E-32
18	GO:0010033	response to organic substance	8.71E-21
18	GO:0051716	cellular response to stimulus	4.12E-65
19	GO:0007268	chemical synaptic transmission	6.65E-28
19	GO:0098916	anterograde trans-synaptic signaling	6.65E-28
19	GO:0099537	trans-synaptic signaling	3.33E-28
19	GO:0007267	cell-cell signaling	1.26E-38
19	GO:0099536	synaptic signaling	5.04E-27
20	GO:0007165	signal transduction	2.11E-80
20	GO:0007166	cell surface receptor signaling pathway	5.23E-44
20	GO:0007167	enzyme linked receptor protein signaling pathway	1.80E-22
20	GO:0035556	intracellular signal transduction	1.98E-38
20	GO:0009968	negative regulation of signal transduction	5.73E-27
20	GO:0009966	regulation of signal transduction	3.11E-40
20	GO:1902531	regulation of intracellular signal transduction	1.34E-22
21	GO:0051049	regulation of transport	5.39E-21
22	GO:0051179	localization	5.75E-44
22	GO:0051234	establishment of localization	3.14E-21
22	GO:0006810	transport	7.36E-20
22	GO:0051674	localization of cell	1.00E-26
22	GO:0051641	cellular localization	6.86E-20
23	GO:0007163	establishment or maintenance of cell polarity	6.18E-22
23	GO:0030029	actin filament-based process	2.36E-39
23	GO:0008037	cell recognition	4.72E-21
23	GO:0007154	cell communication	1.76E-99
23	GO:0006928	movement of cell or subcellular component	1.68E-53
23	GO:0009987	cellular process	1.90E-84
24	GO:0048523	negative regulation of cellular process	3.93E-62
24	GO:0010648	negative regulation of cell communication	6.02E-29
24	GO:0010646	regulation of cell communication	7.77E-49
24	GO:0048522	positive regulation of cellular process	1.13E-69
24	GO:0050794	regulation of cellular process	7.43E-140
25	GO:0016477	cell migration	4.81E-35
25	GO:0048870	cell motility	2.39E-27
25	GO:0001667	ameboidal-type cell migration	1.67E-26
25	GO:0010631	epithelial cell migration	1.83E-24
26	GO:0032879	regulation of localization	6.82E-39
27	GO:0050803	regulation of synapse structure or activity	7.75E-27
27	GO:0065008	regulation of biological quality	5.94E-67

Table 7.15: Table continued on the next page.

Table continued from the previous page.

GO Cluster	GO ID	Term	p-value
28	GO:0050890	cognition	1.03E-26
28	GO:0090130	tissue migration	8.84E-25
28	GO:0090132	epithelium migration	8.16E-25
28	GO:0032501	multicellular organismal process	9.76E-84
28	GO:0007610	behavior	3.28E-35
28	GO:0007613	memory	5.83E-21
28	GO:0007611	learning or memory	1.03E-26
29	GO:0050896	response to stimulus	2.89E-61
29	GO:0048585	negative regulation of response to stimulus	1.25E-26
29	GO:0048583	regulation of response to stimulus	3.58E-45
29	GO:0040011	locomotion	1.02E-64
29	GO:0006935	chemotaxis	2.78E-47
29	GO:0042330	taxis	2.15E-51
29	GO:0009605	response to external stimulus	6.37E-35
30	GO:0051241	negative regulation of multicellular organismal process	3.62E-23
30	GO:0051239	regulation of multicellular organismal process	8.63E-72
30	GO:0051240	positive regulation of multicellular organismal process	1.30E-28
30	GO:0023052	signaling	5.22E-97
30	GO:0023057	negative regulation of signaling	1.04E-28
30	GO:0023051	regulation of signaling	7.77E-49
30	GO:0051704	multi-organism process	2.20E-28
30	GO:0065007	biological regulation	6.90E-159
30	GO:0048519	negative regulation of biological process	2.19E-67
30	GO:0048518	positive regulation of biological process	1.20E-70
30	GO:0050789	regulation of biological process	9.23E-143

Table 7.15: Clustered biological process gene ontology terms significantly enriched in genes encoding transcripts containing 3'UTR ABCD motif combinations. 1570 GO terms were enriched. Here, only the 205 most significant GO terms are displayed.

GO Cluster	GO ID	Term	p-value
1	GO:0035770	ribonucleoprotein granule	0.000437312
1	GO:0010494	cytoplasmic stress granule	0.001268343
1	GO:0036464	cytoplasmic ribonucleoprotein granule	0.00068298
2	GO:0036379	myofilament	0.003058483
2	GO:0030016	myofibril	2.48E-09
2	GO:0043292	contractile fiber	1.14E-09
2	GO:0030018	Z disc	7.12E-05
2	GO:0030017	sarcomere	1.14E-08
2	GO:0031674	I band	3.64E-05
3	GO:0005700	polytene chromosome	1.78E-11
3	GO:0005704	polytene chromosome band	0.001394026
3	GO:0005703	polytene chromosome puff	0.008022601
3	GO:0098687	chromosomal region	0.005580194
4	GO:0099513	polymeric cytoskeletal fiber	5.16E-05
4	GO:0005884	actin filament	1.36E-07
4	GO:0031941	filamentous actin	0.000275045
5	GO:0016459	myosin complex	0.001312904
5	GO:0016460	myosin II complex	0.003548888
5	GO:0001725	stress fiber	0.007210599
5	GO:0032432	actin filament bundle	0.000515982
5	GO:0097517	contractile actin filament bundle	0.007210599
6	GO:0005818	aster	0.007210599
6	GO:0005819	spindle	0.002469228
6	GO:0072686	mitotic spindle	0.006518083
6	GO:0000803	sex chromosome	0.005807622
6	GO:0043229	intracellular organelle	1.23E-10
6	GO:0045170	spectrosome	0.006887874
6	GO:0030863	cortical cytoskeleton	0.000519365
6	GO:0099569	presynaptic cytoskeleton	0.002129247
6	GO:0015629	actin cytoskeleton	4.85E-11
6	GO:0042641	actomyosin	0.007210599
6	GO:0005856	cytoskeleton	4.17E-09
6	GO:0043232	intracellular non-membrane-bounded organelle	0.002395599
7	GO:0012506	vesicle membrane	5.07E-09
7	GO:0098588	bounding membrane of organelle	9.41E-05
7	GO:0031982	vesicle	8.13E-16
7	GO:0043227	membrane-bounded organelle	1.26E-07
7	GO:0043226	organelle	1.37E-10
7	GO:0043228	non-membrane-bounded organelle	0.002395599
8	GO:0005654	nucleoplasm	0.000273088
8	GO:0016604	nuclear body	0.000224928
8	GO:0016607	nuclear speck	0.001486402
8	GO:0035062	omega speckle	0.001147788
9	GO:0070732	spindle envelope	0.000371148
9	GO:0005794	Golgi apparatus	0.009667242
9	GO:0031519	PcG protein complex	0.009237648
9	GO:0005634	nucleus	1.87E-13
9	GO:0043231	intracellular membrane-bounded organelle	0.001342603
10	GO:0030140	trans-Golgi network transport vesicle	0.002748014
10	GO:0008021	synaptic vesicle	2.27E-09
10	GO:0030133	transport vesicle	4.57E-12
10	GO:0070382	exocytic vesicle	1.16E-10
10	GO:0005768	endosome	4.33E-05
10	GO:0005769	early endosome	0.006093334
10	GO:0030136	clathrin-coated vesicle	2.14E-06
10	GO:0030135	coated vesicle	0.001598458
10	GO:0031410	cytoplasmic vesicle	6.93E-15
10	GO:0097708	intracellular vesicle	6.93E-15
10	GO:0030139	endocytic vesicle	0.002789453
10	GO:0099503	secretory vesicle	6.37E-10
11	GO:0030285	integral component of synaptic vesicle membrane	0.005807622
11	GO:0030672	synaptic vesicle membrane	8.14E-05
11	GO:0098563	intrinsic component of synaptic vesicle membrane	0.005807622
12	GO:0030120	vesicle coat	0.005890488
12	GO:0030125	clathrin vesicle coat	0.003149244
12	GO:0030658	transport vesicle membrane	1.21E-05
12	GO:0099501	exocytic vesicle membrane	3.17E-05
12	GO:0030659	cytoplasmic vesicle membrane	3.72E-07
12	GO:0030662	coated vesicle membrane	0.002145461
12	GO:0030665	clathrin-coated vesicle membrane	0.001367133

Table 7.16: Table continued on the next page.

Table continued from the previous page.

GO Cluster	GO ID	Term	p-value
13	GO:0009898	cytoplasmic side of plasma membrane	2.49E-08
13	GO:0031234	extrinsic component of cytoplasmic side of plasma membrane	0.000556589
13	GO:0071944	cell periphery	3.83E-59
13	GO:0019897	extrinsic component of plasma membrane	1.15E-06
13	GO:0044298	cell body membrane	0.003548888
13	GO:0005886	plasma membrane	4.16E-54
13	GO:0016328	lateral plasma membrane	0.000581562
14	GO:0045211	postsynaptic membrane	2.33E-06
14	GO:0042734	presynaptic membrane	1.31E-06
14	GO:0097060	synaptic membrane	1.09E-10
15	GO:0016324	apical plasma membrane	4.57E-13
15	GO:0098592	cytoplasmic side of apical plasma membrane	0.001074544
15	GO:0032154	cleavage furrow	2.59E-05
15	GO:0005905	clathrin-coated pit	0.007210599
15	GO:0009925	basal plasma membrane	0.001268343
15	GO:0016327	apicolateral plasma membrane	0.005663969
15	GO:0016323	basolateral plasma membrane	2.22E-06
15	GO:0098590	plasma membrane region	2.51E-16
16	GO:0016021	integral component of membrane	3.21E-05
16	GO:0031224	intrinsic component of membrane	7.91E-06
16	GO:0005887	integral component of plasma membrane	3.67E-18
16	GO:0031226	intrinsic component of plasma membrane	2.50E-19
17	GO:0098552	side of membrane	4.29E-09
17	GO:0098562	cytoplasmic side of membrane	2.42E-08
17	GO:0098805	whole membrane	0.000136629
17	GO:0048475	coated membrane	0.000777475
17	GO:0016020	membrane	3.39E-25
17	GO:0019898	extrinsic component of membrane	3.73E-06
18	GO:0000307	cyclin-dependent protein kinase holoenzyme complex	0.001367133
18	GO:1902554	serine/threonine protein kinase complex	0.000551856
18	GO:1902911	protein kinase complex	0.001659234
19	GO:1990351	transporter complex	1.70E-07
19	GO:0090443	FAR/SIN/STRIPAK complex	0.006887874
19	GO:0005667	transcription regulator complex	0.001820167
19	GO:0043235	receptor complex	3.35E-09
20	GO:0034702	ion channel complex	4.51E-09
20	GO:1902495	transmembrane transporter complex	1.81E-08
20	GO:0034703	cation channel complex	4.26E-06
20	GO:0034705	potassium channel complex	0.001074544
20	GO:0005891	voltage-gated calcium channel complex	0.000371148
20	GO:0005892	acetylcholine-gated channel complex	0.000275045
20	GO:0008076	voltage-gated potassium channel complex	0.002748014
21	GO:0030118	clathrin coat	0.000581562
21	GO:0030117	membrane coat	0.000777475
21	GO:0030119	AP-type membrane coat adaptor complex	0.007086298
22	GO:0098796	membrane protein complex	0.000629073
22	GO:0098797	plasma membrane protein complex	7.61E-16
22	GO:0098802	plasma membrane signaling receptor complex	9.56E-09
23	GO:0030175	filopodium	0.001268343
23	GO:0098858	actin-based cell projection	0.001317168
23	GO:0035230	cytoneme	0.000413231
23	GO:0042995	cell projection	4.41E-15
23	GO:0016028	rhabdomere	0.000152259
23	GO:0120025	plasma membrane bounded cell projection	5.89E-15
24	GO:0061174	type I terminal bouton	4.52E-06
24	GO:0043195	terminal bouton	9.22E-14
24	GO:0043679	axon terminus	1.37E-15
25	GO:0030426	growth cone	2.58E-05
25	GO:0044295	axonal growth cone	1.36E-07
25	GO:0030424	axon	4.91E-19
25	GO:0150034	distal axon	1.31E-19
26	GO:0043197	dendritic spine	0.001147788
26	GO:0044309	neuron spine	0.001147788
26	GO:0043005	neuron projection	4.02E-19
26	GO:0044306	neuron projection terminus	5.64E-16

Table 7.16: Table continued on the next page.

Table continued from the previous page.

GO Cluster	GO ID	Term	p-value
27	GO:0005622	intracellular	4.91E-22
27	GO:0099568	cytoplasmic region	5.83E-05
27	GO:0016528	sarcoplasm	0.007210599
27	GO:0048471	perinuclear region of cytoplasm	6.16E-06
27	GO:0045169	fusome	2.24E-07
27	GO:0005737	cytoplasm	3.06E-11
27	GO:0005829	cytosol	2.72E-12
28	GO:0048788	cytoskeleton of presynaptic active zone	0.002129247
28	GO:0098831	presynaptic active zone cytoplasmic component	0.000769639
28	GO:0005938	cell cortex	2.19E-21
28	GO:0045180	basal cortex	9.74E-05
28	GO:0045179	apical cortex	3.72E-10
28	GO:0099738	cell cortex region	3.31E-15
29	GO:0098595	perivitelline space	0.002952154
29	GO:0045171	intercellular bridge	3.95E-06
29	GO:0035324	female germline ring canal	0.007086298
29	GO:0045172	germline ring canal	8.09E-06
30	GO:0099080	supramolecular complex	1.42E-11
30	GO:0099081	supramolecular polymer	2.04E-10
30	GO:0099512	supramolecular fiber	3.96E-10
31	GO:0044297	cell body	0.001594144
31	GO:0036477	somatodendritic compartment	2.06E-05
31	GO:0043025	neuronal cell body	0.003416975
31	GO:0070938	contractile ring	0.000113501
31	GO:0045178	basal part of cell	2.51E-05
31	GO:0045177	apical part of cell	5.43E-21
31	GO:0032153	cell division site	8.55E-06
31	GO:0030427	site of polarized growth	0.000120514
31	GO:0012505	endomembrane system	0.000434809
31	GO:0009986	cell surface	0.001598458
31	GO:0110165	cellular anatomical entity	1.67E-25
32	GO:0014069	postsynaptic density	0.003149244
32	GO:0099572	postsynaptic specialization	0.009237648
32	GO:0032279	asymmetric synapse	0.003149244
32	GO:0098984	neuron to neuron synapse	0.003149244
32	GO:0048786	presynaptic active zone	7.41E-11
32	GO:0098793	presynapse	4.73E-28
32	GO:0031594	neuromuscular junction	2.28E-10
32	GO:0045202	synapse	2.57E-36
32	GO:0098794	postsynapse	4.51E-09
33	GO:0005918	septate junction	0.000219548
33	GO:0043296	apical junction complex	2.10E-06
33	GO:0070160	tight junction	0.000361593
33	GO:0005911	cell-cell junction	2.58E-16
33	GO:0005912	adherens junction	8.74E-09
34	GO:0030054	cell junction	5.42E-53
34	GO:0070161	anchoring junction	1.20E-20
34	GO:0005927	muscle tendon junction	0.000515982
34	GO:0005925	focal adhesion	0.000265709
34	GO:0030055	cell-substrate junction	5.43E-06

Table 7.16: Clustered cellular component gene ontology terms significantly enriched in genes encoding transcripts containing 3'UTR ABCD motif combinations.

GO Cluster	GO ID	Term	p-value
1	GO:0072509	divalent inorganic cation transmembrane transporter activity	0.000183099
1	GO:0015085	calcium ion transmembrane transporter activity	0.000163972
1	GO:0022890	inorganic cation transmembrane transporter activity	4.88E-06
1	GO:0046873	metal ion transmembrane transporter activity	1.77E-05
1	GO:0015079	potassium ion transmembrane transporter activity	0.000121524
1	GO:0005249	voltage-gated potassium channel activity	0.00108606
1	GO:0005267	potassium channel activity	0.001324928
1	GO:0005245	voltage-gated calcium channel activity	3.11E-05
1	GO:0022843	voltage-gated cation channel activity	4.88E-07
1	GO:0005261	cation channel activity	9.09E-06
1	GO:0005262	calcium channel activity	3.93E-05
2	GO:0005215	transporter activity	0.006320746
2	GO:0015318	inorganic molecular entity transmembrane transporter activity	0.00025908
2	GO:0005326	neurotransmitter transmembrane transporter activity	0.008359133
2	GO:0005342	organic acid transmembrane transporter activity	0.005091913
2	GO:0046943	carboxylic acid transmembrane transporter activity	0.005091913
2	GO:0008509	anion transmembrane transporter activity	0.000116228
2	GO:0008324	cation transmembrane transporter activity	4.72E-06
2	GO:0015075	ion transmembrane transporter activity	0.000345654
3	GO:0005244	voltage-gated ion channel activity	2.91E-05
3	GO:0022832	voltage-gated channel activity	2.91E-05
3	GO:0005231	excitatory extracellular ligand-gated ion channel activity	0.001022089
3	GO:0022848	acetylcholine-gated cation-selective channel activity	0.000222674
4	GO:0061631	ubiquitin conjugating enzyme activity	0.008359133
4	GO:0004842	ubiquitin-protein transferase activity	3.24E-07
4	GO:0019787	ubiquitin-like protein transferase activity	1.22E-06
4	GO:0061630	ubiquitin protein ligase activity	1.38E-05
4	GO:0061659	ubiquitin-like protein ligase activity	2.49E-05
5	GO:0004675	transmembrane receptor protein serine/threonine kinase activity	0.003138763
5	GO:0019199	transmembrane receptor protein kinase activity	0.000767825
5	GO:0004713	protein tyrosine kinase activity	0.00042341
5	GO:0004672	protein kinase activity	7.95E-12
5	GO:0004674	protein serine/threonine kinase activity	1.66E-07
6	GO:0071617	lysophospholipid acyltransferase activity	0.002524847
6	GO:0016773	phosphotransferase activity, alcohol group as acceptor	5.06E-14
6	GO:0016740	transferase activity	0.00037364
6	GO:0016301	kinase activity	6.96E-13
6	GO:0016772	transferase activity, transferring phosphorus-containing groups	7.04E-09
7	GO:0008188	neuropeptide receptor activity	0.003419165
7	GO:0008528	G protein-coupled peptide receptor activity	0.002905516
7	GO:0004930	G protein-coupled receptor activity	1.96E-07
7	GO:0008227	G protein-coupled amine receptor activity	0.00020467
8	GO:0030594	neurotransmitter receptor activity	0.008187001
8	GO:0004952	dopamine neurotransmitter receptor activity	0.003138763
8	GO:0098960	postsynaptic neurotransmitter receptor activity	0.001561668
8	GO:0060089	molecular transducer activity	0.002423171
8	GO:0003707	steroid hormone receptor activity	0.000450979
8	GO:0001653	peptide receptor activity	0.005855327
8	GO:0038023	signaling receptor activity	0.002423171
9	GO:0003712	transcription coregulator activity	0.002825246
9	GO:0140110	transcription regulator activity	5.47E-23
9	GO:0004879	nuclear receptor activity	1.54E-05
9	GO:0098531	ligand-activated transcription factor activity	1.54E-05
9	GO:0001217	DNA-binding transcription repressor activity	1.19E-07
9	GO:0001227	DNA-binding transcription repressor activity, RNA polymerase II-specific	2.55E-07
9	GO:0000981	DNA-binding transcription factor activity, RNA polymerase II-specific	3.52E-21
9	GO:0003700	DNA-binding transcription factor activity	4.98E-24
9	GO:0001216	DNA-binding transcription activator activity	3.86E-09
9	GO:0001228	DNA-binding transcription activator activity, RNA polymerase II-specific	1.51E-08
10	GO:0008047	enzyme activator activity	6.22E-05
10	GO:0030234	enzyme regulator activity	0.002012163
10	GO:0005096	GTPase activator activity	0.001754019
10	GO:0030695	GTPase regulator activity	0.000378495
10	GO:0060589	nucleoside-triphosphatase regulator activity	0.000447256
11	GO:0019207	kinase regulator activity	0.001122251
11	GO:0016538	cyclin-dependent protein serine/threonine kinase regulator activity	0.004022002
11	GO:0019887	protein kinase regulator activity	0.008480513
11	GO:0019209	kinase activator activity	0.00711949
11	GO:0030295	protein kinase activator activity	0.004495475
11	GO:0043539	protein serine/threonine kinase activator activity	0.005515885

Table 7.17: Table continued on the next page.

Table continued from the previous page.

