THE UNIVERSITY
of EDINBURGH

Molecular and functional characterisation of an extracellular Argonaute protein secreted by a gastrointestinal nematode

Kyriaki Neophytou

Thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2022

Declaration

I, the undersigned, declare that this thesis has been composed solely by myself. The work presented has been conducted by myself, unless otherwise acknowledged or stated throughout the thesis. Data from Chapter 3 have previously been published in Chow *et al*, (2019) for which I am a co-author. Some of the data in Chapter 6 are included in the MSc thesis of Ms Chanel Naar, who worked under my supervision on specific elements (which are stated in the thesis). Some data in Chapter 6 have been submitted as part of a patent (Patent publication no.: WO/2021/038250). This work has not been submitted for any other degree or professional qualification.

Kyriaki Neophytou

Date: 05/02/2022

Contributions

Dr Cei Abreu-Goodger analysed the small RNA-seq data and generated some of the figures presented in Chapter 3 (through discussion with the author). The small RNA-seq was performed by the Wellcome Trust Clinical Research Facility in Edinburgh.

The CLASH protocol in cells was optimised by Dr Katrina Gordon and Dr Sarah Ressel. Bioinformatic analysis of the non-chimeric data generated by the CLASH experiment (Chapter 4) was done by Dr Sujai Kumar. The sRNA guides generated by the aforementioned experiment were analysed by Dr Jose Roberto Bermudez-Barrientos and involved the input of Dr Cei Abreu-Goodger. The libraries were sequenced by Edinburgh Genomics, at the University of Edinburgh. Some of the figures presented in Chapter 4 were generated by Dr Cei Abreu-Goodger.

The recombinant MODE-K cell line expressing FLAG-tagged exWAGO was generated by Dr Franklin Chow and Mr Jonathan Wild.

LC-MS/MS from IPs from infected gut and the recombinant exWAGO-MODE-K cell line presented in Chapter 5 was performed by Dr Dominic Thekkedath Kurian at the Proteomics and Metabolomics facility of the Roslin Institute (Edinburgh) who also provided the details for the corresponding methods text. LC-MS/MS from IPs from worm lysate and infected gut (Chapter 5) was carried out by the EdinOmics facility (Ms Lisa Imrie) at The University of Edinburgh, who also curated the data and provided a method description. Immunofluorescence detection of exWAGO in MODE-K cells was done in collaboration with Dr Tom Fenton.

Successful completion of the vaccination experiments discussed in Chapter 6 required the involvement of several people including: the author, Ms Elaine Robertson, Ms Yvonne Harcus, staff from the Ashworth animal unit, Ms Ruby White, Ms Chanel Naar, Ms Rowan Bancroft, Mr Sam Hillman, Ms Nicola Logan and Mr Xiaochen Du. The author, ER, YH, RW, RB and SH helped with enumerating worm burdens. The author, ER, YH and XD helped with egg counts. The author, ER, YH and staff from animal unit were involved in vaccination of the animals. ER, YH and NL harvested the mice. ELISAs for the analysis of the immune responses of the vaccinated mice in Chapter 6 were performed under the author's guidance by ER, CN and the author. Western blot analysis examining the IgG response of the vaccinated mice were performed by YH with guidance from the author.

Acknowledgements

This journey would have not been possible if it wasn't for my supervisor, my colleagues, my family, my partner, and of course funding.

Amy, thank you for all the advice, guidance, support, and opportunities you offered me during the PhD. From travelling to Boulder and working with the Parker Lab, to attending conferences and retreats, to spending time with the girls. Your enthusiasm, determination, and passion are truly infectious! It has been a pleasure working with you these past years!

To my collaborator Cei – Thank you for all your help and advice on all the bioinformatic matters. Along with the bioinformatic expertise of Sujai and Beto, this project would have not been the same. Thank you all for your patience, for answering all my silly questions and requests, and the numerous discussions you happily had with me.

To my thesis committee, Liz and Keith, you have always been there whenever I asked to speak to you and for that I thank you for your support these 3 or so years.

This project would have not flourished to the same extend if it wasn't for the groundwork done by Franklin. Thank you for teaching me all things exWAGO and equipping me with adequate knowledge and skills to take forward this project.