GO Cluster	GO ID	Term	p-value
12	GO:0004114	3',5'-cyclic-nucleotide phosphodiesterase activity	0.005058434
12	GO:0034596	phosphatidylinositol phosphate 4-phosphatase activity	0.003138763
12	GO:0008138	protein tyrosine/serine/threonine phosphatase activity	0.005515885
12	GO:0004721	phosphoprotein phosphatase activity	5.41E-05
12	GO:0004725	protein tyrosine phosphatase activity	0.000130521
13	GO:0005246	calcium channel regulator activity	0.0009905
13	GO:0099106	ion channel regulator activity	0.000122027
13	GO:0016247	channel regulator activity	5.48E-06
13	GO:0098772	molecular function regulator	7.61E-29
14	GO:0008307	structural constituent of muscle	0.005515885
14	GO:0030371	translation repressor activity	2.92E-05
14	GO:0045182	translation regulator activity	0.006859733
14	GO:0003924	GTPase activity	5.57E-05
14	GO:0008239	dipeptidyl-peptidase activity	0.005906052
15	GO:0001786	phosphatidylserine binding	0.000878512
15	GO:0070300	phosphatidic acid binding	0.003138763
15	GO:0035091	phosphatidylinositol binding	0.003516696
15	GO:0005543	phospholipid binding	6.88E-09
15	GO:0005544	calcium-dependent phospholipid binding	0.000122027
16	GO:0005509	calcium ion binding	4.27E-12
16	GO:0043169	cation binding	0.008883036
16	GO:0008201	heparin binding	0.004022002
16	GO:0043167	ion binding	8.84E-10
16	GO:0043168	anion binding	3.68E-10
17	GO:0051371	muscle alpha-actinin binding	0.003138763
17	GO:0017022	myosin binding	5.73E-06
17	GO:0032036	myosin heavy chain binding	0.001810023
17	GO:0008017	microtubule binding	3.88E-06
17	GO:0015631	tubulin binding	0.000102838
17	GO:0019894	kinesin binding	0.006097191
17	GO:0003779	actin binding	9.04E-14
17	GO:0008092	cytoskeletal protein binding	1.64E-17
18	GO:0008134	transcription factor binding	7.64E-11
18	GO:0001221	transcription cofactor binding	0.006097191
18	GO:0001223	transcription coactivator binding	0.005058434
18	GO:0070491	repressing transcription factor binding	0.00020467
18	GO:0061629	RNA polymerase II-specific DNA-binding transcription factor binding	0.00711949
18	GO:0140297	DNA-binding transcription factor binding	8.11E-07
19	GO:0019899	enzyme binding	3.15E-11
19	GO:0019900	kinase binding	1.04E-09
19	GO:0019901	protein kinase binding	4.69E-09
19	GO:0005085	guanyl-nucleotide exchange factor activity	0.002708587
19	GO:0031267	small GTPase binding	2.13E-05
19	GO:0051020	GTPase binding	1.24E-05
20	GO:0031208	POZ domain binding	0.001810023
20	GO:0017124	SH3 domain binding	0.000638497
20	GO:0019904	protein domain specific binding	1.25E-08
20	GO:0046983	protein dimerization activity	9.30E-05
20	GO:0042802	identical protein binding	6.57E-07
20	GO:0042803	protein homodimerization activity	6.23E-07
20	GO:0001784	phosphotyrosine residue binding	0.005058434
20	GO:0046332	SMAD binding	0.000638497
20	GO:0070411	I-SMAD binding	0.0009905
20	GO:0005102	signaling receptor binding	6.95E-09
20	GO:0005112	Notch binding	0.002151131
20	GO:0050839	cell adhesion molecule binding	3.89E-11
20	GO:0030276	clathrin binding	1.65E-05
20	GO:0019955	cytokine binding	0.002524847
20	GO:0000149	SNARE binding	0.004995941
20	GO:0005515	protein binding	1.57E-62
21	GO:0060090	molecular adaptor activity	2.85E-07
21	GO:0030674	protein-macromolecule adaptor activity	7.30E-07
21	GO:0035591	signaling adaptor activity	0.00047019
22	GO:0044877	protein-containing complex binding	0.001829371
22	GO:0051015	actin filament binding	2.42E-05
22	GO:0097367	carbohydrate derivative binding	0.000152027
22	GO:0042562	hormone binding	0.002483239
22	GO:0036094	small molecule binding	7.36E-05
22	GO:0005488	binding	3.93E-40
22	GO:0008289	lipid binding	1.48E-07
23	GO:0001882	nucleoside binding	8.49E-05
23	GO:0001883	purine nucleoside binding	3.01E-05
23	GO:0032549	ribonucleoside binding	3.01E-05
23	GO:0032550	purine ribonucleoside binding	1.31E-05

Table 7.17: Table continued on the next page.

Table continued from the previous page.

GO Cluster	GO ID	Term	p-value
24	GO:0030554	adenyl nucleotide binding	2.78E-05
24	GO:0019001	guanyl nucleotide binding	1.90E-05
24	GO:0000166	nucleotide binding	1.28E-06
24	GO:0017076	purine nucleotide binding	1.59E-08
24	GO:0005525	GTP binding	1.31E-05
24	GO:0019003	GDP binding	0.000878512
24	GO:0005524	ATP binding	3.93E-05
24	GO:0032561	guanyl ribonucleotide binding	1.97E-05
24	GO:0032553	ribonucleotide binding	7.35E-08
24	GO:0032555	purine ribonucleotide binding	3.33E-08
24	GO:0032559	adenyl ribonucleotide binding	5.43E-05
25	GO:0000976	transcription regulatory region sequence-specific DNA binding	2.53E-19
25	GO:0000977	RNA polymerase II transcription regulatory region sequence-specific DNA binding	7.41E-19
25	GO:0000978	RNA polymerase II cis-regulatory region sequence-specific DNA binding	2.06E-14
25	GO:0000987	cis-regulatory region sequence-specific DNA binding	1.77E-14
25	GO:0003677	DNA binding	7.43E-12
25	GO:0003690	double-stranded DNA binding	1.22E-17
25	GO:0043565	sequence-specific DNA binding	1.16E-20
25	GO:1990837	sequence-specific double-stranded DNA binding	1.79E-19
26	GO:0003729	mRNA binding	3.19E-11
26	GO:0003730	mRNA 3'-UTR binding	2.03E-08
26	GO:0008187	poly-pyrimidine tract binding	0.005515885
26	GO:0008266	poly(U) RNA binding	7.69E-05
26	GO:0003727	single-stranded RNA binding	0.00474141
26	GO:0070717	poly-purine tract binding	0.000878512
27	GO:0000900	translation repressor activity, mRNA regulatory element binding	0.000329734
27	GO:0008494	translation activator activity	0.0009905
27	GO:0035639	purine ribonucleoside triphosphate binding	1.77E-08
27	GO:1901265	nucleoside phosphate binding	1.28E-06
27	GO:1901363	heterocyclic compound binding	2.67E-14
27	GO:0097159	organic cyclic compound binding	2.68E-14
27	GO:0001067	regulatory region nucleic acid binding	3.51E-19
27	GO:0003676	nucleic acid binding	1.28E-09

Table 7.17: Clustered molecular function gene ontology terms significantly enriched in genes encoding transcripts containing 3'UTR ABCD motif combinations.

		MEME-4 (A)	MEME-2 (B)	MEME-7 (C)	MEME-3 (D)	MEME-1	MEME-15	MEME-21
EV only plus EV enriched transcripts	5'UTR	49.34	17.69	40.44	18.40	6.98	8.80	0.00
	CDS	69.69	55.43	0.30	0.10	0.20	0.00	0.00
	3'UTR	47.70	18.54	50.90	39.08	35.27	0.00	9.12
Cell only transcripts	5'UTR	15.16	3.31	9.35	6.40	0.96	0.00	0.00
	CDS	17.00	0.00	0.00	0.14	0.00	0.00	0.00
	3'UTR	6.18	2.35	6.18	6.11	5.12	0.00	0.00
EV only plus EV enriched genes	5'UTR	43.61	17.70	35.08	21.64	8.52	10.16	0.00
	CDS	60.95	44.13	0.32	0.32	0.32	0.00	0.00
	3'UTR	37.58	11.15	40.13	30.89	28.03	0.00	6.05
Cell only genes	5'UTR	13.33	2.99	9.25	7.76	1.09	0.00	0.00
	CDS	14.45	0.00	0.00	0.26	0.00	0.00	0.00
	3'UTR	4.77	1.68	4.90	5.67	4.77	0.00	0.00

Table 7.18: The prevalence of 5'UTR and CDS *Drosophila* motif occurrences determined within the 'EV only plus EV enriched' and 'Cell only' active mRNA transcripts. Values shown as a percentage of the total number of genes or transcripts investigated within that group (shown in **Supplementary Table 7.19**).

			Number of sequences available
5'UTR	**Cell only**	*Transcripts*	1359
		Genes	735
	EV only plus EV enriched	*Transcripts*	989
		Genes	305
CDS	**Cell only**	*Transcripts*	1406
		Genes	775
	EV only plus EV enriched	*Transcripts*	1003
		Genes	315
3'UTR	**Cell only**	*Transcripts*	1407
		Genes	776
	EV only plus EV enriched	*Transcripts*	998
		Genes	314

Table 7.19: Number of genes and transcripts for which 5'UTR, CDS and 3'UTR sequences were available from Ensemble Biomart and used for FIMO motif analysis.

7.4 Supplementary Listings

```
#!/usr/bin/env python
# coding: utf-8

# In[186]:


import pandas as pd
from glob import glob
import matplotlib as mpl
import matplotlib.pyplot as plt
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
from sklearn import decomposition
from sklearn import datasets
from sklearn.decomposition import PCA


path = "/Users/clairehill/Documents/Oxford/PhD_project/RNA-seq_with_Lizzie/AugustwithJana/2_DESEQ2/"


#Read in csv
df = pd.read_csv (path+"Normalised_coutns.csv", sep=",", index_col=0)
df


#Transpose table so different samples make up different rows
df1_transposed = df.T
df1_transposed

#Extract row names from transposed tabled and make this into a new dataframe - this will be used later
indexNamesArr = df1_transposed.index.values
names = pd.DataFrame(data = indexNamesArr)
names.rename(columns={0: "Sample"}, inplace=True)
names

#Scaling data to normalise
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
data_rescaled = scaler.fit_transform(df1_transposed)



#Determining the 95% of variance - how many components do you need to include to account for 95% of variance
pca = PCA(n_components = 0.95)
pca.fit(data_rescaled)
reduced2 = pca.transform(data_rescaled)
X=reduced2
X

#Plot the analysis of how many components and % variance
pca = PCA().fit(data_rescaled)
get_ipython().run_line_magic("matplotlib", "inline")

plt.rcParams["figure.figsize"] = (12,6)

fig, ax = plt.subplots()
xi = np.arange(1, 7, step=1)
y = np.cumsum(pca.explained_variance_ratio_)

plt.ylim(0.0,1.1)
plt.plot(xi, y, marker="o", linestyle="--", color="b")

plt.xlabel("Number of Components", fontsize=15)
plt.xticks(np.arange(0, 7, step=1)) #change from 0-based array index to 1-based human-readable label
plt.ylabel("Cumulative variance (%)", fontsize=15)
plt.title("The number of components needed to explain variance", fontsize=15)

ax.xaxis.labelpad = 15
ax.yaxis.labelpad = 15

plt.axhline(y=0.95, color="r", linestyle="-")
plt.text(0.5, 0.85, "95% cut-off threshold", color = "red", fontsize=16)

ax.grid(axis="x")
plt.savefig(path+"normalized_count_PCA_number_of_components", dpi=300, bbox_inches = "tight")

plt.show()

#PCA analysis - 2 component
pca = PCA(n_components = 2)
pca.fit(data_rescaled)
reduced = pca.transform(data_rescaled)

principalDf = pd.DataFrame(data = reduced
             , columns = ["principal component 1", "principal component 2"])

#Adding the PCA analysis data to the names extracted earlier
finalDf = pd.concat([principalDf, names[["Sample"]]], axis = 1)

```

```
#Renaming names as you wish them to appear on the graph
finalDf.iloc[0, finalDf.columns.get_loc("Sample")] = "Cell Rep 1"
finalDf.iloc[1, finalDf.columns.get_loc("Sample")] = "Cell Rep 2"
finalDf.iloc[2, finalDf.columns.get_loc("Sample")] = "Cell Rep 3"
finalDf.iloc[3, finalDf.columns.get_loc("Sample")] = "EV Rep 1"
finalDf.iloc[4, finalDf.columns.get_loc("Sample")] = "EV Rep 2"
finalDf.iloc[5, finalDf.columns.get_loc("Sample")] = "EV Rep 3"


#Visualise 2 component PCA

fig = plt.figure(figsize = (8,8))
ax = fig.add_subplot(1,1,1)
ax.set_xlabel("Principal component 1", fontsize = 15)
ax.set_ylabel("Principal component 2", fontsize = 15)
targets = ["Cell Rep 1", "Cell Rep 2", "Cell Rep 3", "EV Rep 1", "EV Rep 2", "EV Rep 3"]
colors = ["g", "g", "g", "blue", "blue", "blue"]
markers = ["s","P","o","s","P","o"]
for target, color, marker in zip(targets,colors, markers):
    indicesToKeep = finalDf["Sample"] == target
    ax.scatter(finalDf.loc[indicesToKeep, "principal component 1"]
               , finalDf.loc[indicesToKeep, "principal component 2"]
               , c = color
               , s = 300
               , marker = marker
               , alpha = 0.6)

ax.xaxis.labelpad = 20
ax.yaxis.labelpad = 20

ax.legend(targets, fontsize=20)
plt.savefig(path+"normalized_count_PCA", dpi=300, bbox_inches = "tight")
ax.grid()



#PCA analysis - 3 component
pca = PCA(n_components = 3)
pca.fit(data_rescaled)
reduced3 = pca.transform(data_rescaled)

principalDfnew = pd.DataFrame(data = reduced3
             , columns = ["principal component 1", "principal component 2", "principal component 3"])


finalDfnew = pd.concat([principalDfnew, names[["Sample"]]], axis = 1)

finalDfnew.iloc[0, finalDfnew.columns.get_loc("Sample")] = "Cell Rep 1"
finalDfnew.iloc[1, finalDfnew.columns.get_loc("Sample")] = "Cell Rep 2"
finalDfnew.iloc[2, finalDfnew.columns.get_loc("Sample")] = "Cell Rep 3"
finalDfnew.iloc[3, finalDfnew.columns.get_loc("Sample")] = "EV Rep 1"
finalDfnew.iloc[4, finalDfnew.columns.get_loc("Sample")] = "EV Rep 2"
finalDfnew.iloc[5, finalDfnew.columns.get_loc("Sample")] = "EV Rep 3"

#3D plot of 3 component analysis
fig = plt.figure(figsize = (15,10))
ax = fig.add_subplot(1, 1, 1, projection="3d")

targets = ["Cell Rep 1", "Cell Rep 2", "Cell Rep 3", "EV Rep 1", "EV Rep 2", "EV Rep 3"]
colors = ["g", "g", "g", "blue", "blue", "blue"]
markers = ["s","P","o","s","P","o"]
for target, color, marker in zip(targets,colors, markers):
    indicesToKeep = finalDfnew["Sample"] == target
# Data for three-dimensional scattered points
    zdata = finalDfnew.loc[indicesToKeep, "principal component 3"]
    xdata = finalDfnew.loc[indicesToKeep, "principal component 1"]
    ydata = finalDfnew.loc[indicesToKeep, "principal component 2"]
    ax.scatter3D(xdata, ydata, zdata, c=color, s=200, cmap="Greens", marker=marker, alpha=0.5)

ax.xaxis.labelpad = 15
ax.yaxis.labelpad = 15
ax.zaxis.labelpad = 15

ax.legend(targets, fontsize=20, loc="center left", bbox_to_anchor=(1, 0.6))

ax.set_xlabel("Principal component 1", fontsize = 15)
ax.set_ylabel("Principal component 2", fontsize = 15)
ax.set_zlabel("Principal component 3", fontsize = 15)

#use if you want to change the direction which you look at 3D graph
#ax.view_init(3, 3)


plt.savefig(path+"normalized_count_PCA_3D", dpi=300, bbox_inches = "tight")
```

Listing 7.1: Principal component analysis and visualisation for normalised read counts.

```
import pandas as pd

# creating dataframe with assigned colours and motif names
path="/Users/Claire/Documents/Oxford/PhD_project/RNA-seq_with_Lizzie/AugustwithJana/7_RNAplfold/RNAfold/24_dec_19/28_dec_19/"

d=[["red","A","MEME-4"],
   ["green","B","MEME-2"],
   ["blue","C","MEME-7"],
   ["orange","D","MEME-3"]]


info_df = pd.DataFrame(d,columns=("color","motif_name","motif_alt_id"))
info_df

# merging motifs table with newly created dataframe on motif id to have the correct color and correct motif name next to each coordinate
merged_df=pd.merge(info_df,dedup_motif_df,on="motif_alt_id")
merged_df["motif_length"]=merged_df["stop"]-merged_df["start"]
merged_df
#merged_df.to_csv("./Desktop/motif_probability_in_loop_BEFORE_FILTRATION.csv",index=False,header=True,sep=",")

# Removing stacked identical motifs and specific regions
# because some motifs create a stacked  features. They need to be removed and only one value will be kept
# by grouping the motifs name and transcripts and subtracting the "start" values by the previous value and because these values are sorted they will create only small values if they are located next to each other

merged_df["diff"] = merged_df.sort_values(["start"]).groupby(["motif_name","Trans_ID"])["start"].diff()
# because first number will never have any value it was assigned NAN but in order to keep this number I have assigned the value 50 to all NAN values

merged_df["diff"]=merged_df["diff"].fillna(50)
# filtering out those lines with value less the 10

filter_for_probability=merged_df[(merged_df["diff"] > 10)]
#filter_for_probability.to_csv(path+ "motif_probability_TEST.csv",index=False,header=True,sep=",")
filter_for_probability
```

Listing 7.2: Removing overlapping motif hits from the FIMO output data.[44]

[44]Code developed by Jana Terenova

```
import matplotlib.pyplot as plt
import pandas as pd
from matplotlib.lines import Line2D
from matplotlib.patches import Patch

path="/Users/Claire/Documents/Oxford/PhD_project/RNA-seq_with_Lizzie/AugustwithJana/6_FIMO/__EV/
     positive_hits_only/"


#import FIMO data
df = pd.read_csv(path+"positive_trans_ABCD_position_SHORTLIST.csv", low_memory=False, header=None)

#import list of single transcripts with ABCD motifs
df_transcripts = pd.read_csv(path+"positive_trans_ABCD_SHORTLIST.csv")
#change DF into list
##print(df_transcripts.head(10))
Transcripts_list = df_transcripts["Trans_ID"].values.tolist()

df_transcriptID_geneID = pd.read_csv(path+"positive_trans_ABCD_label_SHORTLIST.csv")
print(df_transcriptID_geneID.head(3))
TranscriptID_geneID_list = df_transcriptID_geneID["Trans_ID"].values.tolist()
print(TranscriptID_geneID_list)
##print(df.head(3))

new_header = df.iloc[0] #grab the first row for the header
df = df[1:] #take the data less the header row
df.columns = new_header #set the header row as the df header

df[["Begin", "End"]] = df[["Begin", "End"]].apply(pd.to_numeric)
types = df.dtypes
##print(types)

new_header_transcripts = df_transcripts.iloc[0] #grab the first row for the header
df_transcripts = df_transcripts[1:] #take the data less the header row
df_transcripts.columns = new_header_transcripts #set the header row as the df header

### Preview the first 5 lines of the loaded data
#print(df.head(30))
#print(df_transcripts.head(30))


#Fill a dictionary with y axis heights for each sequence name
#Used to separate the sequences along the y axis
seq_name_y = {}
i = 0
i_list = []
for t in Transcripts_list:
  seq_name_y[t] = i
  i_list.append(i)
  i += 1


#Make different colours for each motif
colours = {}
colours["MEME-3"] = "orange"
colours["MEME-4"] = "red"
colours["MEME-7"] = "blue"
colours["MEME-2"] = "green"


# iterate over rows with iterrows()

#dropna gets rid of missing data, nan. SOme beginning and end values were missing
df = df.dropna()

fig, ax = plt.subplots(figsize=(20, 25))

#Loop over all rows in df, and plot it with a specific colour and y axis height
for index, row in df.iterrows():

    #using error bar plots to make it easier to plot each row as you go along
    #The box width we want is (row["End"]-row["Begin"])
    #The error bar plot gives you a plus and a minus error bar
    #So to make the error bars cover the whole box width, we set each error bar to be half of the box width
    error_bar = float(row["End"]-row["Begin"])/2
    #We set the error bar point to be the middle of the box, so that our error bars will cover the box width
     we want
    middle = 0.5*(row["Begin"]+row["End"])
    #This is the unique y height for each sequence name, which we defined in seq_name_y above. It goes from 0
     to 23
    y = seq_name_y["%s" % row["Trans_ID"]]
    #This is the colour of the bar, based on what the motif number is. It uses the colours dictionary we made
     above.
    colour = colours["%s" % row["motif_alt_id"]]
    #Linewdith=4 plots the bar as a fat line! Which makes it look like a box :)
    #markersize=0 means it doesnt plot the middle error bar point, which we dont need
    #capsize=0 means it doesn"t put endcaps onto the error bars, which we dont need either
    plt.errorbar(middle, y, xerr=error_bar,fmt="o",markersize=0,ecolor=colour,linewidth=4, capsize=0)

    #Set the y axis names
    ax.set_yticks(i_list)
    ax.set_yticklabels(TranscriptID_geneID_list, fontsize=10)
    #Set the axis levels
```

```
90      ax.set(xlabel="Normalised 3UTR postion", ylabel="Identifier")
91      ax.xaxis.label.set_size(15)
92      ax.yaxis.label.set_size(15)
93
94  # Create the figure legend
95
96  legend_elements = [Line2D([0], [0], color="red", lw=4, label="MEME-4 (A)"), Line2D([0], [0], color="green", lw
        =4, label="MEME-2 (B)"), Line2D([0], [0], color="blue", lw=4, label="MEME-7 (C)"), Line2D([0], [0],
        color="orange", lw=4, label="MEME-3 (D)")]
97  ax.legend(handles=legend_elements, loc="upper right", scatterpoints=1, ncol=12, fontsize=15)
98
99  plt.tight_layout()
100 fig.savefig("figure_longlist_labels.pdf")
101 #plt.show()
```

Listing 7.3: Plotting motifs along the normalised 3'UTR position example.

```
import glob as glob
filter_for_probability # deduplicated and unstacked dataframe with values

# subsetting the dataset for suther analysis
selected_new=filter_for_probability[["motif_name","start","Trans_ID"]]

# grouping dataframe by transcript name
gp = selected_new.groupby('Trans_ID')

# creating new text files for each transcript which will include motif and start position and transcript name
gp.apply(lambda x: x.to_csv('/Users/clairehill/Documents/Oxford/PhD_project/RNA-seq_with_Lizzie/AugustwithJana
    /9_motif_order/April_2020/output/' + str(x.name) + '.txt',index=False,header=False))

import itertools
from pprint import pprint
import csv

# import all the files from directory where they were created with the ending of txt

motifs_files= glob.glob('/Users/clairehill/Documents/Oxford/PhD_project/RNA-seq_with_Lizzie/AugustwithJana/9
    _motif_order/April_2020/output/*.txt')

# create a function that would return all the combinations of motifs order per transcript in position order
class Motif:

    def __init__(self, symbol,position,transcript):  # using three arguments from the input file:
        self.symbol = symbol                          # 1, motif symbol
        self.position = position                      # 2, motif start position
        self.transcript = transcript                  # 3, transcript name

    def __repr__(self):
        return ("Motif[\"%s\",%d,\"%s\"]" % (self.symbol,self.position,self.transcript))
final_ls=[]
for file in motifs_files:                          # looping through the file names
    file_name=file.rpartition("/")[2]              # selecting file name and using the file name opening files
    data=[]
    csv=pd.read_csv("/Users/clairehill/Documents/Oxford/PhD_project/RNA-seq_with_Lizzie/AugustwithJana/9
     _motif_order/April_2020/output/{}".format(file_name),header=None) # opening files


    for id, line in csv.iterrows():               # iterating over each dataframe and selecting data from
     three columns
        file_data=(line[0],line[1],line[2])
        data.append(file_data)



    dic_motif = dict();                           # create empty dictionary



    for item in data:

        dic_motif.setdefault(item[0],[]).append(Motif(item[0],item[1],item[2]))      # Apply the Motif
     function


    dic_as_list = list(dic_motif.values())                # create list from values in dictionary

    result = list(itertools.product(*dic_as_list))        # create motifs unique combination


    for i in result:
        orderedList = sorted(i, key=lambda x: x.position) # order motifs based on their position
        final=([x.symbol for x in orderedList],[x.transcript for x in orderedList]) # output symbol
     combinations per transcript and transcript name
        final_ls.append(final)
          #append data to the list

df_final=pd.DataFrame(final_ls)                # create dataframe from list with motifs and transcript name
print(df_final)
df_final.to_csv("/Users/clairehill/Documents/Oxford/PhD_project/RNA-seq_with_Lizzie/AugustwithJana/9
    _motif_order/April_2020/motif_script_output_April_final_1.txt", sep='\t',index=False,header=False)

# import motifs combination file
motif_file=pd.read_csv('/Users/clairehill/Documents/Oxford/PhD_project/RNA-seq_with_Lizzie/AugustwithJana/9
    _motif_order/April_2020/motif_script_output_April_final_1.txt',sep="\t",header=None)

ls_motif=[]
for s, line in motif_file.iterrows():
    motif_combination=line[0].replace("'","").replace("[","").replace("]","") # removing quotes and square
     brackets
    transcript_ID=line[1].replace("'","").replace("[","").replace("]","").split(",")[1] # removing quotes and
     square brackets and selecting only one column of transcript name
    result=(motif_combination,transcript_ID)
    ls_motif.append(result)                       # appending data to list


df_motif1=pd.DataFrame(ls_motif,columns=("motifs_combinations","transcript_name"))    # create a new dataframe

df_motif1.groupby(["motifs_combinations"])["transcript_name"].nunique() # motifs located in unique transcripts
df_motif1.groupby(["motifs_combinations"])["transcript_name"].size()    # motifs counts occurances in all
    transcripts
```

```
83
84 no_unique_trans = df_motif1.groupby(["motifs_combinations"])["transcript_name"].nunique() # motifs located in
       unique transcripts
85 no_all_trans = df_motif1.groupby(["motifs_combinations"])["transcript_name"].size()    # motifs counts
       occurances in all transcripts
86
87 no_unique_trans
```

Listing 7.4: Determining motif order.[45]

[45] Code developed by Jana Terenova

```
import pandas as pd
import numpy as np
from io import BytesIO

final_list=[]
path="/Users/clairehill/Documents/Oxford/PhD_project/RNA-seq_with_Lizzie/AugustwithJana/7_RNAplfold/RNAfold/24_dec_19/"
# importing RNAFold csv files
rnafold_inputs= glob.glob(path+"*.csv")
#print(rnafold_inputs)

colNames = ("transcript_name", "motif_alt_id", "motif_length", "start", "stop", "count", "% bases paired")
masterDF = pd.DataFrame(columns = colNames)

for rnafoldfile in rnafold_inputs: # looping through the file names
    rnafold_file_name=rnafoldfile[0:-4].rpartition("/")[-1] # extracting  transcript names from the path
    print(rnafold_file_name)
    df_list=[]
    with open (path+"{}.csv".format(rnafold_file_name),"r") as input_file:
        # using the format function importing only the above seleted fold files
        df= pd.read_csv(input_file,header=0)     # opening rnafold files

                df.dropna(inplace=True   # removing NA if any present
        df["name"]= rnafold_file_name
        ########################################
        # DOT is unpaired. Bracked it paired
        # 0 is unpaired. 1 it paired
        ########################################
        df.replace(".", 0, inplace=True)
        df.replace(")", 1, inplace=True)
        df.replace("(", 1, inplace=True)
        length = df["Coordinate"].iloc[-1]
        length_int = int(length)
        df_list.append(df) # appending everything into one list
    array1 = [None] * length_int            # creating empty array of the length of transcript
    iterating_df = pd.concat(df_list,axis=0)   # creating new dataframe
    iterating_df[["gene_name", "transcript_name"]] = iterating_df["name"].str.split("_",expand=True) # splitting name column content into two new columns
##################################################
    for id, line in iterating_df.iterrows():

        array1[line[1]]=line[2]        # storing transcript nucleotide position and probabily values in the arrays per each transcript
    df_combined = pd.DataFrame(columns=["count"])
    probabilities=[]
    # iterating over the file with motif names, positions, IDs, transcript names and motif lengths
    for line in zip(filter_for_probability["motif_name"],filter_for_probability["motif_alt_id"],
     filter_for_probability["start"],filter_for_probability["stop"],filter_for_probability["Trans_ID"],
     filter_for_probability["motif_length"]):

        for item in zip(iterating_df ["Coordinate"],iterating_df ["Paired/Unpaired"],iterating_df ["transcript_name"]):   # iterating over the file with transcript nucleotide positions and probability values and transcript name

            if line[4] == item[2] and item[0] in range (line[2],line[3]):                   # if transcript name from motif file equals to the transcript name in probability file and if the trancript nucleotide position is in the range of motif start and end positions
                output=(item[0],item[1], item[2], line[1],line[2],line[3],line[4],line[5],line[0])    # then save transcript position, position probability value, transcript ID, motif length, motif start position, motif end position, motif ID, motif name
                probabilities.append(output)


    df_probabilities=pd.DataFrame(probabilities,columns=("Coordinate","Paired_Unpaired","transcript_name","motif_alt_id","start","stop","Trans_ID","motif_length","motif_name"))  # creating new dataframe with the selected values
    df_paired_count=df_probabilities.groupby(["transcript_name","motif_alt_id", "motif_length", "start", "stop"])["Paired_Unpaired"].apply(lambda x: (x==1).sum()).reset_index(name="count")
    df_paired_count["% bases paired"] = (df_paired_count["count"]/df_paired_count["motif_length"])*100
    masterDF = masterDF.append(df_paired_count,ignore_index=True)
##################################################
masterDF
```

Listing 7.5: Determine percentage paired.[46]

[46]Code developed in collaboration with Jana Terenova

References

[1] Bernard Perbal. “Communication is the key”. In: *Cell Communication and Signaling* 3.1 (2003).

[2] Björn L.D.M. Brücher and Ijaz S. Jamall. “Cell-cell communication in the tumor microenvironment, carcinogenesis, and anticancer treatment”. In: *Cellular Physiology and Biochemistry* 34.2 (2014), pp. 213–243.

[3] H. Spemann and B. Geinitz. “Über Weckung organisatorischer Fähigkeiten durch Verpflanzung in organisatorische Umgebung”. In: *Wilhelm Roux’ Archiv für Entwicklungsmechanik der Organismen* 109.2 (Apr. 1927), pp. 129–175.

[4] Roy Sougata and Thomas B. Kornberg. “Paracrine signaling mediated at cell-cell contacts”. In: *Bioessays* 37.1 (2015), pp. 25–33.

[5] L. Wolpert. “Positional information and the spatial pattern of cellular differentiation”. In: *Journal of Theoretical Biology* 25.1 (1969), pp. 1–47.

[6] Christiane Nüsslein-volhard and Eric Wieschaus. “Mutations affecting segment number and polarity in Drosophila”. In: *Nature* 287.5785 (1980), pp. 795–801.

[7] Frans Rijsewijk et al. “The Drosophila homology of the mouse mammary oncogene int-1 is identical to the segment polarity gene wingless”. In: *Cell* 50.4 (1987), pp. 649–657.

[8] Philip W. Ingham, Yoshiro Nakano, and Claudia Seger. “Mechanisms and functions of Hedgehog signalling across the metazoa”. In: *Nature Reviews Genetics* 12.6 (2011), pp. 393–406.

[9] Ken M. Cadigan and Roel Nusse. “Wnt signaling: A common theme in animal development”. In: *Genes and Development* 11.24 (1997), pp. 3286–3305.

[10] Hye Sook Han et al. “Regulation of glucose metabolism from a liver-centric perspective”. In: *Experimental and Molecular Medicine* 48.3 (2016).

[11] P Graham and L. Pick. “Drosophila as a model for diabetes and diseases of insulin resistance”. In: *Current Topics in Developmental Biology* 121 (2017), pp. 397–419.

[12] K. G. Guruharsha, Mark W. Kankel, and Spyros Artavanis-Tsakonas. “The Notch signalling system: Recent insights into the complexity of a conserved pathway”. In: *Nature Reviews Genetics* 13.9 (2012), pp. 654–666.

[13] Mark E. Fortini. “Notch Signaling: The core pathway and its posttranslational regulation”. In: *Developmental Cell* 16.5 (2009), pp. 633–647.

[14] Benjamin Mattes and Steffen Scholpp. “Emerging role of contact-mediated cell communication in tissue development and diseases”. In: *Histochemistry and Cell Biology* 150.5 (2018), pp. 431–442.

[15] Michael Cohen et al. “Dynamic filopodia transmit intermittent Delta-Notch signaling to drive pattern refinement during lateral inhibition”. In: *Developmental Cell* 19.1 (2010), pp. 78–89.

[16] Cyrille De Joussineau et al. “Delta-promoted filopodia mediate long-range lateral inhibition in Drosophila”. In: *Nature* 426.6966 (2003), pp. 555–559.

[17] Laura González-Méndez, Ana Citlali Gradilla, and Isabel Guerrero. “The cytoneme connection: Direct long-distance signal transfer during development”. In: *Development* 146.9 (2019).

[18] Yukiko M. Yamashita, Mayu Inaba, and Michael Buszczak. "Specialized Intercellular Communications via Cytonemes and Nanotubes". In: *Annual Review of Cell and Developmental Biology* 34 (2018), pp. 59–84.

[19] J. Miller, S. E. Fraser, and D. McClay. "Dynamics of thin filopodia during sea urchin gastrulation". In: *Development* 121.8 (1995), pp. 2501–2511.

[20] T. B. Kornberg and S. Roy. "Cytonemes as specialized signaling filopodia". In: *Development* 141.4 (2014), pp. 729–736.

[21] Felipe Andrés Ramírez-Weber and Thomas B. Kornberg. "Cytonemes: Cellular processes that project to the principal signaling center in Drosophila imaginal discs". In: *Cell* 97.5 (1999), pp. 599–607.

[22] Mayu Inaba, Michael Buszczak, and Yukiko M. Yamashita. "Nanotubes mediate niche-stem-cell signalling in the Drosophila testis". In: *Nature* 523.7560 (2015), pp. 329–332.

[23] Amin Rustom et al. "Nanotubular Highways for Intercellular Organelle Transport". In: *Science* 303.5660 (2004), pp. 1007–1010.

[24] Hans Hermann Gerdes and Raquel Negrão Carvalho. "Intercellular transfer mediated by tunneling nanotubes". In: *Current Opinion in Cell Biology* 20.4 (2008), pp. 470–475.

[25] Sergio Casas-tintó and Marta Portela. "Cytonemes, Their Formation, Regulation, and Roles in Signaling and Communication in Tumorigenesis". In: *International Journal of Molecular Sciences* 20.22 (2019).

[26] Magnus Wiger Austefjord, Hans Hermann Gerdes, and Xiang Wang. "Tunneling nanotubes: Diversity in morphology and structure". In: *Communicative and Integrative Biology* 7.1 (2014).

[27] Jordi Cayuso et al. "EphrinB1/EphB3b Coordinate Bidirectional Epithelial-Mesenchymal Interactions Controlling Liver Morphogenesis and Laterality". In: *Developmental Cell* 39.3 (2016), pp. 316–328.

[28] Marcus Bischoff et al. "Cytonemes are required for the establishment of a normal Hedgehog morphogen gradient in Drosophila epithelia". In: *Nature Cell Biology* 15.11 (2013), pp. 1269–1281.

[29] Laura González-Méndez, Irene Seijo-Barandiarán, and Isabel Guerrero. "Cytoneme-mediated cell-cell contacts for Hedgehog reception". In: *eLife* 6 (2017).

[30] Weitao Chen et al. "Essential basal cytonemes take up hedgehog in the Drosophila wing imaginal disc". In: *Development* 144.17 (2017), pp. 3134–3144.

[31] Timothy A. Sanders, Esther Llagostera, and Maria Barna. "Specialized filopodia direct long-range transport of SHH during vertebrate tissue patterning". In: *Nature* 497.7451 (2013), pp. 628–632.

[32] Hai Huang and Thomas B. Kornberg. "Myoblast cytonemes mediate Wg signaling from the wing imaginal disc and Delta-Notch signaling to the air sac primordium". In: *eLife* 4 (2015).

[33] Laura González-méndez et al. "Polarized sorting of Patched enables cytoneme-mediated Hedgehog reception in the Drosophila wing disc". In: *EMBO Journal* 39 (2020).

[34] Ceren Korkut et al. “Trans-Synaptic Transmission of Vesicular Wnt Signals through Evi/Wntless”. In: *Cell* 139.2 (2009), pp. 393–404.

[35] Sougata Roy et al. “Cytoneme-mediated contact-dependent transport of the Drosophila decapentaplegic signaling protein”. In: *Science* 343.6173 (2014).

[36] Ilse Hurbain et al. “Microvilli-derived Extracellular Vesicles Govern Morphogenesis in Drosophila wing epithelium”. In: *bioRxiv* (2020).

[37] Michael Danilchik, Melissa Williams, and Elizabeth Brown. “Blastocoel-spanning filopodia in cleavage-stage Xenopus laevis: Potential roles in morphogen distribution and detection”. In: *Developmental Biology* 382.1 (2013), pp. 70–81.

[38] Wolf Heusermann et al. “Exosomes surf on filopodia to enter cells at endocytic hot spots, traffic within endosomes, and are targeted to the ER”. In: *Journal of Cell Biology* 213.2 (2016), pp. 173–184.

[39] F T Borges, L A Reis, and N Schor. “Extracellular vesicles: structure, function, and potential clinical uses in renal diseases”. In: *Brazilian Journal of Medical and Biological Research* 46.10 (2013).

[40] T. Pisitkun, R.-F. Shen, and M. A. Knepper. “Identification and proteomic profiling of exosomes in human urine”. In: *Proceedings of the National Academy of Sciences* 101.36 (2004), pp. 13368–13373.

[41] Charlotte Admyre et al. “Exosomes with immune modulatory features are present in human breast milk”. In: *Journal of Immunology* 179.3 (2007).

[42] H F Heijnen et al. “Activated platelets release two types of membrane vesicles: microvesicles by surface shedding and exosomes derived from exocytosis of multivesicular bodies and alpha-granules.” In: *Blood* 94.11 (1999), pp. 3791–9.

[43] M.-P. Caby et al. “Exosomal-like vesicles are present in human blood plasma”. In: *International Immunology* 17.7 (2005), pp. 879–887.

[44] Mercedes Tkach and Clotilde Thery. “Communication by Extracellular Vesicles: Where we are and where we need to go”. In: *Cell* 164.6 (2016), pp. 1226–1232.

[45] Mea M. Holm, Julia Kaiser, and Martin E. Schwab. “Extracellular Vesicles: Multimodal Envoys in Neural Maintenance and Repair”. In: *Trends in Neurosciences* 41.6 (2018), pp. 360–372.

[46] Karen Beckett et al. “Drosophila S2 cells secrete wingless on exosome-like vesicles but the wingless gradient forms independently of exosomes”. In: *Traffic* 14.1 (2013), pp. 82–96.

[47] Julia Christina Gross et al. “Active Wnt proteins are secreted on exosomes”. In: *Nature Cell Biology* 14.10 (2012), pp. 1036–1045.

[48] Paul D. Robbins and Adrian E. Morelli. “Regulation of immune responses by extracellular vesicles”. In: *Nature Reviews Immunology* 14.3 (2014), pp. 195–208.

[49] Valentina R Minciacchi et al. “Extracellular Vesicles in Cancer: Exosomes, Microvesicles and the Emerging Role of Large Oncosomes”. In: *Seminars in Cell and Developmental Biology* 40 (2015), pp. 41–51.

[50] Rong Xu et al. “Extracellular vesicles in cancer — implications for future improvements in cancer care”. In: *Nature Reviews Clinical Oncology* 15.10 (2018), pp. 617–638.

[51] Edit I. Buzas et al. “Emerging role of extracellular vesicles in inflammatory diseases”. In: *Nature Reviews Rheumatology* 10.6 (2014), pp. 356–364.

[52] E Chargaff and R West. “The biological significance of the thromboplastic protein of blood”. eng. In: *The Journal of Biological Chemistry* 166.1 (1946), pp. 189–197.

[53] P. Wolf. “The nature and significance of platelet products in human plasma”. In: *British Journal of Haematology* 13.3 (1967), pp. 269–288.

[54] H. C. Anderson. “Vesicles associated with calcification in the matrix of epiphyseal cartilage”. In: *The Journal of Cell Biology* 41.1 (1969), pp. 59–72.

[55] Irving M. Shapiro, William J. Landis, and Makarand V. Risbud. “Matrix vesicles: Are they anchored exosomes?” In: *Bone* 79 (2015), pp. 29–36.

[56] Eberhard G. Trams et al. “Exfoliation of membrane ecto-enzymes in the form of micro-vesicles”. In: *Biochimica et Biophysica Acta - Biomembranes* 645.1 (1981), pp. 63–70.

[57] Bin Tao Pan and Rose M. Johnstone. “Fate of the transferrin receptor during maturation of sheep reticulocytes in vitro: Selective externalization of the receptor”. In: *Cell* 33.3 (1983), pp. 967–978.

[58] C Harding, J Heuser, and P Stahl. “Receptor-mediated endocytosis of transferrin and recycling of the transferrin receptor in rat reticulocytes”. In: *The Journal of Cell Biology* 97.2 (1983), pp. 329–39.

[59] C Harding, J Heuser, and P Stahl. “Endocytosis and intracellular processing of transferrin and colloidal gold-transferrin in rat reticulocytes: demonstration of a pathway for receptor shedding.” eng. In: *European Journal of Cell Biology* 35.2 (1984), pp. 256–263.

[60] B T Pan et al. “Electron microscopic evidence for externalization of the transferrin receptor in vesicular form in sheep reticulocytes”. In: *The Journal of Cell Biology* 101.3 (1985), pp. 942–8.

[61] G Raposo et al. “B lymphocytes secrete antigen-presenting vesicles”. In: *The Journal of Experimental Medicine* 183.3 (Mar. 1996), pp. 1161–72.

[62] Francesc Baixauli, Carlos López-Otín, and Maria Mittelbrunn. “Exosomes and autophagy: Coordinated mechanisms for the maintenance of cellular fitness”. In: *Frontiers in Immunology* 5 (2014).

[63] Marina Colombo et al. “Analysis of ESCRT functions in exosome biogenesis, composition and secretion highlights the heterogeneity of extracellular vesicles”. In: *Journal of Cell Science* 126.24 (2013), pp. 5553–5565.

[64] Charles E. Jackson et al. “Effects of inhibiting VPS4 support a general role for ESCRTs in extracellular vesicle biogenesis”. In: *Biophysical Journal* 113.6 (2017), pp. 1342–1352.

[65] Keiichi Tamai et al. “Exosome secretion of dendritic cells is regulated by Hrs, an ESCRT-0 protein”. In: *Biochemical and Biophysical Research Communications* 399.3 (2010), pp. 384–390.

[66] Maria Francesca Baietti et al. “Syndecan-syntenin-ALIX regulates the biogenesis of exosomes”. In: *Nature Cell Biology* 14.7 (2012), pp. 677–685.

[67] Charles Géminard et al. “Degradation of AP2 during reticulocyte maturation enhances binding of hsc70 and Alix to a common site on TfR for sorting in exosomes”. In: *Traffic* 5.3 (2004), pp. 181–193.

[68] Hadi Valadi et al. “Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells”. In: *Lipid Insights* 9.6 (2007), pp. 654–659.

[69] J. Ratajczak et al. “Embryonic stem cell-derived microvesicles reprogram hematopoietic progenitors: Evidence for horizontal transfer of mRNA and protein delivery”. In: *Leukemia* 20.5 (2006), pp. 847–856.

[70] Johan Skog et al. “Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers”. In: *Nature Cell Biology* 10.12 (2008), pp. 1470–1476.

[71] Ni-ting Chiou et al. “Selective export into extracellular vesicles and function of tRNA fragments during T cell activation”. In: *Cell Reports* 25.12 (2018), 3356–3370.e4.

[72] Maoshan Chen et al. “Transcriptome and long noncoding RNA sequencing of three extracellular vesicle subtypes released from the human colon cancer LIM1863 cell line”. In: *Scientific Reports* 6 (2016), pp. 1–14.

[73] Joanna Kowal et al. “Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes”. In: *Proceedings of the National Academy of Sciences of the United States of America* 113.8 (2016), E968–E977.

[74] Morayma M. Temoche-Diaz et al. “Distinct mechanisms of microRNA sorting into cancer cell-derived extracellular vesicle subtypes”. In: *eLife* 8 (2019).

[75] Haiying Zhang et al. “Identification of distinct nanoparticles and subsets of extracellular vesicles by asymmetric flow field-flow fractionation”. In: *Nature Cell Biology* 20.3 (2018), pp. 332–343.