Kat, your immense support during my rotation and at the beginning of the PhD set me good for the rest of this journey. Thank you for encouraging me and being there for me whatever the matter was. Along with Sarah, Ruby, Elaine, Yvonne, Patricia, Jeroen & Alíz, your friendship and words of wisdom have resonated with me, and I truly appreciate your physical and mental support during this journey. Along with many other friends & colleagues from the Buck, Macias and Zamoyska groups, thank you all for sharing reagents, teaching me new techniques, helping me design/troubleshoot experiments, and bearing with all Latin telenovela things related!

The vaccination experiments would have not been executed without the expertise of Elaine, Yvonne, and many other people. Thank you for helping me organise these and for your perseverance throughout those long harvest days. Elaine and Chanel, thank you for being my ELISA queens. Yvonne and David W., thank you for your hard work on the monoclonal. Here, I would also like to thank Al Nisbet, Rick Maizels, Matt Taylor, the Pedersen group, and others for your physical &/or intellectual contributions to the vaccination experiments.

A big thank you to the HPGH programme & administrators, and the Wellcome Trust for funding my PhD. The programme allowed me to meet and work with lots of amazing scientists. To my first-year host labs – Pedro, Vincent, Katy, Liam, Finn, Pete and Stephen – thank you all for helping me become a better scientist and for being part of this journey.

To the university counsellor Rhona H. – Thank you for listening to me and equipping me with the tools to understanding my feelings and managing anxiety and stress. You inspired me to persevere and get through the hard times.

Στους γονείς μου και στο μπεϊπάκι μου – Τα λόγια αυτά είναι πολύ φτωχά για να σας πουν ευχαριστώ και να εκφράσουν το πόσο εκτιμώ που είστε πάντα δίπλα μου και που πάντοτε με υποστηρίζατε ενεργά να επιδιώξω αυτό που ήθελα. Τα δάκρυα λύπης, θυμού και άγχους ας γίνουν επιτέλους δάκρυα χαράς! Σας αγαπώ πολύ! Ελπίζω να σας κάνω πάντα περήφανους.

To my fiancé & soon-to-be husband – Words will never be enough to thank you for your advice, support, and encouragement. Thank you for being there to pop the champagne in my highs and to catch me in my lows. You are my always, forever.

Δόξα Σοι, Κύριε!

Lay Summary

Gastrointestinal worms infect one in four people, particularly in countries where sanitation, hygiene and access to clean water are poor. One reason that gastrointestinal worms are so successful at infecting and surviving in their hosts is because they make molecules that change host cells, so that host cells do not expel the parasite. The parasite molecules that change cells can include ribonucleic acids (RNAs). There are various types of RNAs. Messenger RNAs (mRNAs) are sequences carrying instructions that tell the cell to make specific proteins. Smaller sequences of RNA can also be made that do not produce proteins, but instead regulate how much protein gets produced. To do this, the small RNAs work together with a protein that belongs to the Argonaute protein family. The small RNA directs the Argonaute protein to the mRNA and this interaction causes the mRNA to be destroyed, shutting down gene expression and protein production. In this way, small RNAs are important in regulating gene expression and this controls how cells function and respond to infections.

Our group studies a parasitic worm called *Heligmosomoides bakeri* (*H. bakeri*) which infects mice and lives in their guts. This model is used to study how the parasites are able to establish infections and to test new ways to treat or control infection that could be relevant to human and animal gastrointestinal worm infections. As with other parasitic worms, *H. bakeri* releases proteins and RNA molecules, some of which are exported in packages known as vesicles. *H. bakeri* vesicles can be transferred to host cells and contain small RNAs and an Argonaute protein, which we named exWAGO. As exWAGO is an Argonaute protein and since Argonaute proteins are involved in controlling gene expression, we predict that exWAGO and the small RNAs released by the parasite can control the expression of genes in the host. The small RNAs and exWAGO are found both inside and outside the vesicles. A goal of this research is to understand if and how the two forms of exWAGO existing inside and outside vesicles are functional in regulating host gene expression. In particular, key questions addressed here are whether both forms of exWAGO bind the same set of small RNAs and finding which proteins exWAGO interacts with that could enable its predicted function in gene regulation. Additionally, we test whether exWAGO could be used as a vaccine candidate against infections caused by other exWAGO-producing parasitic worms.

Our results show that both forms of exWAGO interact with a particular type of small RNAs, called secondary small interfering RNAs (siRNAs). Direct comparison of the secondary siRNAs interacting with the two forms of exWAGO shows that while the two exWAGO forms interact with the same secondary siRNAs, they each also interact with unique secondary siRNAs. This would suggest that the two forms of exWAGO might be directed to different mRNAs and therefore control different host genes.