[76] Dennis K. Jeppesen et al. “Reassessment of Exosome Composition”. In: *Cell* 177.2 (2019), 428–445.e18.

[77] Elisa Lázaro-Ibáñez et al. “Distinct prostate cancer-related mRNA cargo in extracellular vesicle subsets from prostate cell lines”. In: *BMC Cancer* 17 (2017).

[78] Juan Pablo Tosar et al. “Assessment of small RNA sorting into different extracellular fractions revealed by high-throughput sequencing of breast cell lines”. In: *Nucleic Acids Research* 43.11 (2015), pp. 5601–5616.

[79] Eduard Willms et al. “Cells release subpopulations of exosomes with distinct molecular and biological properties”. In: *Scientific Reports* 6 (2016), pp. 1–12.

[80] Guillaume Van Niel et al. “Exosomes: A common pathway for a specialized function”. In: *Journal of Biochemistry* 140.1 (2006), pp. 13–21.

[81] Jean Gruenberg and Harald Stenmark. “The biogenesis of multivesicular endosomes”. In: *Nature Reviews Molecular Cell Biology* 5.4 (2004), pp. 317–323.

[82] Eduard Willms et al. “Extracellular vesicle heterogeneity: Subpopulations, isolation techniques, and diverse functions in cancer progression”. In: *Frontiers in Immunology* 9 (2018), p. 738.

[83] Sarah Caruso and Ivan K.H. Poon. "Apoptotic cell-derived extracellular vesicles: More than just debris". In: *Frontiers in Immunology* 9 (2018), p. 1486.

[84] J F Kerr, A H Wyllie, and A R Currie. "Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics". In: *British Journal of Cancer* 26.4 (1972), pp. 239–57.

[85] Georgia K. Atkin-Smith et al. "A novel mechanism of generating extracellular vesicles during apoptosis via a beads-on-a-string membrane structure". In: *Nature Communications* 6 (2015).

[86] E. E. Morrison, M. A. Bailey, and J. W. Dear. "Renal extracellular vesicles: from physiology to clinical application". In: *Journal of Physiology* 594.20 (2016), pp. 5735–5748.

[87] Hurley James H. "ESCRTs are everywhere". In: *EMBO Journal* 34.19 (2015), pp. 2398–2407.

[88] Clotilde Théry et al. "Proteomic Analysis of Dendritic Cell-Derived Exosomes: A Secreted Subcellular Compartment Distinct from Apoptotic Vesicles". In: *The Journal of Immunology* 166.12 (2001), pp. 7309–7318.

[89] William M. Henne, Nicholas J. Buchkovich, and Scott D. Emr. "The ESCRT Pathway". In: *Developmental Cell* 21.1 (2011), pp. 77–91.

[90] Oliver Schmidt and David Teis. "The ESCRT machinery". In: *Current Biology* 22.4 (2012), R116–R120.

[91] Camilla Raiborg et al. "Hrs recruits clathrin to early endosomes". In: *EMBO Journal* 20.17 (2001), pp. 5008–5021.

[92] Susan C. Shih et al. "Epsins and Vps27p/Hrs contain ubiquitin-binding domains that function in receptor endocytosis". In: *Nature Cell Biology* 4.5 (2002), pp. 389–393.

[93] Patricia S. Bilodeau et al. "The Vps27p-Hse1p complex binds ubiquitin and mediates endosomal protein sorting". In: *Nature Cell Biology* 4.7 (2002), pp. 534–539.

[94] Kristi G. Bache et al. "Hrs regulates multivesicular body formation via ESCRT recruitment to endosomes". In: *Journal of Cell Biology* 162.3 (2003), pp. 435–442.

[95] David J. Katzmann, Markus Babst, and Scott D. Emr. "Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I". In: *Cell* 106.2 (2001), pp. 145–155.

[96] James H. Hurley and Phyllis I. Hanson. "Membrane budding and scission by the ESCRT machinery: it's all in the neck". In: *Nature Reviews Molecular Cell Biology* 11.8 (2010), pp. 556–566.

[97] Markus Babst et al. "ESCRT-III: An endosome-associated heterooligomeric protein complex required for MVB sorting". In: *Developmental Cell* 3.2 (2002), pp. 271–282.

[98] Markus Babst et al. "Endosome-associated complex, ESCRT-II, recruits transport machinery for protein sorting at the multivesicular body". In: *Developmental Cell* 3.2 (2002), pp. 283–289.

[99] Suraj Saksena et al. “Functional Reconstitution of ESCRT-III Assembly and Disassembly”. In: *Cell* 136.1 (2009), pp. 97–109.

[100] Markus Babst et al. “Endosomal transport function in yeast requires a novel AAA-type ATPase, Vps4p”. In: *EMBO Journal* 16.8 (1997), pp. 1820–1831.

[101] Hsiangling Teo et al. “ESCRT-II, an endosome-associated complex required for protein sorting: Crystal structure and interactions with ESCRT-III and membranes”. In: *Developmental Cell* 7.4 (2004), pp. 559–569.

[102] David Teis, Suraj Saksena, and Scott D. Emr. “Ordered Assembly of the ESCRT-III Complex on Endosomes Is Required to Sequester Cargo during MVB Formation”. In: *Developmental Cell* 15.4 (2008), pp. 578–589.

[103] Rania Ghossoub et al. “Syntenin-ALIX exosome biogenesis and budding into multivesicular bodies are controlled by ARF6 and PLD2”. In: *Nature Communications* 5 (2014).

[104] Bart Roucourt et al. “Heparanase activates the syndecan-syntenin-ALIX exosome pathway”. In: *Cell Research* 25.4 (2015), pp. 412–428.

[105] Susanne Stuffers et al. “Multivesicular endosome biogenesis in the absence of ESCRTs”. In: *Traffic* 10.7 (2009), pp. 925–937.

[106] Zoraida Andreu and María Yáñez-Mó. “Tetraspanins in extracellular vesicle formation and function”. In: *Frontiers in Immunology* 5 (2014).

[107] Fedor Berditchevski and Elena Odintsova. “Tetraspanins as regulators of protein trafficking”. In: *Traffic* 8.2 (2007), pp. 89–96.

[108] Arthit Chairoungdua et al. “Exosome release of β-catenin: A novel mechanism that antagonizes Wnt signaling”. In: *Journal of Cell Biology* 190.6 (2010), pp. 1079–1091.

[109] Sonja I. Buschow et al. “MHC II In dendritic cells is targeted to lysosomes or t cell-induced exosomes via distinct multivesicular body pathways”. In: *Traffic* 10.10 (2009), pp. 1528–1542.

[110] Guillaume van Niel et al. “The Tetraspanin CD63 Regulates ESCRT-Independent and -Dependent Endosomal Sorting during Melanogenesis”. In: *Developmental Cell* 21.4 (2011), pp. 708–721.

[111] Jack Taylor and Mary Bebawy. “Proteins Regulating Microvesicle Biogenesis and Multidrug Resistance in Cancer”. In: *Proteomics* 19.1-2 (2019).

[112] Ian John McGough and Jean-Paul Vincent. “Exosomes in developmental signalling”. In: *Development* 143.14 (2016), pp. 2482–2493.

[113] Anshul S. Jadli et al. “Inside(sight) of tiny communicator: exosome biogenesis, secretion, and uptake”. In: *Molecular and Cellular Biochemistry* 467.1-2 (2020), pp. 77–94.

[114] Maarten P. Bebelman et al. “Biogenesis and function of extracellular vesicles in cancer”. In: *Pharmacology and Therapeutics* 188 (2018), pp. 1–11.

[115] Nina Pettersen Hessvik and Alicia Llorente. “Current knowledge on exosome biogenesis and release”. In: *Cellular and Molecular Life Sciences* 75.2 (2018), pp. 193–208.

[116] Hirotami Matsuo et al. “Role of LBPA and Alix in Multivesicular Liposome Formation and Endosome Organization”. In: *Science* 303.5657 (2004), pp. 531–534.

[117] Erich Gulbins and Richard Kolesnick. “Raft ceramide in molecular medicine”. In: *Oncogene* 22 (2003), pp. 7070–7077.

[118] Katarina Trajkovic et al. “Ceramide triggers budding of exosome vesicles into multivesicular endosomes”. In: *Science* 319.5867 (2008).

[119] Huaqing Cai, Karin Reinisch, and Susan Ferro-Novick. “Coats, Tethers, Rabs, and SNAREs Work Together to Mediate the Intracellular Destination of a Transport Vesicle”. In: *Developmental Cell* 12.5 (2007), pp. 671–682.

[120] Benjamin Short et al. “The Rab6 GTPase regulates recruitment of the dynactin complex to Golgi membranes”. In: *Current Biology* 12.20 (2002), pp. 1792–1795.

[121] Theodoros Matanis et al. “Bicaudal-D regulates COPI-independent Golgi-ER transport by recruiting the dynein-dynactin motor complex”. In: *Nature Cell Biology* 4.12 (2002), pp. 986–992.

[122] John A. Hammer and Xufeng S. Wu. “Rabs grab motors: Defining the connections between Rab GTPases and motor proteins”. In: *Current Opinion in Cell Biology* 14.1 (2002), pp. 69–75.

[123] F. P. Finger and P. Novick. “Synthetic interactions of the post-Golgi sec mutations of Saccharomyces cerevisiae”. In: *Genetics* 156.3 (2000), pp. 943–951.

[124] R. A. Bacon et al. “The GTP-binding protein Ypt1 is required for transport in vitro: The Golgi apparatus is defective in ypt1 mutants”. In: *Journal of Cell Biology* 109.3 (1989), pp. 1015–1022.

[125] Kate Koles et al. “Mechanism of evenness interrupted (Evi)-exosome release at synaptic boutons”. In: *Journal of Biological Chemistry* 287.20 (2012), pp. 16820–16834.

[126] Ariel Savina, Michel Vidal, and Maria I. Colombo. “The exosome pathway in K562 cells is regulated by Rab11”. In: *Journal of Cell Science* 115.12 (2002), pp. 2505–2515.

[127] Matias Ostrowski et al. “Rab27a and Rab27b control different steps of the exosome secretion pathway”. In: *Nature Cell Biology* 12.1 (2010), pp. 19–30.

[128] Véronique Proux-Gillardeaux et al. “Expression of the Longin domain of TI-VAMP impairs lysosomal secretion and epithelial cell migration”. In: *Biology of the Cell* 99.5 (2007), pp. 261–271.

[129] Reinhard Jahn and Richard H. Scheller. “SNAREs - Engines for membrane fusion”. In: *Nature Reviews Molecular Cell Biology* 7.9 (2006), pp. 631–643.

[130] Bénédicte Hugel et al. “Membrane microparticles: Two sides of the coin”. In: *Physiology* 20.1 (2005), pp. 22–27.

[131] Bretscher Mark, S. “Asymmetrical Lipid Bilayer Structure for Biological Membranes”. In: *Nature New Biology* 236 (1972), pp. 11–12.

[132] Edouard M. Bevers et al. “Lipid translocation across the plasma membrane of mammalian cells”. In: *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* 1439.3 (1999), pp. 317–330.

[133] Edouard M. Bevers, Paul Comfurius, and Robert F.A. Zwaal. "Changes in membrane phospholipid distribution during platelet activation". In: *Biochimica et Biophysica Acta - Biomembranes* 736.1 (1983), pp. 57–66.

[134] David L. Daleke. "Regulation of transbilayer plasma membrane phospholipid asymmetry". In: *Journal of Lipid Research* 44.2 (2003), pp. 233–242.

[135] François Bassé et al. "Isolation of an erythrocyte membrane protein that mediates Ca2+- dependent transbilayer movement of phospholipid". In: *Journal of Biological Chemistry* 271.29 (1996), pp. 17205–17210.

[136] Crislyn D'Souza-Schorey and Philippe Chavrier. "ARF proteins: Roles in membrane traffic and beyond". In: *Nature Reviews Molecular Cell Biology* 7.5 (2006), pp. 347–358.

[137] Vandhana Muralidharan-Chari et al. "ARF6-Regulated Shedding of Tumor Cell-Derived Plasma Membrane Microvesicles". In: *Current Biology* 19.22 (2009), pp. 1875–1885.

[138] B. Li et al. "RhoA triggers a specific signaling pathway that generates transforming microvesicles in cancer cells". In: *Oncogene* 31.45 (2012), pp. 4740–4749.

[139] Katharina B. Beer and Ann Marie Wehman. "Mechanisms and functions of extracellular vesicle release in vivo—What we can learn from flies and worms". In: *Cell Adhesion and Migration* 11.2 (2017), pp. 135–150.

[140] Ann M. Wehman et al. "The P4-ATPase TAT-5 inhibits the budding of extracellular vesicles in C. elegans embryos". In: *Current Biology* 21.23 (2011), pp. 1951–1959.

[141] Joseph F. Nabhan et al. "Formation and release of arrestin domain-containing protein 1-mediated microvesicles (ARMMs) at plasma membrane by recruitment of TSG101 protein". In: *Proceedings of the National Academy of Sciences of the United States of America* 109.11 (2012), pp. 4146–4151.

[142] L. Michael Thomas and Russell D. Salter. " Activation of Macrophages by P2X 7 -Induced Microvesicles from Myeloid Cells Is Mediated by Phospholipids and Is Partially Dependent on TLR4 ". In: *The Journal of Immunology* 185.6 (2010), pp. 3740–3749.

[143] Fabio Bianco et al. "Acid sphingomyelinase activity triggers microparticle release from glial cells". In: *EMBO Journal* 28.8 (2009), pp. 1043–1054.

[144] Kerstin Menck et al. "Neutral sphingomyelinases control extracellular vesicles budding from the plasma membrane". In: *Journal of Extracellular Vesicles* 6.1 (2017).

[145] Carl D. Bortner, Maria I. Sifre, and John A. Cidlowski. "Cationic gradient reversal and cytoskeleton-independent volume regulatory pathways define an early stage of apoptosis". In: *Journal of Biological Chemistry* 283.11 (2008), pp. 7219–7229.

[146] Olivier Morel et al. "Cellular mechanisms underlying the formation of circulating microparticles". In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 31.1 (2011), pp. 15–26.

[147] Sarah J. Storr et al. "The calpain system and cancer". In: *Nature Reviews Cancer* 11.5 (2011), pp. 364–374.

[148] J. Li and J. Yuan. "Caspases in apoptosis and beyond". In: *Oncogene* 27.48 (2008), pp. 6194–6206.

[149] Lorenzo Galluzzi et al. "Molecular mechanisms of cell death: Recommendations of the Nomenclature Committee on Cell Death 2018". In: *Cell Death and Differentiation* 25.3 (2018), pp. 486–541.

[150] Isabelle Sirois et al. "Caspase activation regulates the extracellular export of autophagic vacuoles". In: *Autophagy* 8.6 (2012).

[151] I. Sirois et al. "Caspase-3-dependent export of TCTP: A novel pathway for antiapoptotic intercellular communication". In: *Cell Death and Differentiation* 18.3 (2011), pp. 549–562.

[152] Soo Jeong Park et al. "Molecular mechanisms of biogenesis of apoptotic exosome-like vesicles and their roles as damage-associated molecular patterns". In: *Proceedings of the National Academy of Sciences of the United States of America* 115.50 (2018), E11721–E11730.

[153] Michaël Sebbagh et al. "Caspase-3-mediated cleavage of ROCK I induces MLC phosphorylation and apoptotic membrane blebbing". In: *Nature Cell Biology* 3 (2001), pp. 346–352.

[154] Michael Sebbagh et al. "Direct cleavage of ROCK II by granzyme B induces target cell membrane blebbing in a caspase-independent manner". In: *Journal of Experimental Medicine* 201.3 (2005), pp. 465–471.

[155] Jon D. Lane, Victoria J. Allan, and Philip G. Woodman. "Active relocation of chromatin and endoplasmic reticulum into blebs in late apoptotic cells". In: *Journal of Cell Science* 118.17 (2005), pp. 4059–4071.

[156] Xuebo Xu, Yueyang Lai, and Zi Chun Hua. "Apoptosis and apoptotic body: Disease message and therapeutic target potentials". In: *Bioscience Reports* 39.1 (2019), pp. 1–17.

[157] Cédric Sapet et al. "Thrombin-induced endothelial microparticle generation: Identification of a novel pathway involving ROCK-II activation by caspase-2". In: *Blood* 108.6 (2006), pp. 1868–1876.

[158] Sang Bum Kim et al. "Caspase-8 controls the secretion of inflammatory lysyl-tRNA synthetase in exosomes from cancer cells". In: *The Journal of Cell Biology* 216.7 (2017), pp. 1–22.

[159] Michael Ibba and S Dieter. "Aminoacyl-tRNA Synthesis". In: *Annual Review of Biochemistry* 69 (2000), pp. 617–650.

[160] Min Guo and Paul Schimmel. "Essential nontranslational functions of tRNA synthetases". In: *Nature chemical biology* 9.3 (2013), pp. 145–53.

[161] S. G. Park et al. "From The Cover: Human lysyl-tRNA synthetase is secreted to trigger proinflammatory response". In: *Proceedings of the National Academy of Sciences of the United States of America* 102.18 (2005), pp. 6356–6361.

[162] Véronique Friand, Guido David, and Pascale Zimmermann. "Syntenin and syndecan in the biogenesis of exosomes". In: *Biology of the Cell* 107.10 (2015), pp. 331–341.

[163] Catherine Rabouille. "KRS: A cut away from release in exosomes". In: *The Journal of Cell Biology* 216.7 (2017), pp. 1891–1893.

[164] Megan Stevens and Sebastian Oltean. "Modulation of the Apoptosis Gene Bcl-x Function Through Alternative Splicing". In: *Frontiers in Genetics* 10 (2019).

[165] Ioulia Vardaki et al. "Caspase-3-dependent cleavage of Bcl-xL in the stroma exosomes is required for their uptake by hematological malignant cells". In: *Blood* 128.23 (2016), pp. 2655–2665.

[166] Luis Alberto Baena-Lopez et al. "Non-apoptotic Caspase regulation of stem cell properties". In: *Seminars in Cell and Developmental Biology* (2017).

[167] María Yáñez-Mó et al. "Biological properties of extracellular vesicles and their physiological functions". In: *Journal of Extracellular Vesicles* 4 (2015).

[168] Erik R. Abels and Xandra O. Breakefield. "Introduction to Extracellular Vesicles: Biogenesis, RNA Cargo Selection, Content, Release, and Uptake". In: *Cellular and Molecular Neurobiology* 36.3 (2016), pp. 301–312.

[169] Tatyana Vagner et al. "Protein Composition Reflects Extracellular Vesicle Heterogeneity". In: *Proteomics* 19.8 (2019).

[170] Mercedes Tkach et al. "Qualitative differences in T-cell activation by dendritic cell-derived extracellular vesicle subtypes". In: *EMBO Journal* 36.20 (2017), pp. 3012–3028.

[171] R. M. Johnstone et al. "Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes)." In: *Journal of Biological Chemistry* 262.19 (1987), pp. 9412–9420.

[172] D.-K. Kim et al. "EVpedia: a community web portal for extracellular vesicles research". In: *Bioinformatics* 31.6 (Mar. 2015), pp. 933–939.

[173] Suresh Mathivanan et al. "ExoCarta 2012: Database of exosomal proteins, RNA and lipids". In: *Nucleic Acids Research* 40.D1 (2012), pp. 1241–1244.

[174] Suresh Mathivanan and Richard J. Simpson. "ExoCarta: A compendium of exosomal proteins and RNA". In: *Proteomics* 9.21 (2009), pp. 4997–5000.

[175] Shivakumar Keerthikumar et al. "ExoCarta: A Web-Based Compendium of Exosomal Cargo". In: *Journal of Molecular Biology* 428.4 (2016), pp. 688–692.

[176] Hina Kalra et al. "Vesiclepedia: A Compendium for Extracellular Vesicles with Continuous Community Annotation". In: *PLoS Biology* 10.12 (2012), pp. 8–12.

[177] Andrew M. Leidal et al. "The LC3-conjugation machinery specifies the loading of RNA-binding proteins into extracellular vesicles". In: *Nature Cell Biology* 22.2 (2020), pp. 187–199.

[178] Jamie M. Rimer et al. "Long-range function of secreted small nucleolar RNAs thatrecieve direct 2-O-methylation". In: *Journal of Biological Chemistry* 293.34 (2018), pp. 13284–13296.

[179] Michelle Demory Beckler et al. "Proteomic analysis of exosomes from mutant KRAS colon cancer cells identifies intercellular transfer of mutant KRAS". In: *Molecular and Cellular Proteomics* 12.2 (2013), pp. 343–355.

[180] Lilla Turiák et al. "Proteomic characterization of thymocyte-derived microvesicles and apoptotic bodies in BALB/c mice". In: *Journal of Proteomics* 74.10 (2011), pp. 2025–2033.

[181] Mireya Gonzalez-Begne et al. "Proteomic Analysis of Human Parotid Gland Exosomes by Multidimensional Protein Identification Technology (MudPIT) Mireya". In: *Journal of Proteome Research* 8.3 (2009), pp. 1304–1314.

[182] Javier Conde-Vancells et al. "Characterization and comprehensive proteome profiling of exosomes secreted by hepatocytes". In: *Journal of Proteome Research* 7.12 (2008), pp. 5157–5166.

[183] Michael W. Graner et al. "Proteomic and immunologic analyses of brain tumor exosomes". In: *The FASEB Journal* 23.5 (2009), pp. 1541–1557.

[184] Yi Fang et al. "Higher-order oligomerization targets plasma membrane proteins and HIV Gag to exosomes". In: *PLoS Biology* 5.6 (2007), pp. 1267–1283.

[185] Jean Michel Escola et al. "Selective enrichment of tetraspan proteins on the internal vesicles of multivesicular endosomes and on exosomes secreted by human B-lymphocytes". In: *Journal of Biological Chemistry* 273.32 (1998), pp. 20121–20127.

[186] Stéphanie Charrin et al. "Lateral organization of membrane proteins: Tetraspanins spin their web". In: *Biochemical Journal* 420.2 (2009), pp. 133–154.

[187] Jr Ming Yang and Stephen J. Gould. "The cis-acting signals that target proteins to exosomes and microvesicles". In: *Biochemical Society Transactions* 41.1 (2013), pp. 277–282.

[188] Carolina Villarroya-beltri, Francesc Baixauli, and Cristina Gutiérrez-vázquez. "Sorting it out: Regulation of exosome loading". In: *Seminars in Cancer Biology* 2.28 (2015), pp. 3–13.

[189] Hong Zhu et al. "Mutation of SIMPLE in Charcot-Marie-Tooth 1C alters production of exosomes". In: *Molecular Biology of the Cell* 24.11 (2013), pp. 1619–1637.

[190] Ulrich Putz et al. "Nedd4 family-interacting protein 1 (Ndfip1) is required for the exosomal secretion of Nedd4 family proteins". In: *Journal of Biological Chemistry* 283.47 (2008), pp. 32621–32627.

[191] Marcel Kunadt et al. "Extracellular vesicle sorting of α-Synuclein is regulated by sumoylation". In: *Acta Neuropathologica* 129.5 (2015), pp. 695–713.

[192] Sushma Anand et al. "Ticket to a bubble ride: Cargo sorting into exosomes and extracellular vesicles". In: *Biochimica et Biophysica Acta - Proteins and Proteomics* 1867.12 (2019), p. 140203.

[193] Santosh Phuyal et al. "Regulation of exosome release by glycosphingolipids and flotillins". In: *FEBS Journal* 281.9 (2014), pp. 2214–2227.

[194] Melanie Meister and Ritva Tikkanen. "Endocytic trafficking of membrane-bound cargo: A flotillin point of view". In: *Membranes* 4.3 (2014), pp. 356–371.

[195] Eren M. Veziroglu and George I. Mias. "Characterizing Extracellular Vesicles and Their Diverse RNA Contents". In: *Frontiers in Genetics* 11 (2020).

[196] Ye Tian et al. "Protein Profiling and Sizing of Extracellular Vesicles from Colorectal Cancer Patients via Flow Cytometry". In: *ACS Nano* 12.1 (2018), pp. 671–680.

[197] Yusuke Yoshioka et al. “Comparative marker analysis of extracellular vesicles in different human cancer types”. In: *Journal of Extracellular Vesicles* 2.1 (2013).

[198] Yahui Ji et al. “Multiplexed profiling of single-cell extracellular vesicles secretion”. In: *Proceedings of the National Academy of Sciences of the United States of America* 116.13 (2019), pp. 5979–5984.

[199] Kévin Carayon et al. “Proteolipidic composition of exosomes changes during reticulocyte maturation”. In: *Journal of Biological Chemistry* 286.39 (2011), pp. 34426–34439.

[200] Yuchen Sun, Kosuke Saito, and Yoshiro Saito. “Lipid profile characterization and lipoprotein comparison of extracellular vesicles from human plasma and serum”. In: *Metabolites* 9.11 (2019), pp. 10–14.

[201] W. J. van Blitterswijk et al. “Comparative lipid analysis of purified plasma membranes and shed extracellular membrane vesicles from normal murine thymocytes and leukemic GRSL cells”. In: *Biochimica et Biophysica Acta - Biomembranes* 688.2 (1982), pp. 495–504.

[202] Alicia Llorente et al. “Molecular lipidomics of exosomes released by PC-3 prostate cancer cells”. In: *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* 1831.7 (2013), pp. 1302–1309.

[203] Todd A. Lydic et al. “Rapid and Comprehensive ‘Shotgun’ Lipidome Profiling of Colorectal Cancer Cell Derived Exosomes”. In: *Methods* 87 (2015), pp. 83–95.

[204] Caroline Subra et al. “Exosome lipidomics unravels lipid sorting at the level of multivesicular bodies”. In: *Biochimie* 89.2 (2007), pp. 205–212.

[205] Karine Laulagnier et al. “PLD2 is enriched on exosomes and its activity is correlated to the release of exosomes”. In: *FEBS Letters* 572.1-3 (2004), pp. 11–14.

[206] Laura Ann Mulcahy, Ryan Charles Pink, and David Raul Francisco Carter. “Routes and mechanisms of extracellular vesicle uptake”. In: *Journal of Extracellular Vesicles* 3.1 (2014).

[207] Nuria Izquierdo-Useros et al. “Capture and transfer of HIV-1 particles by mature dendritic cells converges with the exosome-dissemination pathway”. In: *Blood* 113.12 (2009), pp. 2732–2741.

[208] Aude De Gassart et al. “Lipid raft-associated protein sorting in exosomes”. In: *Blood* 102.13 (2003), pp. 4336–4344.

[209] T. Janas, Teresa Janas, and Michael Yarus. “Specific RNA binding to ordered phospholipid bilayers”. In: *Nucleic Acids Research* 34.7 (2006), pp. 2128–2136.

[210] Arsen O. Batagov and Igor V. Kurochkin. “Exosomes secreted by human cells transport largely mRNA fragments that are enriched in the 3-prime untranslated regions”. In: *Biology Direct* 8.12 (2013).

[211] Mu Li et al. “Analysis of the RNA content of the exosomes derived from blood serum and urine and its potential as biomarkerss”. In: *Philosophical Transactions of the Royal Society B* 369 (2014).

[212] Lorea Manterola et al. “A small noncoding RNA signature found in exosomes of GBM patient serum as a diagnostic tool”. In: *Neuro-Oncology* 16.4 (2014), pp. 520–527.

[213] Lucia Vojtech et al. “Exosomes in human semen carry a distinctive repertoire of small non-coding RNAs with potential regulatory functions”. In: *Nucleic Acids Research* 42.11 (2014), pp. 7290–7304.

[214] Linda Cambier et al. “ Y RNA fragment in extracellular vesicles confers cardioprotection via modulation of IL -10 expression and secretion ”. In: *EMBO Molecular Medicine* 9.3 (2017), pp. 337–352.

[215] Zhiyun Wei et al. “Coding and noncoding landscape of extracellular RNA released by human glioma stem cells”. In: *Nature Communications* 8.1 (2017).

[216] Diana J. Cha et al. “KRAS-dependent sorting of miRNA to exosomes”. In: *eLife* 4 (2015).

[217] Matthew J. Shurtleff et al. “Y-box protein 1 is required to sort microRNAs into exosomes in cells and in a cell-free reaction”. In: *eLife* 5 (2016), pp. 1–23.

[218] Orsolya Biró et al. “Circulating exosomal and Argonaute-bound microRNAs in preeclampsia”. In: *Gene* 692 (2019), pp. 138–144.

[219] Yun Teng et al. “MVP-mediated exosomal sorting of miR-193a promotes colon cancer progression”. In: *Nature Communications* 8 (2017).

[220] James W. Clancy et al. “An ARF6-Exportin-5 Axis Delivers pre-miRNA Cargo to Tumor Microvesicles”. In: *Nature Cell Biology* 21.7 (2019), pp. 856–866.

[221] D. Michiel Pegtel et al. “Functional delivery of viral miRNAs via exosomes”. In: *Proceedings of the National Academy of Sciences of the United States of America* 107.14 (2010), pp. 6328–6333.

[222] Jennifer Pérez-Boza, Michelle Lion, and Ingrid Struman. “Exploring the RNA landscape of endothelial exosomes”. In: *RNA* 24 (2018), pp. 423–435.

[223] Tom A.P. Driedonks et al. “Immune stimuli shape the small non-coding transcriptome of extracellular vesicles released by dendritic cells”. In: *Cellular and Molecular Life Sciences* 75.20 (2018), pp. 3857–3875.

[224] Scott A. Hinger et al. “Diverse Long RNAs Are Differentially Sorted into Extracellular Vesicles Secreted by Colorectal Cancer Cells”. In: *Cell Reports* 25.3 (2018), 715–725.e4.

[225] Mario Leonardo Squadrito et al. “Endogenous RNAs Modulate MicroRNA Sorting to Exosomes and Transfer to Acceptor Cells”. In: *Cell Reports* 8.5 (2014), pp. 1432–1446.

[226] Heedoo Lee et al. “Caveolin-1 selectively regulates microRNA sorting into microvesicles after noxious stimuli”. In: *Journal of Experimental Medicine* 216.9 (2019), pp. 2202–2220.

[227] Fengxia Lin et al. “YBX-1 mediated sorting of miR-133 into hypoxia/reoxygenation-induced EPC-derived exosomes to increase fibroblast angiogenesis and MEndoT”. In: *Stem Cell Research and Therapy* 10 (2019).

[228] Pin Lu et al. “MEX3C interacts with adaptor-related protein complex 2 and involves in miR-451a exosomal sorting”. In: *PLoS ONE* 12.10 (2017).

[229] Juan Pablo Tosar, Kenneth Witwer, and Alfonso Cayota. “Revisiting Extracellular RNA Release, Processing, and Function”. In: *Trends in Biochemical Sciences* (2020), pp. 1–8.

[230] Helena Sork et al. “Heterogeneity and interplay of the extracellular vesicle small RNA transcriptome and proteome”. In: *Scientific Reports* 8.1 (2018), pp. 1–12.

[231] Piroon Jenjaroenpun et al. “Characterization of RNA in exosomes secreted by human breast cancer cell lines using next-generation sequencing”. In: *PeerJ* 1 (2013), e201.

[232] Kevin C. Miranda et al. “Massively parallel sequencing of human urinary exosome/microvesicle RNA reveals a predominance of non-coding RNA”. In: *PLoS ONE* 9.5 (2014).

[233] Danijela Koppers-Lalic et al. “Nontemplated nucleotide additions distinguish the small RNA composition in cells from exosomes”. In: *Cell Reports* 8.6 (2014), pp. 1649–1658.

[234] Carolina Villarroya-Beltri et al. “Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs”. In: *Nature Communications* 4 (2013).

[235] Helena Sork et al. “Profiling of extracellular small RNAs highlights a strong bias towards non-vesicular secretion”. In: *bioRxiv* (2020).

[236] Ugur Gezer et al. “Long non-coding RNAs with low expression levels in cells are enriched in secreted exosomes”. In: *Cell Biology International* 38.9 (2014), pp. 1076–1079.

[237] Teresa Janas et al. “Mechanisms of RNA loading into exosomes”. In: *FEBS Letters* 589.13 (2015), pp. 1391–1398.

[238] Nobuyoshi Kosaka et al. “Secretory mechanisms and intercellular transfer of microRNAs in living cells”. In: *Journal of Biological Chemistry* 285.23 (2010), pp. 17442–17452.

[239] Masanori Fukushima et al. “StAR-related lipid transfer domain 11 (STARD11)-mediated ceramide transport mediates extracellular vesicle biogenesis”. In: *Journal of Biological Chemistry* 293.39 (2018), pp. 15277–15289.

[240] Bahnisikha Barman et al. “Biogenesis of RNA-containing extracellular vesicles at endoplasmic reticulum membrane contact sites. Authors:” in: *bioRxiv* (2020).

[241] Anastasia Khvorova et al. “RNAs that bind and change the permeability of phospholipid membranes”. In: *Proceedings of the National Academy of Sciences of the United States of America* 96.19 (1999), pp. 10649–10654.

[242] A. Vlassov, A. Khvorova, and M. Yarus. “Binding and disruption of phospholipid bilayers by supramolecular RNA complexes”. In: *Proceedings of the National Academy of Sciences of the United States of America* 98.14 (2001), pp. 7706–7711.

[243] Tadeusz Janas and Michael Yarus. “Visualization of membrane RNAs”. In: *RNA* 9.11 (2003), pp. 1353–1361.

[244] Teresa Janas, Tadeusz Janas, and Michael Yarus. “A membrane transporter for tryptophan composed of RNA”. In: *RNA* 10.10 (2004), pp. 1541–1549.

[245] Arsen O. Batagov, Vladimir A. Kuznetsov, and Igor V. Kurochkin. “Identification of nucleotide patterns enriched in secreted RNAs as putative cis-acting elements targeting them to exosome nano-vesicles”. In: *BMC Genomics* 12 (2011), S18.

[246] Natalia Szostak et al. “Sorting signal targeting mRNA into hepatic extracellular vesicles”. In: *RNA Biology* 11.7 (2014), pp. 836–844.

[247] Olga A. Kossinova et al. “Cytosolic YB-1 and NSUN2 are the only proteins recognizing specific motifs present in mRNAs enriched in exosomes”. In: *Biochimica et Biophysica Acta - Proteins and Proteomics* 1865.6 (2017), pp. 664–673.

[248] Mehmet Fatih Bolukbasi et al. “miR-1289 and “Zipcode”-like sequence enrich mRNAs in microvesicles”. In: *Molecular Therapy–Nucleic Acids* 1.2 (2012).

[249] Laura Santangelo et al. “The RNA-Binding Protein SYNCRIP Is a Component of the Hepatocyte Exosomal Machinery Controlling MicroRNA Sorting”. In: *Cell Reports* 17.3 (2016), pp. 799–808.

[250] Kamalika Mukherjee et al. “ Reversible HuR-micro RNA binding controls extracellular export of miR-122 and augments stress response ”. In: *EMBO Reports* 17.8 (2016), pp. 1184–1203.

[251] Suvendra N. Bhattacharyya et al. “Relief of microRNA-Mediated Translational Repression in Human Cells Subjected to Stress”. In: *Cell* 125.6 (2006), pp. 1111–1124.

[252] Zhelong Li et al. “In Vitro and in Vivo RNA Inhibition by CD9-HuR Functionalized Exosomes Encapsulated with miRNA or CRISPR/dCas9”. In: *Nano Letters* 19.1 (2019), pp. 19–28.

[253] Matthew J. Shurtleff et al. “Broad role for YBX1 in defining the small noncoding RNA composition of exosomes”. In: *Proceedings of the National Academy of Sciences of the United States of America* 114.43 (2017), E8987–E8995.

[254] Ann L. Wozniak et al. “The RNA binding protein FMR1 controls selective exosomal miRNA cargo loading during inflammation”. In: *The Journal of Cell Biology* 219.10 (2020).

[255] Forrest Weghorst et al. “Caspase-3 Cleaves Extracellular Vesicle Proteins During Auditory Brainstem Development”. In: *Frontiers in Cellular Neuroscience* 14 (2020).

[256] Andrew J. McKenzie et al. “KRAS-MEK Signaling Controls Ago2 Sorting into Exosomes”. In: *Cell Reports* 15.5 (2016), pp. 978–987.

[257] Andrey Turchinovich et al. “Characterization of extracellular circulating microRNA”. In: *Nucleic Acids Research* 39.16 (2011), pp. 7223–7233.

[258] Terence N. Bukong et al. “Exosomes from Hepatitis C Infected Patients Transmit HCV Infection and Contain Replication Competent Viral RNA in Complex with Ago2-miR122-HSP90”. In: *PLoS Pathogens* 10.10 (2014).

[259] Pierre Yves Mantel et al. “Infected erythrocyte-derived extracellular vesicles alter vascular function via regulatory Ago2-miRNA complexes in malaria”. In: *Nature Communications* 7 (2016).

[260] Sonia A. Melo et al. “Cancer Exosomes Perform Cell-Independent MicroRNA Biogenesis and Promote Tumorigenesis”. In: *Cancer Cell* 26.5 (2014), pp. 707–721.

[261] Anna M. Eiring et al. “miR-328 Functions as an RNA Decoy to Modulate hnRNP E2 Regulation of mRNA Translation in Leukemic Blasts”. In: *Cell* 140.5 (2010), pp. 652–665.

[262] Menghui Yin et al. “HNRNPA2B1 as an RNA Switch Trigger Modulate MicroRNA-Mediated Regulation of CDK6”. In: *Cell Press Sneak Peak* (2020).

[263] Andreas Zietzer et al. “The RNA-binding protein hnRNPU regulates the sorting of microRNA-30c-5p into large extracellular vesicles”. In: *Journal of Extracellular Vesicles* 9.1 (2020).

[264] Rossella Crescitelli et al. “Distinct RNA profiles in subpopulations of extracellular vesicles: Apoptotic bodies, microvesicles and exosomes”. In: *Journal of Extracellular Vesicles* 2.1 (2013).

[265] Cecilia Lässer et al. “Two distinct extracellular RNA signatures released by a single cell type identified by microarray and next-generation sequencing”. In: *RNA Biology* 14.1 (2017), pp. 58–72.

[266] Guillaume Van Niel, Gisela D’Angelo, and Graça Raposo. “Shedding light on the cell biology of extracellular vesicles”. In: *Nature Reviews Molecular Cell Biology* 19.4 (2018), pp. 213–228.

[267] Mathilde Mathieu et al. “Specificities of secretion and uptake of exosomes and other extracellular vesicles for cell-to-cell communication”. In: *Nature Cell Biology* 21.1 (2019), pp. 9–17.

[268] John R. Chevillet et al. “Quantitative and stoichiometric analysis of the microRNA content of exosomes”. In: *Proceedings of the National Academy of Sciences of the United States of America* 111.41 (2014), pp. 14888–14893.

[269] Manuel Albanese et al. “Micro RNAs are minor constituents of extracellular vesicles and are hardly delivered to target cells”. In: *bioRxiv* (2020).

[270] Wei Ying et al. “Adipose Tissue Macrophage-Derived Exosomal miRNAs Can Modulate in Vivo and in Vitro Insulin Sensitivity”. In: *Cell* 171.2 (2017), 372–384.e12.

[271] Erik R. Abels et al. “Glioblastoma-Associated Microglia Reprogramming Is Mediated by Functional Transfer of Extracellular miR-21”. In: *Cell Reports* 28.12 (2019), 3105–3119.e7.

[272] Farah Ghamloush et al. “The PAX3-FOXO1 oncogene alters exosome miRNA content and leads to paracrine effects mediated by exosomal miR-486”. In: *Scientific Reports* 9.1 (2019).

[273] Rocco Lucero et al. “Glioma-Derived miRNA-Containing Extracellular Vesicles Induce Angiogenesis by Reprogramming Brain Endothelial Cells”. In: *Cell Reports* 30.7 (2020), 2065–2074.e4.

[274] Meng Shen et al. “Chemotherapy-induced extracellular vesicle miRNAs promote breast cancer stemness by targeting OneCUT2”. In: *Cancer Research* 79.14 (2019), pp. 3608–3621.

[275] Kasey C Vickers et al. “MicroRNAs are Transported in Plasma and Delivered to Recipient Cells by High-Density Lipoproteins”. In: *Nature Cell Biology* 13.4 (2011), pp. 423–33.

[276] Jason D. Arroyo et al. “Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.12 (2011), pp. 5003–5008.

[277] Qin Zhang et al. “Transfer of Functional Cargo in Exomeres”. In: *Cell Reports* 27.3 (2019), 940–954.e6.

[278] Nasibeh Karimi et al. “Detailed analysis of the plasma extracellular vesicle proteome after separation from lipoproteins”. In: *Cellular and Molecular Life Sciences* 75.15 (2018), pp. 2873–2886.

[279] Sudipto K. Chakrabortty et al. “Extracellular vesicle-mediated transfer of processed and functional RNY5 RNA”. In: *RNA* 21.11 (2015), pp. 1966–1979.

[280] Katrin J. Svensson et al. “Exosome uptake depends on ERK1/2-heat shock protein 27 signaling and lipid raft-mediated endocytosis negatively regulated by caveolin-1”. In: *Journal of Biological Chemistry* 288.24 (2013), pp. 17713–17724.

[281] Helena Costa Verdera et al. “Cellular uptake of extracellular vesicles is mediated by clathrin-independent endocytosis and macropinocytosis”. In: *Journal of Controlled Release* 266 (2017), pp. 100–108.

[282] Sayo Horibe et al. “Mechanism of recipient cell-dependent differences in exosome uptake”. In: *BMC Cancer* 18 (2018).

[283] Dirk Fitzner et al. “Selective transfer of exosomes from oligodendrocytes to microglia by macropinocytosis”. In: *Journal of Cell Science* 124.3 (2011), pp. 447–458.

[284] Mathilde Chivet et al. “Exosomes secreted by cortical neurons upon glutamatergic synapse activation specifically interact with neurons”. In: *Journal of Extracellular Vesicles* 3.1 (2014).

[285] Cristina Grange et al. “Biodistribution of mesenchymal stem cell-derived extracellular vesicles in a model of acute kidney injury monitored by optical imaging”. In: *International Journal of Molecular Medicine* 33.5 (2014), pp. 1055–1063.

[286] Stefania Bruno et al. “Mesenchymal stem cell-derived microvesicles protect against acute tubular injury”. In: *Journal of the American Society of Nephrology* 20.5 (2009), pp. 1053–1067.

[287] Héctor Peinado et al. “Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET”. In: *Nature Medicine* 18.6 (2012), pp. 883–891.

[288] Kinsley C. French, Marc A. Antonyak, and Richard A. Cerione. “Extracellular vesicle docking at the cellular port: Extracellular vesicle binding and uptake”. In: *Seminars in Cell and Developmental Biology* 67 (2017), pp. 48–55.

[289] Sanyukta Rana and Margot Zöller. “Exosome target cell selection and the importance of exosomal tetraspanins: A hypothesis”. In: *Biochemical Society Transactions* 39.2 (2011), pp. 559–562.

[290] A. E. Morelli et al. “Endocytosis, intracellular sorting, and processing of exosomes by dendritic cells”. In: *Blood* 104.10 (2004), pp. 3257–3266.

[291] Irina Nazarenko et al. “Cell surface tetraspanin Tspan8 contributes to molecular pathways of exosome-induced endothelial cell activation”. In: *Cancer Research* 70.4 (2010), pp. 1668–1678.

[292] Sanyukta Rana et al. “Toward tailored exosomes: The exosomal tetraspanin web contributes to target cell selection”. In: *International Journal of Biochemistry and Cell Biology* 44.9 (2012), pp. 1574–1584.

[293] Helena C. Christianson et al. “Cancer cell exosomes depend on cell-surface heparan sulfate proteoglycans for their internalization and functional activity”. In: *Proceedings of the National Academy of Sciences of the United States of America* 110.43 (2013), pp. 17380–17385.