Using a biochemical technique that allows us to capture exWAGO, the small RNA, and the mRNA from infected mouse gut tissue, we aimed to identify the mouse mRNAs that are potentially regulated by exWAGO. The mRNA sequences detected have unusual characteristics that are reminiscent of the characteristics of parasite small RNAs. For this reason, further experiments are required to test if we have detected true mouse mRNAs. Research is currently underway to understand if and how exWAGO may regulate the expression of mouse mRNAs.

One way that Argonaute proteins can regulate gene expression is by slicing/chopping the mRNA, which results in its degradation. We therefore tested the ability of exWAGO to slice. Our data show that exWAGO cannot slice mRNA and this suggests that exWAGO might employ other mechanisms to regulate gene expression. To understand if this hypothesis stands, we set out to identify the proteins it interacts with inside host cells. The results led to identification of some potential protein interactors. Some of these have been reported to interact with other Argonaute proteins, whereas others have not been previously identified to interact with Argonaute proteins. From these data we can now hypothesise that exWAGO might be localised to specific compartments in host cells to perform its function. We also discovered a candidate protein that might be involved in the internalisation of exWAGO inside host cells. Further experiments are required to validate the protein interactions identified.

Finally, we tested if exWAGO can serve as a vaccine candidate against infection with *H. bakeri*. Vaccination of mice with exWAGO generated antibody responses against exWAGO and led to a significant clearance of worms from the mouse gut. These data suggest that vaccination with exWAGO reduces the ability of the parasite to survive. As exWAGO-like proteins are also generated by other gastrointestinal worms, including the human hookworm and sheep parasites, we propose that exWAGO could be a vaccine candidate for these parasitic worms.

In summary, this thesis presents a first insight into the mechanism of function of exWAGO in terms of what small RNAs and what proteins it interacts with. We also explored the properties and characteristics of the two forms of exWAGO, inside and outside vesicles, for the first time. By understanding if and how exWAGO may facilitate communication between parasite small RNAs and host mRNAs, we expand our understanding of how parasitic worms manipulate their hosts. This thesis also offers a possible intervention for gastrointestinal worm infections via exWAGO vaccination.

Scientific Abstract

Parasites manipulate their hosts to promote infection by secreting bioactive molecules including proteins, lipids and RNAs. Helminths (parasitic worms) secrete a plethora of such molecules possessing immunogenic and immunomodulatory properties and some of these are co-packaged within extracellular vesicles (EVs) and can be directly transferred to host cells. We have found that the mouse-infective gastrointestinal nematode *Heligmosomoides bakeri*, a close relative of the human hookworm *Necator americanus* and the animal-infective nematodes *Teladorsagia circumcincta* and *Haemonchus contortus*, secretes an RNA-binding protein which belongs to the family of Argonaute (AGO) proteins. AGO proteins are at the heart of RNA interference, a mechanism involved in gene regulation. By associating with small RNA guides, AGO proteins are directed to messenger RNA targets. This interaction usually leads to gene silencing. The AGO protein secreted by *H. bakeri*, termed exWAGO, is found in two forms: a vesicular form (detected inside EVs) and a non-vesicular form that does not co-purify with EVs. One goal in this thesis is to understand whether and how these two forms of exWAGO are different in terms of their sRNA guides and potential targets. It is known that *H. bakeri* EVs are internalised by mouse host cells, such that the parasite-derived cargo could directly interact with and interfere with host gene expression. The EVs predominantly contain 5'PPP secondary short interfering RNAs (siRNAs) and EV uptake by mouse cells was shown to result in suppression of host genes involved in immunity and inflammation. As AGO proteins coordinate gene silencing mechanisms, we hypothesise that exWAGO is directly involved in mediating changes in host gene targets. The goal in this thesis is to build understanding on the putative role of exWAGO in mediating cross-species gene silencing by identifying the RNA and protein interaction partners of exWAGO. A further goal is to determine the importance of exWAGO in parasite survival by blocking it through vaccination and testing the consequence to subsequent infections.

To identify the small RNAs (sRNAs) that associate with exWAGO we immunopurified exWAGO and generated sRNA libraries from three different sample types: adult worms (to detect the intra-parasite exWAGO), exWAGO from the excretory-secretory products that is found in EVs, and exWAGO from the excretory-secretory products that does not co-purify with EVs. sRNA sequencing analyses show that the intra-parasite exWAGO and the two extracellular forms of exWAGO bind 22-23G 5'PPP secondary siRNAs originating from transposable elements and novel repeats,

but are depleted from microRNAs (miRNAs), Y-RNAs, transfer RNAs and ribosomal RNAs. This suggests that exWAGO is bound specifically to secondary siRNAs, and we hypothesise that exWAGO is required for the export of these sequences from the parasite. To study the binding selectivity of exWAGO, we used gel shift assays and found that exWAGO binds 5'PPP guide RNAs with high affinity in contrast to the mammalian mouse AGO2 (mAGO2). Direct comparison of the secondary siRNA sequences bound by the vesicular and non-vesicular exWAGO indicates that the two exWAGO forms have some overlap in which secondary siRNAs they bind, however there are differences in the relative abundance of different siRNAs and there are some siRNAs only found in one exWAGO form. This suggests that the vesicular and non-vesicular forms of exWAGO might target different genes and have different functional properties.

Identification of the host transcripts targeted by exWAGO is crucial in understanding the role that exWAGO might play in cross-species gene silencing. To examine this, we developed a method to immunopurify the exWAGO protein *in vivo* from the gut of *H. bakeri*-infected mice and sequenced the RNAs with which it associates. Our initial dataset using the mAGO2 as a positive control suggests that we can successfully detect gene targets with this method. The preliminary results indicate that RNAs associated with exWAGO map to intronic regions in the mouse transcriptome, in contrast to mAGO2 where reads map to 3' untranslated regions and coding regions (consistent with location of mAGO2 target sites). Further experiments are required to test if the putative targets of exWAGO identified are true targets. Development of a bioinformatics pipeline to examine this complex dataset would also allow us to test if the method enabled formation of guide-target chimeric reads, hence permitting direct identification of guide sRNA-host target interactions.

To understand the mechanism by which exWAGO might mediate gene silencing in host cells, we tested whether exWAGO possesses the ability to cleave "slice" targets *in vitro*. Our results show that exWAGO does not possess slicer activity, consistent with its lack of a catalytic motif required by other AGOs for slicer activity. To explore what other gene regulation mechanism(s) might be employed by exWAGO inside host cells we set out to identify the protein interactors of exWAGO *in vitro* and *in vivo* using liquid chromatography-tandem mass spectrometry. The proteomic analysis identified some putative interactors that have been reported to interact with other known mammalian AGO proteins as well as some proteins with no previous literature linking

these to gene silencing. From these data we can now formulate hypotheses that exWAGO might be localised to specific compartments in host cells to perform its function. We also discovered a putative candidate protein that might be involved in the internalisation of exWAGO inside host cells and in host AGO trafficking. Further experiments are required to validate the protein interactions identified.

Finally, to understand the importance of exWAGO in infection and test if exWAGO can serve as a vaccine candidate, we immunised mice with recombinant exWAGO protein. Vaccination with recombinant exWAGO protein led to a strong induction of IgG1 antibodies against the protein and resulted in decreased egg and worm burdens (59.0% and 66.7% respectively, calculated as average across three experiments). These data suggest that blocking exWAGO *in vivo* reduces the ability of the parasite to survive. As exWAGO is highly conserved amongst Clade V gastrointestinal worms, including the human hookworm and sheep parasites, we propose that exWAGO could be a vaccine candidate for these nematodes.

In summary, this thesis presents a first insight into the mechanistic basis of exWAGO in terms of what sRNA guides it binds, what its putative target host genes are, and what proteins it interacts with. We also explore the properties and characteristics of the two extracellular forms of exWAGO for the first time. By understanding if and how exWAGO may facilitate communication between parasite sRNAs and host targets, we expand our understanding of how helminths manipulate their hosts. This thesis offers a putative intervention for gastrointestinal worm infections via exWAGO vaccination, but further studies could also lead to development of other therapeutic strategies or molecular tools based on exWAGO.