[294] Juan J. Garcia-Vallejo and Yvette van Kooyk. “The physiological role of DC-SIGN: A tale of mice and men”. In: *Trends in Immunology* 34.10 (2013), pp. 482–486.

[295] Cristina Escrevente et al. “Interaction and uptake of exosomes by ovarian cancer cells”. In: *BMC Cancer* 11.108 (2011).

[296] Sushrut Kamerkar et al. “Exosomes Facilitate Therapeutic Targeting of Oncogenic Kras in Pancreatic Cancer”. In: *Nature* 546.7659 (2017), pp. 498–503.

[297] Bong Hwan Sung et al. “A live cell reporter of exosome secretion and uptake reveals pathfinding behavior of migrating cells”. In: *Nature Communications* 11 (2020).

[298] Angela Montecalvo et al. “Mechanism of transfer of functional microRNAs between mouse dendritic cells via exosomes”. In: *Blood* 119.3 (2012), pp. 756–766.

[299] Qiuwei Pan et al. “Hepatic cell-to-cell transmission of small silencing RNA can extend the therapeutic reach of RNA interference (RNAi)”. In: *Gut* 61.9 (2012), pp. 1330–1339.

[300] M D Flanagan and S Lin. “Cytochalasins block actin filament elongation by binding to high affinity sites associated with F-actin”. In: *Journal of Biological Chemistry* 255 (1980), pp. 835–839.

[301] Christophe Lamaze et al. “The actin cytoskeleton is required for receptor-mediated endocytosis in mammalian cells”. In: *Journal of Biological Chemistry* 272.33 (1997), pp. 20332–20335.

[302] Gary J. Doherty and Harvey T. McMahon. “Mechanisms of endocytosis”. In: *Annual Review of Biochemistry* 78 (2009), pp. 857–902.

[303] Masaharu Somiya. “Where does the cargo go?: Solutions to provide experimental support for the “extracellular vesicle cargo transfer hypothesis””. In: *Journal of Cell Communication and Signaling* 14.2 (2020), pp. 135–146.

[304] Michelle E. Hung and Joshua N. Leonard. “A platform for actively loading cargo RNA to elucidate limiting steps in EV-mediated delivery”. In: *Journal of Extracellular Vesicles* 5.1 (2016).

[305] Emeline Bonsergent et al. “Quantitative characterization of extracellular vesicle uptake and content delivery within mammalian cells”. In: *Nature Communications* 12.1 (2021), pp. 1–11.

[306] Dhruvitkumar S. Sutaria et al. “Low active loading of cargo into engineered extracellular vesicles results in inefficient miRNA mimic delivery”. In: *Journal of Extracellular Vesicles* 6.1 (2017).

[307] Masamitsu Kanada et al. “Differential fates of biomolecules delivered to target cells via extracellular vesicles”. In: *Proceedings of the National Academy of Sciences of the United States of America* 112.12 (2015), E1433–E1442.

[308] Avash Das et al. “Red blood cell-derived extracellular vesicles mediate intercellular communication in ischemic heart failure”. In: *bioRxiv* (2019).

[309] Kirsten Ridder et al. “Extracellular vesicle-mediated transfer of functional RNA in the tumor microenvironment”. In: *OncoImmunology* 4.6 (2015), e1008371.

[310] Kirsten Ridder et al. “Extracellular Vesicle-Mediated Transfer of Genetic Information between the Hematopoietic System and the Brain in Response to Inflammation”. In: *PLoS Biology* 12.6 (2014).

[311] Anoek Zomer et al. “Studying extracellular vesicle transfer by a Cre-loxP method”. In: *Nature Protocols* 11.1 (2015), pp. 87–101.

[312] Anoek Zomer et al. “In Vivo imaging reveals extracellular vesicle-mediated phenocopying of metastatic behavior.” In: *Cell* 161.5 (2015), pp. 1046–57.

[313] Ulrich Sterzenbach et al. “Engineered Exosomes as Vehicles for Biologically Active Proteins”. In: *Molecular Therapy* 25.6 (2017), pp. 1269–1278.

[314] Ling Li et al. “Extracellular vesicles carry miR-195 to intrahepatic cholangiocarcinoma and improve survival in a rat model”. In: *Hepatology* 65.2 (2017), pp. 501–514.

[315] Shin Ichiro Ohno et al. “Systemically injected exosomes targeted to EGFR deliver antitumor microrna to breast cancer cells”. In: *Molecular Therapy* 21.1 (2013), pp. 185–191.

[316] Lydia Alvarez-Erviti et al. “Delivery of siRNA to the mouse brain by systemic injection of targeted exosomes”. In: *Nature Biotechnology* 29.4 (2011), pp. 341–345.

[317] Marie Cécile Didiot et al. “Exosome-mediated delivery of hydrophobically modified siRNA for huntingtin mRNA silencing”. In: *Molecular Therapy* 24.10 (2016), pp. 1836–1847.

[318] Olivier G. de Jong et al. “A CRISPR-Cas9-based reporter system for single-cell detection of extracellular vesicle-mediated functional transfer of RNA”. In: *Nature Communications* 11.1 (2020), pp. 1–13.

[319] Margaret Alexander et al. “Exosome-delivered microRNAs modulate the inflammatory response to endotoxin”. In: *Nature Communications* 6 (2015).

[320] Nambin Yim et al. “Exosome engineering for efficient intracellular delivery of soluble proteins using optically reversible protein-protein interaction module”. In: *Nature Communications* 7 (2016), pp. 1–9.

[321] Arda Mizrak et al. “Genetically engineered microvesicles carrying suicide mRNA/protein inhibit schwannoma tumor growth”. In: *Molecular Therapy* 21.1 (2013), pp. 101–108.

[322] Michael J. Munson et al. “A high-throughput Galectin-9 imaging assay for quantifying nanoparticle uptake, endosomal escape and functional RNA delivery”. In: *Communications Biology* 4 (2021).

[323] Tomoya Hikita et al. “Sensitive and rapid quantification of exosomes by fusing luciferase to exosome marker proteins”. In: *Scientific Reports* 8 (2018).

[324] Serena Rubina Baglio et al. “Blocking tumor-educated MSC paracrine activity halts osteosarcoma progression”. In: *Clinical Cancer Research* 23.14 (2017), pp. 3721–3733.

[325] Paula Carpintero-Fernández, Juan Fafián-Labora, and Ana O’Loghlen. “Technical Advances to Study Extracellular Vesicles”. In: *Frontiers in Molecular Biosciences* 4 (2017).

[326] Charles P. Lai et al. “Visualization and tracking of tumour extracellular vesicle delivery and RNA translation using multiplexed reporters”. In: *Nature Communications* 6.May (2015), pp. 1–12.

[327] Du Feng et al. “Cellular internalization of exosomes occurs through phagocytosis”. In: *Traffic* 11.5 (2010), pp. 675–687.

[328] Carrie A. Franzen et al. “Characterization of uptake and internalization of exosomes by bladder cancer cells”. In: *BioMed Research International* 2014 (2014).

[329] H. D. Roberts-Dalton et al. “Fluorescence labelling of extracellular vesicles using a novel thiol-based strategy for quantitative analysis of cellular delivery and intracellular traffic”. In: *Nanoscale* 9.36 (2017), pp. 13693–13706.

[330] Alessandra Lo Cicero et al. “Exosomes released by keratinocytes modulate melanocyte pigmentation”. In: *Nature Communications* 6 (2015).

[331] Frederik Johannes Verweij et al. “Quantifying exosome secretion from single cells reveals a modulatory role for GPCR signaling”. In: *Journal of Cell Biology* (2018).

[332] Frederik J. Verweij et al. “Live Tracking of Inter-organ Communication by Endogenous Exosomes In Vivo”. In: *Developmental Cell* 48.4 (2019), 573–589.e4.

[333] Víctor Toribio et al. “Development of a quantitative method to measure EV uptake”. In: *Scientific Reports* 9.1 (2019), p. 10522.

[334] Killian O’Brien et al. “RNA delivery by extracellular vesicles in mammalian cells and its applications”. In: *Nature Reviews Molecular Cell Biology* (2020).

[335] Brandon M. Lehrich et al. “Fetal bovine serum-derived extracellular vesicles persist within vesicle-depleted culture media”. In: *International Journal of Molecular Sciences* 19.11 (2018).

[336] Juan Pablo Tosar et al. “Ribonucleic artefacts: Are some extracellular RNA discoveries driven by cell culture medium components?” In: *Journal of Extracellular Vesicles* 6 (2017).

[337] Jan Lötvall et al. “Minimal experimental requirements for definition of extracellular vesicles and their functions: A position statement from the International Society for Extracellular Vesicles”. In: *Journal of Extracellular Vesicles* 3 (2014).

[338] Kazuyo Fujita et al. “Induction of lipid droplets in non-macrophage cells as well as macrophages by liposomes and exosomes”. In: *Biochemical and Biophysical Research Communications* 510.1 (2019), pp. 184–190.

[339] Ádám Oszvald et al. “Fibroblast-Derived Extracellular Vesicles Induce Colorectal Cancer Progression by Transmitting Amphiregulin”. In: *Frontiers in Cell and Developmental Biology* 8 (2020).

[340] Anikó Zeöld et al. “Shared extracellular vesicle miRNA profiles of matched ductal pancreatic adenocarcinoma organoids and blood plasma samples show the power of organoid technology”. In: *Cellular and Molecular Life Sciences* 78.6 (2020).

[341] Divya B. Patel et al. “Impact of cell culture parameters on production and vascularization bioactivity of mesenchymal stem cell-derived extracellular vesicles”. In: *Bioengineering & Translational Medicine* 2.2 (2017), pp. 170–179.

[342] Albert Van Den Berg et al. “Personalised organs-on-chips: functional testing for precision medicine”. In: *Lab on a Chip* 19.2 (2019), pp. 198–205.

[343] Charles P. Lai, Bakhos A. Tannous, and Xandra O. Breakefield. “Noninvasive In Vivo Monitoring of Extracellular Vesicles”. In: *Methods in Molecular Biology* 1098 (2014), pp. 249–258.

[344] Veronica Go et al. “Extracellular vesicles from mesenchymal stem cells reduce microglial-mediated neuroinflammation after cortical injury in aged Rhesus monkeys”. In: *GeroScience* 42 (2020).

[345] Mallikarjun Bidarimath et al. “Extracellular vesicle mediated intercellular communication at the porcine maternal-fetal interface: A new paradigm for conceptus-endometrial cross-talk”. In: *Scientific Reports* 7 (2017), pp. 1–14.

[346] Jorge V Beira and Renato Paro. “The legacy of Drosophila imaginal discs.” In: *Chromosoma* 125.4 (2016), pp. 573–92.

[347] Seth S. Blair. “Imaginal Discs”. In: *Encyclopedia of Insects* (2009), pp. 489–492.

[348] Thomas Flatt. “Life-history evolution and the genetics of fitness components in drosophila melanogaster”. In: *Genetics* 214.1 (2020), pp. 3–48.

[349] Brigitte de Saint Phalle. “Immunostaining of whole-mount imaginal discs.” In: *Methods in Molecular Biology* 247 (2004), pp. 373–387.

[350] Barbara H. Jennings. “Drosophila-a versatile model in biology & medicine”. In: *Materials Today* 14.5 (2011), pp. 190–195.

[351] Melda Tozluoglu and Yanlan Mao. “On folding morphogenesis, a mechanical problem: Folding morphogenesis”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1809 (2020).

[352] Rachel Smith-Bolton. “Drosophila Imaginal Discs as a Model of Epithelial Wound Repair and Regeneration”. In: *Advances in Wound Care* 5.6 (2016), pp. 251–261.

[353] Joseph A. Waldron and Sarah F. Newbury. “The roles of miRNAs in wing imaginal disc development in Drosophila”. In: *Biochemical Society Transactions* 40.4 (2012), pp. 891–895.

[354] David Foronda, Ainhoa Pérez-Garijo, and Francisco A. Martín. “Dpp of posterior origin patterns the proximal region of the wing”. In: *Mechanisms of Development* 126.3-4 (2009), pp. 99–106.

[355] Caitlin E. Fogarty et al. “Extracellular Reactive Oxygen Species Drive Apoptosis-Induced Proliferation via Drosophila Macrophages”. In: *Current Biology* 26.5 (2016), pp. 575–584.

[356] Ginés Morata and Manuel Calleja. "Cell competition and tumorigenesis in the imaginal discs of Drosophila". In: *Seminars in Cancer Biology* (2019).

[357] Michael F. Wangler, Shinya Yamamoto, and Hugo J. Bellen. "Fruit flies in biomedical research". In: *Genetics* 199.3 (2015), pp. 639–653.

[358] Tinri Aegerter-Wilmsen et al. "Model for the regulation of size in the wing imaginal disc of Drosophila". In: *Mechanisms of Development* 124.4 (2007), pp. 318–326.

[359] Marco Milán, Sonsoles Campuzano, and Antonio García-Bellido. "Cell cycling and patterned cell proliferation in the wing primordium of Drosophila". In: *Proceedings of the National Academy of Sciences of the United States of America* 93.2 (1996), pp. 640–645.

[360] Marcos Nahmad and Angelike Stathopoulos. "Dynamic interpretation of hedgehog signaling in the Drosophila wing disc". In: *PLoS Biology* 7.9 (2009).

[361] Brent Wells, Eri Yoshida, and Laura A. Johnston. "Compensatory Proliferation in Drosophila Imaginal Discs Requires Dronc-Dependent p53 Activity". In: *Current Biology* 16.16 (2006), pp. 1606–1615.

[362] Ana Citlali Gradilla et al. "Exosomes as Hedgehog carriers in cytoneme-mediated transport and secretion". In: *Nature Communications* 5 (2014).

[363] Miranda J. Butler et al. "Discovery of genes with highly restricted expression patterns in the Drosophila wing disc using DNA oligonucleotide microarrays". In: *Development* 130.4 (2003), pp. 659–670.

[364] Ghulam Hassan Dar et al. "GAPDH controls extracellular vesicle biogenesis and enhances therapeutic potential of EVs in silencing the Huntingtin gene in mice via siRNA delivery". In: *bioRxiv* (2020).

[365] Ruth E. Thomas et al. "Glucocerebrosidase deficiency promotes protein aggregation through dysregulation of extracellular vesicles". In: *PLoS Genetics* 14.9 (2018), pp. 1–32.

[366] Fabio Alexis Lefebvre et al. "Comparative transcriptomic analysis of human and Drosophila extracellular vesicles". In: *Scientific Reporters* 6 (2016).

[367] A. Nern et al. "Multiple new site-specific recombinases for use in manipulating animal genomes". In: *Proceedings of the National Academy of Sciences of the United States of America* 108.34 (2011), pp. 14198–14203.

[368] Luis Alberto Baena-Lopez et al. "Novel initiator caspase reporters uncover unknown features of caspase-activating cells". In: *Development* (2018), dev.170811.

[369] Johannes Bischof et al. "An optimized transgenesis system for Drosophila using germ-line-specific ϕC31 integrases". In: *Proceedings of the National Academy of Sciences of the United States of America* 104.9 (2007), pp. 3312–3317.

[370] Luis Alberto Baena-Lopez et al. "Accelerated homologous recombination and subsequent genome modification in Drosophila". In: *Development* 140.23 (2013), pp. 4818–4825.

[371] Lucy Cherbas et al. "The transcriptional diversity of 25 Drosophila cell lines". In: *Genome Research* 21.2 (2011), pp. 301–314.

[372] Warren Strober. "Trypan Blue Exclusion Test of Cell Viability". In: *Current Protocols in Immunology* 21.1 (2001), A.3B.1–A.3B.2.

[373] Maria Arantzazu Sanchez-Fernandez et al. "Transgenic mice for a tamoxifen-induced, conditional expression of the Cre recombinase in osteoclasts". In: *PLoS ONE* 7.5 (2012).

[374] Daijun Ling and Paul M. Salvaterra. "Robust RT-qPCR data normalization: Validation and selection of internal reference genes during post-experimental data analysis". In: *PLoS ONE* 6.3 (2011).

[375] Lewis Arthurton et al. "Non-apoptotic caspase activation preserves Drosophila intestinal progenitor cells in quiescence". In: *EMBO Reports* 44 (2020).

[376] Fernández-Moreno MA et al. "Drosophila melanogaster as a Model System to Study Mitochondrial Biology". In: *Physiology & Behavior* 372 (2007), pp. 33–49.

[377] Carrie M. Spratford and Justin P. Kumar. "Dissection and immunostaining of imaginal discs from drosophila melanogaster". In: *Journal of Visualized Experiments* 91 (2014).

[378] Johannes Schindelin et al. "Fiji: An open-source platform for biological-image analysis". In: *Nature Methods* 9.7 (2012), pp. 676–682.

[379] Caroline A. Schneider, Wayne S. Rasband, and Kevin W. Eliceiri. "NIH Image to ImageJ: 25 years of image analysis". In: *Nature Methods* 9.7 (2012), pp. 671–675.

[380] Michael R. Lamprecht, David M. Sabatini, and Anne E. Carpenter. "CellProfiler™: Free, versatile software for automated biological image analysis". In: *BioTechniques* 42.1 (2007), pp. 71–75.

[381] Thomas A Caswell et al. *Matplotlib (v. 3.2.1)*. 2020.

[382] Michael Waskom and the seaborn development team. *mwaskom/seaborn*. Version latest. Sept. 2020. URL: `https://doi.org/10.5281/zenodo.592845`.

[383] Bob Dransfield and Bob Brightwell. *Binomial & Poisson Distributions - Principles*. URL: `https://influentialpoints.com/Training/binomial_and_poisson_distributions-principles-properties-assumptions.htm`.

[384] Bob Dransfield and Bob Brightwell. *z-test for independent proportions - Principles*. URL: `https://influentialpoints.com/Training/z-test_for_independent_proportions-principles-properties-assumptions.htm`.

[385] Graphpad. *QuickCalcs*. URL: `https://www.graphpad.com/quickcalcs/`.

[386] Enis Afgan et al. "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update". In: *Nucleic Acids Research* 44.W1 (2016), W3–W10.

[387] Martin Marcel. "Cutadapt removes adapter sequences from high-throughput sequencing reads". In: *Advances in Environmental Biology* 7.10 (2013), pp. 2803–2809.

[388] Simon Andrews et al. *FastQC*. Babraham Institute. Babraham, UK, Jan. 2010.

[389] Philip Ewels et al. "MultiQC: Summarize analysis results for multiple tools and samples in a single report". In: *Bioinformatics* 32.19 (2016), pp. 3047–3048.

[390] Fiona Cunningham et al. “Ensembl 2019”. In: *Nucleic Acids Research* 47.D1 (2018), pp. D745–D751.

[391] Brian Bushnell. *BBMap*. 2014. URL: https://sourceforge.net/projects/bbmap/.

[392] Alexander Dobin et al. “STAR: Ultrafast universal RNA-seq aligner”. In: *Bioinformatics* 29.1 (2013), pp. 15–21.

[393] Simon Anders, Paul Theodor Pyl, and Wolfgang Huber. “HTSeq-A Python framework to work with high-throughput sequencing data”. In: *Bioinformatics* 31.2 (2015), pp. 166–169.

[394] Michael I. Love, Wolfgang Huber, and Simon Anders. “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. In: *Genome Biology* 15.12 (2014), pp. 1–21.

[395] Scikit-learn developers. *scikit-learn*. 2019. URL: https://scikit-learn.org/stable/whats%7B%5C_%7Dnew/v0.23.html.

[396] Bartosz Mikulski. *PCA — how to choose the number of components?* 2019. URL: https://www.mikulskibartosz.name/pca-how-to-choose-the-number-of-components/.

[397] Wes McKinney. “Data Structures for Statistical Computing in Python”. In: Python in Science Conference, 2010, pp. 56–61.

[398] The Python3 development Team. *Python3*. 2019. URL: https://docs.python.org/3.8/.

[399] The Pandas Development Team. *Pandas*. 2020. URL: https://doi.org/10.5281/zenodo.3509134.

[400] Jupyter Development Team. *Jupyter Notebook*. 2015. URL: https://jupyter.org/index.html.

[401] Stan Brown. *Measures of Shape: Skewness and Kurtosis*. 2016. URL: https://brownmath.com.

[402] Murray R. Spiegel and Larry J. Stephens. *Theory and Problems of Statistics*. 3rd. 1999.

[403] M. G Bulmer. *Principles of Statistics*. Dover, 1979.

[404] Duncan Cramer. *Basic Statistics for Social Research*. 1997.

[405] Cole Trapnell et al. “Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation”. In: *Nature Biotechnology* 28.5 (2010), pp. 511–515.

[406] Adam Roberts et al. “Improving RNA-Seq expression estimates by correcting for fragment bias”. In: *Genome Biology* 12.3 (2011).

[407] Adam Roberts et al. “Identification of novel transcripts in annotated genomes using RNA-seq”. In: *Bioinformatics* 27.17 (2011), pp. 2325–2329.

[408] Cole Trapnell et al. “Differential analysis of gene regulation at transcript resolution with RNA-seq”. In: *Nature Biotechnology* 31.1 (2013), pp. 46–53.

[409] Traver Hart et al. “Finding the active genes in deep RNA-seq gene expression studies”. In: *BMC Genomics* 14.778 (2013).

[410] RStudio Team. *RStudio: Integrated Development Environment for R.* RStudio, Inc. Boston, MA, 2015. URL: `http://www.rstudio.com/`.

[411] R Core Team. *R: A language and environment for statistical computing.* Vienna, Austria, 2013. URL: `http://www.r-project.org/`.

[412] Aurélien Brionne, Amélie Juanchich, and Christelle Hennequet-Antier. "ViSEAGO: A Bioconductor package for clustering biological functions using Gene Ontology and semantic similarity". In: *BioData Mining* 12.1 (2019), pp. 1–13.

[413] Aurélien Brionne, Amélie Juanchich, and Hennequet-Antier Christelle. *An overview of ViSEAGO: Visualisation, Semantic similarity, Enrichment Analysis of Gene Ontology.* 2020. URL: `http://bioconductor.org/packages/devel/bioc/vignettes/ViSEAGO/inst/doc/ViSEAGO.html`.