Table of Contents

List of Abbreviations

AGO – Argonaute protein

CDS – coding sequence

CLASH – Cross-linking, ligation and sequencing of hybrids

COMBAR – combatting anthelmintic resistance in ruminants

CPM – Counts Per Million

d – day

Dab2 – Disabled homolog 2 protein

ELISA – Enzyme-Linked Immunosorbent Assay

EMSA – Electrophoretic Mobility Shift Assay

ES – Excretory/Secretory products

EVs – Extracellular Vesicles

exRNA – extracellular RNA

exWAGO – extracellular Worm-specific Argonaute protein excreted/secreted by *Heligmosomoides bakeri*

FDR – False Discovery Rate

h – hour

hAGO2 – human AGO2 protein

HES – *Heligmosomoides bakeri* Excretory/Secretory products

Ig – Immunoglobulin

IP – immunoprecipitation

i.p. injection – Intraperitoneal injection

LC-MS/MS – Liquid Chromatography tandem Mass Spectrometry

LiHRA – Livestock Helminth Research Alliance

mAGO2 – mouse AGO2 protein

miRNA – microRNA

MODE-K cells – Mouse intestinal epithelial cells

mRNA – messenger RNA

nt – nucleotide(s)

p.c. – post-challenge

polyP – Polyphosphatase enzymatic treatment

RdRPs – RNA-dependent RNA polymerases

RISC – RNA Induced Silencing Complex

RITS – RNA Initiation of Transcriptional Silencing

RNAi – RNA interference

rRNA – ribosomal RNA

RT-qPCR – Reverse Transcription-quantitative real-time Polymerase Chain Reaction

S.E.M. – Standard Error of the Mean

SFPQ – Splicing Factor Proline and Glutamine Rich protein

siRNA – small interfering RNA

tRNA – transfer RNA

Ubap2l – Ubiquitin Associated Protein 2 Like protein

UTR – untranslated region

WAGO – Worm-Specific AGO protein

WB – Western Blot

WHO – World Health Organisation

Chapter 1: Introduction

1.1 Parasitic helminths

Worms are invertebrate animals which have a cylindrical or flattened body shape and typically have no limbs or eyes. They range in size from microscopic to greater than a meter in length (Wakelin, 1996). Some types of worms are free-living whereas others are parasitic and can infect humans, animals, and plants. Parasitic worms, usually referred to as helminths, are among the most common parasites in the world. Although helminths are not considered to cause high mortality rates, they are associated with severe morbidity and can establish chronic infections (Hotez *et al*, 2006).

Helminths are divided into four classes: (1) the nematodes/roundworms, (2) the platyhelminths/flatworms which are subdivided into cestodes/tapeworms and trematodes/flukes, (3) nematomorpha/hairworms, and (4) acanthocephala/spiny-headed worms (Fig. 1.1) (Mckay *et al*, 2017). Depending on the species of parasitic worm, the mode of transmission can vary. For example, infection with *Schistosoma* species and hookworms like *Necator americanus* involves skin penetration by larvae while ingestion of parasitic worm stages via undercooked meat from an infected animal permits transmission of tapeworms (*Taenia* species) (Wakelin, 1996). Another common route of transmission is via ingestion of food or water contaminated with worm eggs which will develop into larvae and adult worms inside the host, as exhibited by the parasitic worms *Ascaris lumbricoides* and *Trichuris trichuria* (Wakelin, 1996).

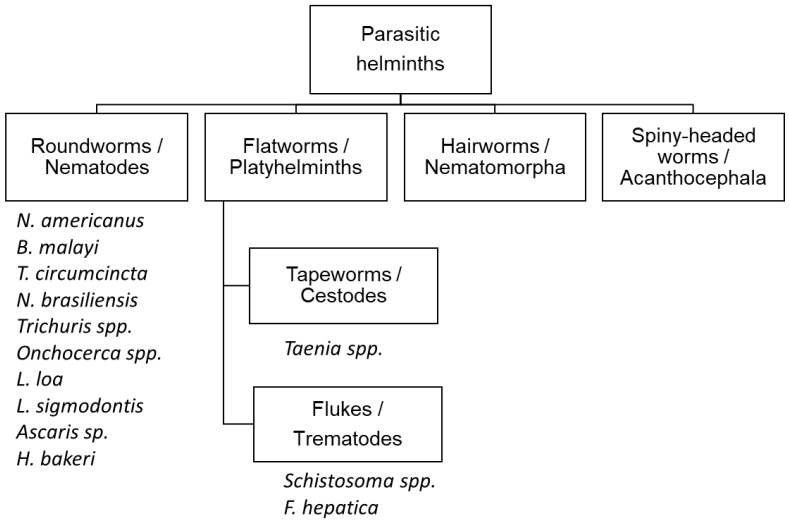


Figure 1.1 | Simplified phylogeny of parasitic helminths. Some of the specific nematode and platyhelminth species that are mentioned in this thesis are noted. Figure adapted from Mckay *et al,* 2017.

1.1.1 Gastrointestinal nematodes

Most mammalian-infective helminth species reside in the host gastrointestinal tract. Intestinal worm infections, some of which are caused by soil-transmitted helminths, are caused by nematodes and platyhelminths (flukes and tapeworms). Among other helminth diseases such as lymphatic filariasis, onchocerciasis, schistosomiasis and foodborne trematode infections, the World Health Organisation (WHO) has declared soil-transmitted helminth infections as a major Neglected Tropical Disease (Casulli, 2021). It is estimated that more than 1.5 billion people worldwide, or almost a quarter of the world's population (WHO, 2020), are specifically infected with intestinal worms and the disease is prevalent mainly in countries where clean water, sanitation and hygiene are poor in terms of accessibility and education (Strunz *et al*, 2014), including countries in sub-Saharan Africa, Latin America and South-East Asia (WHO, 2020). Human intestinal worm infections are predominantly caused by the hookworms *Necator americanus* and *Ancylostoma duodenale*, the whipworm *Trichuris trichiuria* and the roundworm *Ascaris lumbricoides* (Cross, 1996), which have been estimated to infect 439, 465 and 819 million people globally respectively (Hotez *et al*, 2014). Intestinal worms cause an estimated disease burden of 5.18 million disability-adjusted life years and rank first in the list of Neglected Tropical Diseases as published in the Global Burden Disease Study 2010 that quantified the health burden of parasitic diseases (Hotez *et al*, 2014; Pullan *et al*, 2014). This measure reflects the devastating impact that intestinal helminths have on the lives of people, who are caught in a vicious cycle of poverty and disease. Intestinal worms can establish chronic infections and are associated with gastrointestinal upset, loss of nutrition and anaemia as the worms feed on host tissues including blood, predisposition to co-infection, and sometimes death (Mckay *et al*, 2017; WHO, 2020). The effects of intestinal worms are even more devastating in children as the worms cause stunted growth and negatively affect cognitive function and school attendance (Connolly & Kvalsvig, 1993).

Although we lack a global overview of the prevalence of parasitic worms in animals, helminths and intestinal worms are also prevalent in domestic and wild animals, such as dogs, cats, sheep, cattle, goats, pigs, horses, birds, and deer. Studies examining the prevalence of intestinal parasites show that farm animals can be infected with moderate to high intestinal worm infections. For example, around 87% (n = 280) of sheep and 49% (n = 100) of goats tested from three cities in Myanmar were found infected with Trichostrongyle nematodes (Win *et al*, 2020), while approximately 18%

(n = 136) of the pigs tested in a rural area in Korea were infected with *Ascaris suum* (Ismail *et al*, 2010). The economic impact of helminth infections in ruminants is massive. The cost of helminth infections caused by intestinal nematodes including *Ostertagia ostertagi*, *Teladorsagia circumcincta* and *Haemonchus contortus*, the bovine lungworm *Dictyocaulus viviparus* (nematode) and the liver fluke *Fasciola hepatica* (flatworm) on ruminant livestock has recently been estimated in 18 countries of Europe: the cost on ruminant livestock production is estimated at €1.8 billion per year, while the cost of treatment against these helminths is estimated at €38 million per year (Charlier *et al*, 2020). Thus, helminth infections in ruminants lead to agricultural losses and economic damage, and have a negative impact on food security (Charlier *et al*, 2020).

Hart & Hart (2018) predict that almost all wild mammals are infected with intestinal parasites in general and that these animals do not necessarily exhibit pathological effects despite having modest parasite burdens. This suggests that chronic infections with worms have led to host acquisition of disease tolerance to parasitic worms (King & Li, 2018). Helminth infections, however, can have serious ecological consequences on wild populations. To exemplify, various groups report association of intestinal helminths to population crashes in Soay sheep (Gulland, 1992), red grouse (Hudson *et al*, 1998) and wild mice (Pedersen & Greives, 2008). Population crashes and lack of host tolerance to intestinal worms can be linked to a weak immune system and the availability of food resources (Gulland, 1992; Pedersen & Greives, 2008; Knutie *et al*, 2017).