[414] Timothy L. Bailey et al. "The MEME Suite". In: *Nucleic Acids Research* 43.W1 (2015), W39–W49.

[415] Timothy L. Bailey and Charles Elkan. "Fitting a mixture model by expectation maximization to discover motifs in biopolymers". In: *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* 2 (1994), pp. 28–36.

[416] Charles E. Grant, Timothy L. Bailey, and William Stafford Noble. "FIMO: Scanning for occurrences of a given motif". In: *Bioinformatics* 27.7 (2011), pp. 1017–1018.

[417] Daniel Blankenberg et al. "A framework for collaborative analysis of ENCODE data: Making large-scale analyses biologist-friendly". In: *Genome Research* 17.6 (2007), pp. 960–964.

[418] Peter Rice, Lan Longden, and Alan Bleasby. "EMBOSS: The European Molecular Biology Open Software Suite". In: *Trends in Genetics* 16.6 (2000), pp. 276–277.

[419] Timothy L. Bailey and Michael Gribskov. "Combining evidence using p-values: Application to sequence homology searches". In: *Bioinformatics* 14.1 (1998), pp. 48–54.

[420] Debashish Ray et al. "A compendium of RNA-binding motifs for decoding gene regulation". In: *Nature* 499.7457 (2013), pp. 172–177.

[421] Shobhit Gupta et al. "Quantifying similarity between motifs". In: *Genome Biology* 8.2 (2007).

[422] Ronny Lorenz et al. "ViennaRNA Package 2.0". In: *Algorithms for Molecular Biology* 6.1 (2011), pp. 1–14.

[423] Saurabh Agarwal et al. "Sequencing of first-strand cDNA library reveals full-length transcriptomes". In: *Nature Communications* 6 (2015).

[424] Liguo Wang et al. "Measure transcript integrity using RNA-seq data". In: *BMC Bioinformatics* 17.1 (2016), pp. 1–16.

[425] Jonathan Houseley and David Tollervey. "The Many Pathways of RNA Degradation". In: *Cell* 136.4 (2009), pp. 763–776.

[426] Nicole L. Garneau, Jeffrey Wilusz, and Carol J. Wilusz. "The highways and byways of mRNA decay". In: *Nature Reviews Molecular Cell Biology* 8.2 (2007), pp. 113–126.

[427] Ali Mortazavi et al. “Mapping and quantifying mammalian transcriptomes by RNA-Seq”. In: *Nature Methods* 5.7 (2008), pp. 621–628.

[428] Beat Suter. “RNA localization and transport”. In: *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1861.10 (2018), pp. 938–951.

[429] Caspar Glock, Maximilian Heumüller, and Erin M. Schuman. “mRNA transport & local translation in neurons”. In: *Current Opinion in Neurobiology* 45 (2017), pp. 169–177.

[430] Véronique Van De Bor et al. “gurken and the I factor retrotransposon RNAs share common localization signals and machinery”. In: *Developmental Cell* 9.1 (2005), pp. 51–62.

[431] James M. Halstead et al. “Syncrip/hnRNP Q influences synaptic transmission and regulates BMP signaling at the Drosophila neuromuscular synapse”. In: *Biology Open* 3.9 (2014), pp. 839–849.

[432] L. Shannon Holliday, Lorraine Perciliano de Faria, and Wellington J. Rody. “Actin and actin-associated proteins in extracellular vesicles shed by osteoclasts”. In: *International Journal of Molecular Sciences* 21.1 (2020).

[433] Pearl V. Ryder and Dorothy A. Lerit. “RNA localization regulates diverse and dynamic cellular processes”. In: *Traffic* 19.7 (2018), pp. 496–502.

[434] Yaron Shav-Tal and Robert H. Singer. “RNA localisation”. In: *Journal of Cell Science* 118.16 (2005), pp. 4077–4081.

[435] Alexander G. Baltz et al. “The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts”. In: *Molecular Cell* 46.5 (2012), pp. 674–690.

[436] Kai Blin et al. “DoRiNA 2.0-upgrading the dorina database of RNA interactions in post-transcriptional regulation”. In: *Nucleic Acids Research* 43.D1 (2015), pp. D160–D167.

[437] Eric L Van Nostrand et al. “A Large-Scale Binding and Functional Map of Human RNA Binding Proteins”. In: *bioRxiv* (2018), p. 179648.

[438] Flavio Mignone et al. “Untranslated regions of mRNAs”. In: *Genome Biology* 3.3 (2002), pp. 1–10.

[439] Georgi Tushev et al. “Alternative 3-prime UTRs Modify the Localization, Regulatory Potential, Stability, and Plasticity of mRNAs in Neuronal Compartments”. In: *Neuron* 98.3 (2018), 495–511.e6.

[440] Thermo Fisher Scientific. *T042-Technical Bulletin NanoDrop Spectrophotometers.* Tech. rep. 2010.

[441] Philippe Desjardins and Deborah Conklin. “NanoDrop microvolume quantitation of nucleic acids”. In: *Journal of Visualized Experiments* 45 (2010).

[442] Qiagen. *Sample to Insight Application Note: Effects of low A 260 /A 230 ratios in RNA preparations on downstream applications.* Tech. rep. 2018, pp. 1–6.

[443] Bogdan Mateescu et al. “Obstacles and opportunities in the functional analysis of extracellular vesicle RNA - An ISEV position paper”. In: *Journal of Extracellular Vesicles* 6.1 (2017).

[444] Ulrike Lambertz et al. “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”. In: *BMC Genomics* 16.1 (2015).

[445] Tonje Fiskaa et al. “Distinct small RNA signatures in extracellular vesicles derived from breast cancer cell lines”. In: *PLoS ONE* 11.8 (2016).

[446] Zhihua Li et al. “Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs”. In: *Nucleic Acids Research* 40.14 (2012), pp. 6787–6799.

[447] Rebecca L. Hurto. “Unexpected functions of tRNA and tRNA processing enzymes”. In: *Advances in Experimental Medicine and Biology* 722 (2011), pp. 137–155.

[448] Esther N.M. Nolte’T Hoen et al. “Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions”. In: *Nucleic Acids Research* 40.18 (2012), pp. 9272–9285.

[449] Tim R. Mercer et al. “Regulated post-transcriptional RNA cleavage diversifies the eukaryotic transcriptome”. In: *Genome Research* 20.12 (2010), pp. 1639–1650.

[450] Shayne A. Bellingham, Bradley M. Coleman, and Andrew F. Hill. “Small RNA deep sequencing reveals a distinct miRNA signature released in exosomes from prion-infected neuronal cells”. In: *Nucleic Acids Research* 40.21 (2012), pp. 10937–10949.

[451] Tim R. Mercer et al. “Expression of distinct RNAs from 3-prime untranslated regions”. In: *Nucleic Acids Research* 39.6 (2011), pp. 2393–2403.

[452] Marcel E. Dinger et al. “Differentiating protein-coding and noncoding RNA: Challenges and ambiguities”. In: *PLoS Computational Biology* 4.11 (2008).

[453] Katrin Wiederhold and Lori A. Passmore. “Cytoplasmic deadenylation: Regulation of mRNA fate”. In: *Biochemical Society Transactions* 38.6 (2010), pp. 1531–1536.

[454] Jun Cheng et al. “Cis-regulatory elements explain most of the mRNA stability variation across genes in yeast”. In: *RNA* 23.11 (2017), pp. 1648–1659.

[455] Xiaokan Zhang, Anders Virtanen, and Frida E. Kleiman. “To polyadenylate or to deadenylate: That is the question”. In: *Cell Cycle* 9.22 (2010), pp. 4437–4449.

[456] Illumina. *Considerations for RNA-Seq read length and coverage*. Tech. rep. 2020.

[457] Elena S. Martens-Uzunova, Michael Olvedy, and Guido Jenster. “Beyond microRNA - Novel RNAs derived from small non-coding RNA and their implication in cancer”. In: *Cancer Letters* 340.2 (2013), pp. 201–211.

[458] Melissa W. Li et al. “Nuclear export factor 3 regulates localization of small nucleolar RNAs”. In: *Journal of Biological Chemistry* 292.49 (2017), pp. 20228–20239.

[459] Christopher L. Holley et al. “Cytosolic accumulation of small nucleolar RNAs (snoRNAs) is dynamically regulated by NADPH oxidase”. In: *Journal of Biological Chemistry* 290.18 (2015), pp. 11741–11748.

[460] Laura Poliseno et al. “A coding-independent function of gene and pseudogene mRNAs regulates tumour biology”. In: *Nature* 465.7301 (2010), pp. 1033–1038.

[461] Jess Morhayima et al. “Molecular characterization of human osteoblast-derived extracellular vesicle mRNA using next-generation sequencing”. In: *Biochimica et Biophysica Acta - Molecular Cell Research* 1864.7 (2017), pp. 1133–1141.

[462] Darya D. Yanshina et al. “Structural features of the interaction of the 3-prime-untranslated region of mRNA containing exosomal RNA-specific motifs with YB-1, a potential mediator of mRNA sorting”. In: *Biochimie* 144 (2018), pp. 134–143.

[463] Singh Guramrit et al. “The clothes make the mRNA: Past and present trends in mRNP fashion”. In: *Annual Review of Biochemistry* (2015), pp. 229–262.

[464] Patrick C. Gilligan et al. “Conservation defines functional motifs in the squint/nodal-related 1 RNA dorsal localization element”. In: *Nucleic Acids Research* 39.8 (2011), pp. 3340–3349.

[465] Juha Grönholm et al. “Not4 enhances JAK/STAT pathway-dependent gene expression in Drosophila and in human cells”. In: *The FASEB Journal* 26.3 (2012), pp. 1239–1250.

[466] Jenni Kallio et al. “Eye transformer is a negative regulator of Drosophila JAK/STAT signaling”. In: *The FASEB Journal* 24.11 (2010), pp. 4467–4479.

[467] Pascale Gaudet et al. “Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium”. In: *Briefings in Bioinformatics* 12.5 (2011), pp. 449–462.

[468] Jim Thurmond et al. “FlyBase 2.0: The next generation”. In: *Nucleic Acids Research* 47.D1 (2019), pp. D759–D765.

[469] Florencia Cano et al. “A non-proteolytic role for ubiquitin in deadenylation of MHC-I mRNA by the RNA-binding E3-ligase MEX-3C”. In: *Nature Communications* 6 (2015).

[470] Shivakumar Keerthikumar et al. “Proteogenomic analysis reveals exosomes are more oncogenic than ectosomes”. In: *Oncotarget* 6.17 (2015), pp. 15375–15396.

[471] Bing Liang et al. “Characterization and proteomic analysis of ovarian cancer-derived exosomes”. In: *Journal of Proteomics* 80 (2013), pp. 171–182.

[472] Peng Yao et al. “Interplay between miR-574-3p and hnRNP L regulates VEGFA mRNA translation and tumorigenesis”. In: *Nucleic Acids Research* 45.13 (2017), pp. 7950–7964.

[473] Geoffroy Walbrecq et al. “Hypoxia-induced adaptations of mirnomes and proteomes in melanoma cells and their secreted extracellular vesicles”. In: *Cancers* 12.3 (2020).

[474] K. Jin Kim et al. “Inhibition of vascular endothelial growth factor-induced angiogenesis suppresses tumour growth in vivo”. In: *Nature* 362.6423 (1993), pp. 841–844.

[475] Napoleone Ferrara et al. “Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene”. In: *Nature* 380.6573 (1996), pp. 439–442.

[476] Juliane Nguyen and Scott Ferguson. *Compositions and methods for loading extracellular vesicles with chemical and biological agents/molecules (Patent WO 2018/209182 A2)*. 2018.

[477] Kathrin Theil, Margareta Herzog, and Nikolaus Rajewsky. "Post-transcriptional regulation by 3-prime UTRs can be masked by regulatory elements in 5-prime UTRs". In: *Cell Reports* 22.12 (2018), pp. 3217–3226.

[478] Greg J. Bashaw and Bruce S. Baker. "The regulation of the Drosophila msl-2 gene reveals a function for sex- lethal in translational control". In: *Cell* 89.5 (1997), pp. 789–798.

[479] Richard L. Kelley et al. "Sex lethal controls dosage compensation in Drosophila by a non-splicing mechanism". In: *Nature* 387.6629 (1997), pp. 195–199.

[480] Wei Seong Toh et al. "MSC exosome works through a protein-based mechanism of action". In: *Biochemical Society Transactions* 46.4 (2018), pp. 843–853.

[481] Xiaoyi Huang et al. "Characterization of human plasma-derived exosomal RNAs by deep sequencing". In: *BMC Genomics* 14 (2013).

[482] Chi Zhang et al. "Bioinformatics Tools for RNA-seq Gene and Isoform Quantification". In: *Journal of Next Generation Sequencing and Applications* 3.3 (2016).

[483] Takashi Kawashima et al. "Synaptic activity-responsive element in the Arc / Arg3.1 promoter essential for synapse-to-nucleus signaling". In: *Proceedings of the National Academy of Sciences of the United States of America* 106.1 (2008), pp. 316–321.

[484] Robert H. Singer. "The cytoskeleton and mRNA localization". In: *Cell Biology* 4.1 (1992), pp. 15–19.

[485] Caroline Medioni, Kimberly Mowry, and Florence Besse. "Principles and roles of mRNA localization in animal development". In: *Development* 139.18 (2012), pp. 3263–3276.

[486] Edward H Kislauskis and Robert H Singer. "Determinants of mRNA localization". In: *Cell Biology* 4 (1992), pp. 975–978.

[487] Michael D. Blower. "Molecular insights into intracellular RNA localization". In: *International Review of Cell and Molecular Biology* 302 (2013).

[488] Sebastian Baumann et al. "Kinesin-3 and dynein mediate microtubule-dependent co-transport of mrnps and endosomes". In: *Journal of Cell Science* 125.11 (2012), pp. 2740–2752.

[489] Gretchen Dollar et al. "Rab11 polarization of the Drosophila oocyte: A novel link between membrane trafficking, microtubule organization, and oskar mRNA localization and translation". In: *Development* 129.2 (2002), pp. 517–526.

[490] Tamás Matusek et al. "The ESCRT machinery regulates the secretion and long-range activity of Hedgehog". In: *Nature* 516.729 (2014), pp. 99–103.

[491] Muthukumar Ramanathan et al. "Methods to study RNA-protein interactions". In: *Nature Methods* 16.3 (2019), pp. 225–234.

[492] Xuexiu Zheng et al. “Detecting RNA–protein interaction using end-labeled biotinylated RNA oligonucleotides and immunoblotting”. In: *Methods in Molecular Biology* 1421 (2016), pp. 35–44.

[493] Alfredo Castello et al. “System-wide identification of RNA-binding proteins by interactome capture”. In: *Nature Protocols* 8.3 (2013), pp. 491–500.

[494] Alfredo Castello et al. “Identification of RNA-binding domains of RNA-binding proteins in cultured cells on a system-wide scale with RBDmap”. In: *Nature Protocols* 12.12 (2017), pp. 2447–2464.

[495] Kathrin Theil, Koshi Imami, and Nikolaus Rajewsky. “Identification of proteins and miRNAs that specifically bind an mRNA in vivo”. In: *Nature Communications* 10 (2019).

[496] Vikram Agarwal et al. “Predicting microRNA targeting efficacy in Drosophila”. In: *Genome Biology* 19 (2018).

[497] Matthew J. Shurtleff, Morayma M. Temoche-Diaz, and Randy Schekman. “Extracellular Vesicles and Cancer: Caveat Lector”. In: *Annual Review of Cancer Biology* 2.1 (2018), pp. 395–411.

[498] Tamás Baranyai et al. “Isolation of exosomes from blood plasma: Qualitative and quantitative comparison of ultracentrifugation and size exclusion chromatography methods”. In: *PLoS ONE* 10.12 (2015).

[499] Ana Gámez-Valero et al. “Size-Exclusion Chromatography-based isolation minimally alters Extracellular Vesicles’ characteristics compared to precipitating agents”. In: *Scientific Reports* 6 (2016).

[500] D Ter-Ovanesyan. “Investigation of RNA in Extracellular Vesicles”. PhD thesis. 2019.

[501] X. Lin et al. “MicroRNAs and Unusual Small RNAs Discovered in Kaposi’s Sarcoma-Associated Herpesvirus Virions”. In: *Journal of Virology* 86.23 (2012), pp. 12717–12730.

[502] Bogdan Mateescu et al. “Obstacles and opportunities in the functional analysis of extracellular vesicle RNA – an ISEV position paper”. In: *Journal of Extracellular Vesicles* 6.1 (2001).

[503] Andrew F. Hill et al. “ISEV position paper: extracellular vesicle RNA analysis and bioinformatics”. In: *Journal of Extracellular Vesicles* 2.1 (2013), p. 22859.

[504] Cory J Evans et al. “G-TRACE: rapid Gal4-based cell lineage analysis in Drosophila”. In: *Nature Methods* 6.8 (2010), pp. 603–605.

[505] Arslan Akmammedov, Marco Geigges, and Renato Paro. “Single vector non-leaky gene expression system for Drosophila melanogaster”. In: *Scientific Reports* 7.1 (2017), pp. 1–12.

[506] K G Golic and S Lindquist. “The FLP recombinase of yeast catalyzes site-specific recombination in the Drosophila genome.” In: *Cell* 59.3 (1989), pp. 499–509.

[507] Uwe Hinz, Bernd Giebel, and JoséA A. Campos-Ortega. “The basic-helix-loop-helix domain of Drosophila lethal of scute protein is sufficient for proneural function and activates neurogenic genes”. In: *Cell* 76.1 (1994), pp. 77–87.

[508] R G Phillips et al. “The Drosophila segment polarity gene patched is involved in a position-signalling mechanism in imaginal discs.” In: *Development* 110.1 (1990), pp. 105–14.

[509] Justin A. Bosch, Taryn M. Sumabat, and Iswar K. Hariharan. “Persistence of RNAi-mediated knockdown in Drosophila complicates mosaic analysis yet enables highly sensitive lineage tracing”. In: *Genetics* 203.1 (2016), pp. 109–118.

[510] D. Heidmann and C. F. Lehner. “Reduction of Cre recombinase toxicity in proliferating Drosophila cells by estrogen-dependent activity regulation”. In: *Development Genes and Evolution* 211.8-9 (2001), pp. 458–465.

[511] Anne Holz et al. “The two origins of hemocytes in Drosophila”. In: *Development* 130.20 (2003), pp. 4955–4962.

[512] Daniel Bachurski et al. “Extracellular vesicle measurements with nanoparticle tracking analysis–An accuracy and repeatability comparison between NanoSight NS300 and ZetaView”. In: *Journal of Extracellular Vesicles* 8.1 (2019), p. 1596016.

[513] Romain Linares et al. “High-speed centrifugation induces aggregation of extracellular vesicles”. In: *Journal of Extracellular Vesicles* 4.1 (2015).

[514] Joel Z. Nordin et al. “Ultrafiltration with size-exclusion liquid chromatography for high yield isolation of extracellular vesicles preserving intact biophysical and functional properties”. In: *Nanomedicine: Nanotechnology, Biology, and Medicine* 11.4 (2015), pp. 879–883.

[515] Emma A. Mol et al. “Higher functionality of extracellular vesicles isolated using size-exclusion chromatography compared to ultracentrifugation”. In: *Nanomedicine: Nanotechnology, Biology, and Medicine* 13.6 (2017), pp. 2061–2065.

[516] Barbara W. Sódar et al. “Low-density lipoprotein mimics blood plasma-derived exosomes and microvesicles during isolation and detection”. In: *Scientific Reports* 6 (2016).

[517] Kaloyan Takov, Derek M. Yellon, and Sean M. Davidson. “Comparison of small extracellular vesicles isolated from plasma by ultracentrifugation or size-exclusion chromatography: yield, purity and functional potential”. In: *Journal of Extracellular Vesicles* 8.1 (2019).

[518] Muhammad Nawaz and Farah Fatima. “Extracellular vesicles, tunneling nanotubes, and cellular interplay: Synergies and missing links”. In: *Frontiers in Molecular Biosciences* 4 (2017).

[519] V Thayanithy et al. “Tumor-stromal cross-talk: Direct cell-to-cell transfer of oncogenic microRNAs via tunneling nanotubes”. In: *Bone* 23 (2008).

[520] Montserrat Climent et al. “TGFβ triggers miR-143/145 transfer from smooth muscle cells to endothelial cells, thereby modulating vessel stabilization”. In: *Circulation Research* 116.11 (2015), pp. 1753–1764.

[521] Gal Haimovich et al. “Intercellular mRNA trafficking via membrane nanotube-like extensions in mammalian cells”. In: *Proceedings of the National Academy of Sciences of the United States of America* 114.46 (2017), E9873–E9882.

[522] Fidel Nicolás Lolo, Sergio Casas-Tintó, and Eduardo Moreno. “Cell Competition Time Line: Winners Kill Losers, which Are Extruded and Engulfed by Hemocytes”. In: *Cell Reports* 2.3 (2012), pp. 526–539.

[523] B. Mollereau et al. “Compensatory proliferation and apoptosis-induced proliferation: A need for clarification”. In: *Cell Death and Differentiation* 20.1 (2013), p. 181.

[524] Eva Maria Krämer-Albers. “Exosomes deliver ROS for regeneration”. In: *Nature Cell Biology* 20.3 (2018), pp. 225–226.

[525] Arnau Hervera et al. “Reactive oxygen species regulate axonal regeneration through the release of exosomal NADPH oxidase 2 complexes into injured axons”. In: *Nature Cell Biology* 20.3 (2018), pp. 307–319.

[526] Salvador C. Herrera, Raquel Martín, and Ginés Morata. “Tissue Homeostasis in the Wing Disc of Drosophila melanogaster: Immediate Response to Massive Damage during Development”. In: *PLoS Genetics* 9.4 (2013).

[527] Kenneth M Cadigan. “Regulating morphogen gradients in the Drosophila wing”. In: *Seminars in Cell & Developmental Biology* 13.2 (2002), pp. 83–90.

[528] Luis Alberto Baena-López, Antonio Baonza, and Antonio García-Bellido. “The orientation of cell divisions determines the shape of Drosophila organs”. In: *Current Biology* 15.18 (2005), pp. 1640–1644.

[529] B Cohen, E Wimmer, and S M Cohen. “Early development of leg and wing primordia in the Drosophila embryo”. In: *Mechanisms of Development* 33.3 (1991), pp. 229–40.

[530] Peter A. Lawrence and Gines Morata. “The early development of mesothoracic compartments in Drosophila”. In: *Developmental Biology* 56.1 (1977), pp. 40–51.

[531] A Garcia-Bellido, P Ripolll, and G. Morata. “Developmental Compartmentalisation of the Wing Disk of Drosophila”. In: *Nature New Biology* 245 (1973), pp. 251–253.

[532] Natalie A. Dye et al. “Cell dynamics underlying oriented growth of the drosophila wing imaginal disc”. In: *Development* 144.23 (2017), pp. 4406–4421.

[533] Ulrike Nienhaus, Tinri Aegerter-Wilmsen, and Christof M. Aegerter. “In-Vivo Imaging of the Drosophila Wing Imaginal Disc over Time: Novel Insights on Growth and Boundary Formation”. In: *PLoS ONE* 7.10 (2012).

[534] A Garcia-Bellido. “Developmental Genetics of the Venation Pattern of Drosophila”. In: *Annual Review of Genetics* 26.1 (1992), pp. 275–304.

[535] Jaime Resino, Patricia Salama-Cohen, and Antonio Garcia-Bellido. “Determining the role of patterned cell proliferation in the shape and size of the Drosophila wing”. In: *Proceedings of the National Academy of Sciences of the United States of America* 99.11 (2002), pp. 7502–7507.

[536] Alexis V. Forterre et al. “Extracellular vesicle–mediated in vitro transcribed mRNA delivery for treatment of HER2+ breast cancer xenografts in mice by prodrug CB1954 without general toxicity”. In: *Molecular Cancer Therapeutics* 19.3 (2020), pp. 858–867.

[537] Chiara Gamberi, Oona Johnstone, and Paul Lasko. “Drosophila RNA binding proteins”. In: *International Review of Cytology* 248 (2006), pp. 43–139.

[538] Sen Lin Lai and Tzumin Lee. “Genetic mosaic with dual binary transcriptional systems in Drosophila”. In: *Nature Neuroscience* 9.5 (2006), pp. 703–709.

[539] Alberto del Valle Rodríguez, Dominic Didiano, and Claude Desplan. "Power tools for gene expression and clonal analysis in Drosophila". In: *Nature Methods* 9.1 (2012), pp. 47–55.

[540] A. Callejo et al. "Dispatched mediates Hedgehog basolateral release to form the long-range morphogenetic gradient in the Drosophila wing disk epithelium". In: *Proceedings of the National Academy of Sciences of the United States of America* 108.31 (2011), pp. 12591–12598.

[541] Cristina Cruz et al. "A gain-of-function screen identifying genes required for growth and pattern formation of the Drosophila melanogaster wing". In: *Genetics* 183.3 (2009), pp. 1005–1026.

[542] Eva Matoulkova et al. "The role of the 3-prime untranslated region in post-transcriptional regulation of protein expression in mammalian cells". In: *RNA Biology* 9.5 (2012), pp. 563–576.

[543] Carlo Maria Di Liegro, Gabriella Schiera, and Italia Di Liegro. "Extracellular vesicle-associated RNA as a carrier of epigenetic information". In: *Genes* 8.10 (2017).

[544] Ramesh Kakarla et al. "Apoptotic cell-derived exosomes: messages from dying cells". In: *Experimental and Molecular Medicine* 56 (2020), pp. 1–6.

[545] Luis Alberto Baena-Lopez. "All about the caspase-dependent functions without cell death". In: *Seminars in Cell and Developmental Biology* 82 (2018), pp. 77–78.

[546] Derek Cui Xu, Lewis Arthurton, and Luis Alberto Baena-lopez. "Learning on the Fly : The Interplay between Caspases and Cancer". In: *BioMed Research International* (2018).

[547] Alessia Galasso, Daria Iakovleva, and Luis Alberto Baena-Lopez. "Non-apoptotic caspase activity sustains proliferation and differentiation of ovarian somatic cells by modulating Hedgehog-signalling and autophagy". In: *bioRxiv* (2020).

[548] Caitlin E. Fogarty and Andreas Bergmann. "Killers creating new life: Caspases drive apoptosis-induced proliferation in tissue repair and disease". In: *Cell Death and Differentiation* 24.8 (2017), pp. 1390–1400.

[549] Alla Amcheslavsky et al. "Plasma Membrane Localization of Apoptotic Caspases for Non-apoptotic Functions". In: *Developmental Cell* 45.4 (2018), 450–464.e3.

[550] Kajal H. Gupta et al. "Apoptosis and Compensatory Proliferation Signaling Are Coupled by CrkI-Containing Microvesicles". In: *Developmental Cell* 41.6 (2017), 674–684.e5.

[551] Muthukumar Ramanathan et al. "RNA-protein interaction detection in living cells". In: *Nature Methods* 15.3 (2018), pp. 207–212.

[552] Muthukumar Ramanathan, Muthukumar Ramanathan, and Paul Khavari. "RNA-Protein Interaction Detection (RaPID)". In: *Protocol Exchange* (2018), pp. 1–8.

[553] Daniel D. Scott and Chris J. Norbury. "RNA decay via 3-prime uridylation". In: *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1829.6-7 (2013), pp. 654–665.

[554] Bin Tian and James L. Manley. "Alternative polyadenylation of mRNA precursors". In: *Nature Reviews Molecular Cell Biology* 18.1 (2016), pp. 18–30.

[555] Derek Cui Xu, Kenneth M Yamada, and Luis Alberto Baena-lopez. "Widespread non-apoptotic activation of Drosophila Caspase-2/9 limits JNK signaling, macrophage proliferation, and growth of wound-like tumors". In: *bioRxiv* (2020).

[556] Archita Ghoshal et al. "Extracellular vesicle-dependent effect of RNA-binding protein IGF2BP1 on melanoma metastasis". In: *Oncogene* 38.21 (2019), pp. 4182–4196.

[557] Keitaro Hagiwara et al. "Commitment of Annexin A2 in recruitment of microRNAs into extracellular vesicles". In: *FEBS Letters* 589.24 (2015), pp. 4071–4078.

[558] Alessandra Lavello et al. "Role of Alix in miRNA packaging during extracellular vesicle biogenesis". In: *International Journal of Molecular Medicine* 37.4 (2016), pp. 958–966.

[559] A. H. Brand and N. Perrimon. "Targeted gene expression as a means of altering cell fates and generating dominant phenotypes". In: *Development* 118.2 (1993), pp. 401–415.

[560] Jun Ma and Mark Ptashne. "The carboxy-terminal 30 amino acids of GAL4 are recognized by GAL80". In: *Cell* 50.1 (1987), pp. 137–142.

[561] Clotilde Théry et al. "Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines". In: *Journal of Extracellular Vesicles* 7.1 (2018).

[562] Roman Kornilov et al. "Efficient ultrafiltration-based protocol to deplete extracellular vesicles from fetal bovine serum". In: *Journal of Extracellular Vesicles* 7.1 (2018).

[563] Lida A. Beninson and Monika Fleshner. "Exosomes in fetal bovine serum dampen primary macrophage IL-1β response to lipopolysaccharide (LPS) challenge". In: *Immunology Letters* 163.2 (2015), pp. 187–192.

[564] Erez Eitan et al. "Extracellular vesicle-depleted fetal bovine and human sera have reduced capacity to support cell growth". In: *Journal of Extracellular Vesicles* 4.2015 (2015).

[565] Ayuko Hoshino et al. "Extracellular Vesicle and Particle Biomarkers Define Multiple Human Cancers". In: *Cell* 182.4 (2020), 1044–1061.e18.

[566] L. H. Wang, K. G. Rothberg, and R. G.W. Anderson. "Mis-assembly of clathrin lattices on endosomes reveals a regulatory switch for coated pit formation". In: *Journal of Cell Biology* 123.5 (1993), pp. 1107–1117.

[567] Shawn M. Ferguson and Pietro De Camilli. "Dynamin, a membrane remodelling GTPase". In: *Nature Reviews Molecular Cell Biology* 13.2 (2012), pp. 75–88.

[568] J. S. Herskovits et al. "Effects of mutant rat dynamin on endocytosis". In: *Journal of Cell Biology* 122.3 (1993), pp. 565–578.

[569] Céline Barrès et al. "Galectin-5 is bound onto the surface of rat reticulocyte exosomes and modulates vesicle uptake by macrophages". In: *Blood* 115.3 (2010), pp. 696–705.

[570] Ivan R. Nabi and Phuong U. Le. "Caveolae/raft-dependent endocytosis". In: *Journal of Cell Biology* 161.4 (2003), pp. 673–677.

[571] Richard G.W. Anderson. "The caveolae membrane system". In: *Annual Review of Biochemistry* 67 (1998), pp. 199–225.

[572] Teymuras V. Kurzchalia and Robert G. Parton. "Membrane microdomains and caveolae". In: *Current Opinion in Cell Biology* 11.4 (1999), pp. 424–431.

[573] Frank W. Pfrieger and Nicolas Vitale. "Thematic review series: Exosomes and microvesicles: Lipids as Key Components of their Biogenesis and Functions Cholesterol and the journey of extracellular vesicles". In: *Journal of Lipid Research* 59.12 (2018), pp. 2255–2261.

[574] A. Nanbo et al. "Exosomes Derived from Epstein-Barr Virus-Infected Cells Are Internalized via Caveola-Dependent Endocytosis and Promote Phenotypic Modulation in Target Cells". In: *Journal of Virology* 87.18 (2013), pp. 10334–10347.

[575] S. Rudt and R. H. Müller. "In vitro phagocytosis assay of nano- and microparticles by chemiluminescence. III. Uptake of differently sized surface-modified particles, and its correlation to particle properties and in vivo distribution". In: *European Journal of Pharmaceutical Sciences* 1.1 (1993), pp. 31–39.

[576] Len Stephens, Chris Ellson, and Phillip Hawkins. "Roles of PI3Ks in leukocyte chemotaxis and phagocytosis". In: *Current Opinion in Cell Biology* 14.2 (2002), pp. 203–213.

[577] Elizabeth S. Gold et al. "Amphiphysin IIm, a novel amphiphysin II isoform, is required for macrophage phagocytosis". In: *Immunity* 12.3 (2000), pp. 285–292.

[578] Joel A. Swanson. "Shaping cups into phagosomes and macropinosomes". In: *Nature Reviews Molecular Cell Biology* 9.8 (2008), pp. 639–649.

[579] Madhan Masilamani et al. "Uncommon endocytic and trafficking pathway of the natural killer cell CD94/NKG2A inhibitory receptor". In: *Traffic* 9.6 (2008), pp. 1019–1034.

[580] Mamoun Ahram et al. "Rac1-induced endocytosis is associated with intracellular proteolysis during migration through a three-dimensional matrix". In: *Experimental Cell Research* 260.2 (2000), pp. 292–303.

[581] Stine Grimmer, Bo van Deurs, and Kirsten Sandvig. "Membrane ruffling and macropinocytosis in A431 cells require cholesterol". In: *Journal of Cell Science* 115.14 (2002), pp. 2953–2962.

[582] M. Amyere et al. "Constitutive macropinocytosis in oncogene-transformed fibroblasts depends on sequential permanent activation of phosphoinositide 3-kinase and phospholipase C". In: *Molecular Biology of the Cell* 11.10 (2000), pp. 3453–3467.

[583] Markus C. Kerr and Rohan D. Teasdale. "Defining macropinocytosis". In: *Traffic* 10.4 (2009), pp. 364–371.

[584] Manfred Frick et al. "Coassembly of Flotillins Induces Formation of Membrane Microdomains, Membrane Curvature, and Vesicle Budding". In: *Current Biology* 17.13 (2007), pp. 1151–1156.

[585] Grant P. Otto and Benjamin J. Nichols. "The roles of flotillin microdomains-endocytosis and beyond". In: *Journal of Cell Science* 124.23 (2011), pp. 3933–3940.

[586] Oleg O. Glebov, Nicholas A. Bright, and Benjamin J. Nichols. "Flotillin-1 defines a clathrin-independent endocytic pathway in mammalian cells". In: *Nature Cell Biology* 8.1 (2006), pp. 46–54.

[587] Sherif E. Emam et al. "Liposome co-incubation with cancer cells secreted exosomes (extracellular vesicles) with different proteins expressions and different uptake pathways". In: *Scientific Reports* 8.1 (2018), pp. 1–11.

[588] Elodie Segura et al. "ICAM-1 on exosomes from mature dendritic cells is critical for efficient naive T-cell priming". In: *Blood* 106.1 (2005), pp. 216–223.

[589] L Zitvogel et al. "Eradication of established murine tumors using a novel cell-free vaccine: dendritic cell-derived exosomes." In: *Nature Medicine* 4.5 (1998), pp. 594–600.

[590] Clotilde Théry et al. "Indirect activation of naïve CD4+ T cells by dendritic cell-derived exosomes". In: *Nature Immunology* 3.12 (2002), pp. 1156–1162.

[591] Chiara Cossetti et al. "Extracellular vesicles from neural stem cells transfer IFN-γ via Ifngr1 to activate Stat1 signaling in target cells". In: *Molecular Cell* 56.2 (2014), pp. 193–204.

[592] Donald G. Phinney et al. "Mesenchymal stem cells use extracellular vesicles to outsource mitophagy and shuttle microRNAs". In: *Nature Communications* 6 (2015), pp. 1–15.

[593] Muller Fabbri et al. "MicroRNAs bind to Toll-like receptors to induce prometastatic inflammatory response". In: *Proceedings of the National Academy of Sciences of the United States of America* 109.31 (2012).

[594] Quan Lu et al. "TSG101 interaction with HRS mediates endosomal trafficking and receptor down-regulation". In: *Proceedings of the National Academy of Sciences of the United States of America* 100.13 (2003), pp. 7626–7631.

[595] David J. Katzmann et al. "Vps27 recruits ESCRT machinery to endosomes during MVB sorting". In: *Journal of Cell Biology* 162.3 (2003), pp. 413–423.

[596] M. Razi and C. E. Futter. "Distinct Roles for Tsg101 and Hrs in Multivesicular Body Formation and Inward Vesiculation". In: *Molecular Biology of the Cell* 17.August (2006).

[597] John McCullough et al. "Activation of the endosome-associated ubiquitin isopeptidase AMSH by STAM, a component of the multivesicular body-sorting machinery". In: *Current Biology* 16.2 (2006), pp. 160–165.

[598] Kay Hofmann and Laurent Falquet. "A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems". In: *Trends in Biochemical Sciences* 26.6 (2001), pp. 347–350.

[599] Camilla Raiborg et al. "FYVE and coiled-coil domains determine the specific localisation of Hrs to early endosomes". In: *Journal of Cell Science* 114.12 (2001), pp. 2255–2263.

[600] Christopher G. Burd and Scott D. Emr. "Phosphatidylinositol(3)-phosphate signaling mediated by specific binding to RING FYVE domains". In: *Molecular Cell* 2.1 (1998), pp. 157–162.

[601] Jean Michel Gaullier et al. "FYVE fingers bind PtdIns(3)P". In: *Nature* 394.6692 (1998), pp. 432–433.

[602] Ka Fai Leung, Joel B. Dacks, and Mark C. Field. "Evolution of the multivesicular body ESCRT machinery; retention across the eukaryotic lineage". In: *Traffic* 9.10 (2008), pp. 1698–1716.

[603] Eiji Morita et al. "Identification of Human MVB12 Proteins as ESCRT-I Subunits that Function in HIV Budding". In: *Cell Host and Microbe* 2.1 (2007), pp. 41–53.

[604] Tony Chu et al. "New component of ESCRT-I regulates endosomal sorting complex assembly". In: *Journal of Cell Biology* 175.5 (2006), pp. 815–823.

[605] Michael S. Kostelansky et al. "Molecular Architecture and Functional Model of the Complete Yeast ESCRT-I Heterotetramer". In: *Cell* 129.3 (2007), pp. 485–498.

[606] Aitor Hierro et al. "Structure of the ESCRT-II endosomal trafficking complex". In: *Nature* 431.7005 (2004), pp. 221–225.

[607] Young Jun Im and James H. Hurley. "Integrated Structural Model and Membrane Targeting Mechanism of the Human ESCRT-II Complex". In: *Developmental Cell* 14.6 (2008), pp. 902–913.

[608] Thomas Slagsvold et al. "Eap45 in mammalian ESCRT-II binds ubiquitin via a phosphoinositide- interacting GLUE domain". In: *Journal of Biological Chemistry* 280.20 (2005), pp. 19600–19606.

[609] Steven L. Alam et al. "Structural basis for ubiquitin recognition by the human ESCRT-II EAP45 GLUE domain". In: *Nature Structural and Molecular Biology* 13.11 (2006), pp. 1029–1030.

[610] Satoshi Hirano et al. "Structural basis of ubiquitin recognition by mammalian Eap45 GLUE domain". In: *Nature Structural and Molecular Biology* 13.11 (2006), pp. 1031–1032.

[611] Hsiangling Teo et al. "ESCRT-I Core and ESCRT-II GLUE Domain Structures Reveal Role for GLUE in Linking to ESCRT-I and Membranes". In: *Cell* 125.1 (2006), pp. 99–111.

[612] Phyllis I. Hanson et al. "Plasma membrane deformation by circular arrays of ESCRT-III protein filaments". In: *Journal of Cell Biology* 180.2 (2008), pp. 389–402.

[613] Véronique Pons et al. "Hrs and SNX3 functions in sorting and membrane invagination within multivesicular bodies". In: *PLoS Biology* 6.9 (2008), pp. 1942–1956.

[614] Camilla Raiborg and Harald Stenmark. "The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins". In: *Nature* 458.7237 (2009), pp. 445–452.

[615] Mian He et al. "Hepatocellular carcinoma-derived exosomes promote motility of immortalized hepatocyte through transfer of oncogenic proteins and RNAs". In: *Carcinogenesis* 36.9 (2015), pp. 1008–1018.

[616] Rong Xu et al. "Highly-purified exosomes and shed microvesicles isolated from the human colon cancer cell line LIM1863 by sequential centrifugal ultrafiltration are biochemically and functionally distinct". In: *Methods* 87 (2015), pp. 11–25.

[617] Ikrame Lazar et al. "Proteome characterization of melanoma exosomes reveals a specific signature for metastatic cell lines". In: *Pigment Cell and Melanoma Research* 28.4 (2015), pp. 464–475.

[618] Pedram Kharaziha et al. “Molecular profiling of prostate cancer derived exosomes may reveal a predictive signature for response to docetaxel”. In: *Oncotarget* 6.25 (2015), pp. 21740–21754.

[619] Jung Eun Park et al. “Hypoxic tumor cell modulates its microenvironment to enhance angiogenic and metastatic potential by secretion of proteins and exosomes”. In: *Molecular and Cellular Proteomics* 9.6 (2010), pp. 1085–1099.

[620] Gabriel Skogberg et al. “Characterization of Human Thymic Exosomes”. In: *PLoS ONE* 8.7 (2013).

[621] Kasra Hassani and Martin Olivier. “Immunomodulatory Impact of Leishmania-Induced Macrophage Exosomes: A Comparative Proteomic and Functional Analysis”. In: *PLoS Neglected Tropical Diseases* 7.5 (2013).

[622] Sonja I. Buschow et al. “MHC class II-associated proteins in B-cell exosomes and potential functional implications for exosome biogenesis”. In: *Immunology and Cell Biology* 88.8 (2010), pp. 851–856.

[623] Bas W.M. Van Balkom et al. “Quantitative and qualitative analysis of small RNAs in human endothelial cells and exosomes provides insights into localized RNA processing, degradation and sorting”. In: *Journal of Extracellular Vesicles* 4.2015 (2015).

[624] Yuk Kit Chan et al. “Proteomic analysis of exosomes from nasopharyngeal carcinoma cell identifies intercellular transfer of angiogenic proteins”. In: *International Journal of Cancer* 137.8 (2015), pp. 1830–1841.

[625] Xue Zhao et al. “Quantitative proteomic analysis of exosome protein content changes induced by hepatitis B virus in Huh-7 cells using SILAC labeling and LC-MS/MS”. In: *Journal of Proteome Research* 13.12 (2014), pp. 5391–5402.

[626] Sun Lee Hyo, Jaeho Jeong, and Kong Joo Lee. “Characterization of vesicles secreted from insulinoma NIT-1 cells”. In: *Journal of Proteome Research* 8.6 (2009), pp. 2851–2862.

[627] Simon Staubach, Hanieh Razawi, and Franz Georg Hanisch. “Proteomics of MUC1-containing lipid rafts from plasma membranes and exosomes of human breast carcinoma cells MCF-7”. In: *Proteomics* 9.10 (2009), pp. 2820–2835.

[628] Dong Sic Choi et al. “Proteomic analysis of microvesicles derived from human colorectal cancer cells”. In: *Journal of Proteome Research* 6.12 (2007), pp. 4646–4655.

[629] Valbona Luga et al. “Exosomes mediate stromal mobilization of autocrine Wnt-PCP signaling in breast cancer cell migration”. In: *Cell* 151.7 (2012), pp. 1542–1556.

[630] Annika Pienimaeki-Roemer et al. “Lipidomic and proteomic characterization of platelet extracellular vesicle subfractions from senescent platelets”. In: *Transfusion* 55.3 (2015), pp. 507–521.

[631] Patricia A. Gonzales et al. “Large-scale proteomics and phosphoproteomics of urinary exosomes”. In: *Journal of the American Society of Nephrology* 20.2 (2009), pp. 363–379.

[632] S. Mathivanan et al. “Proteomics Analysis of A33 Immunoaffinity-purified Exosomes Released from the Human Colon Tumor Cell Line LIM1215 Reveals a Tissue-specific Protein Signature”. In: *Molecular & Cellular Proteomics* 9.2 (2010), pp. 197–208.