It is not uncommon for the host (humans and animals) to be co-infected with other parasitic, viral, and bacterial infections. This is mainly attributed to the high prevalence of worm infections, the mode of helminth transmission (e.g. due to lack of clean water and sanitation), and the fact that other diseases (such as malaria) share similar geographic distributions (Lello *et al*, 2004; Hotez *et al*, 2007; Steinmann *et al*, 2010; Alemu *et al*, 2020; Wolday *et al*, 2021). Co-infections must be taken into consideration when studying the host-worm interactions and when thinking about the treatments and strategies for disease control that will be required.

1.1.2 Control strategies against helminths

The prevalence of parasitic helminth infections in humans and the wide range of morbidities they cause are recognised by WHO which has set out goals to expand

efforts at prevention and control of Neglected Tropical Diseases, including intestinal worm infections (Casulli, 2021). The targets set to be achieved by 2030 specifically for intestinal helminthiasis include: (1) elimination of morbidities caused in young children, (2) establishment of efficient control programmes for children, adolescents, pregnant and lactating women, and (3) ensuring access to basic sanitation and hygiene in areas where intestinal worms are endemic (WHO, 2020). Along with the aforementioned targets, WHO also recommends administration of medications for prevention and control of intestinal helminthiasis (WHO, 2020). Regarding livestock helminth infections, several organisations, namely the COMBAR (Combatting Anthelmintic Resistance in Ruminants), the LiHRA (Livestock Helminth Research Alliance) and the International Research Consortium STAR-IDAZ have come together to construct research roadmaps for the control of helminth infections in farm ruminants, including gastrointestinal nematodes (COMBAR, 2020). To control for infections in livestock, (1) improved diagnostic tools, (2) better therapeutics and (3) development of vaccines are required to be implemented in an integrative manner where the local epidemiology, farm management and climate change are taken into consideration (Vercruysse *et al*, 2018; COMBAR, 2020).

Currently there are only four classes of drugs against helminths (termed anthelmintic drugs) (Wit *et al*, 2021). These include benzimidazoles, macrocyclic lactones, amino-acetonitrile derivatives and nicotinic acetylcholine receptor agonists (Wit *et al*, 2021). However, these drugs do not prevent re-infection with parasitic worms. In fact, re-infection rates are high, with *A. lumbricoides* prevalence reaching 94% after 12 months of treatment and prevalence with the hookworms *A. duodenale* and *N. americanus* reaching 55% after 6 months of treatment (Jia *et al*, 2012). Moreover, anthelmintic resistance to all these drugs has been reported for livestock-prevalent parasitic worm species while some cases of drug resistance to human parasites, including *N. americanus*, have also been reported (Orr *et al*, 2019; Wit *et al*, 2021). Resistance in human-infective helminths is low but is expected to rise as mass drug administration programmes are established (Wit *et al*, 2021).

The morbidities caused by helminths, the economic and agricultural losses, the emerging drug resistance to anthelmintics and high rates of re-infection, highlight the need for development of other interventions, including vaccines, against human and livestock helminths. Vaccines can act as a prophylactic and/or therapeutic intervention. To date, there are four licensed anti-helminthic vaccines for animal use.

These control for infections with the bovine lungworm nematode *D. viviparus,* the sheep nematode *H. contortus,* the sheep- and goat-infective cestode *Echinococcus granulosus* (Claerebout & Geldhof, 2020), and the pig & human-infective cestode *Taenia solium* (Ouma *et al*, 2021). However, there are no licensed anti-helminthic vaccines for human use (Perera & Ndao, 2021), but there are some vaccine candidates in clinical trials (Zawawi & Else, 2020). It is worth noting that although vaccine resistance might still develop as reported for vaccines against other diseases including the hepatitis B virus (Sheldon & Soriano, 2008) and the bacteria *Streptococcus pneumoniae* (Brueggemann *et al*, 2007), this happens less frequently compared to emergence of drug resistance (Kennedy & Read, 2017).

1.2 Modelling gastrointestinal nematode infections

As stated above, intestinal helminths infect almost a quarter of the human population as well as animals and establish chronic infections that are associated with major economic and global health burdens. To be able to develop new therapeutic interventions we need to understand the biology of the worms, how they are able to survive in their hosts and the impact that they have on the host's immune system. This requires a range of approaches including studies in controlled laboratory experiments. Although studying the worms that cause human intestinal helminthiasis and the immune responses mounted by the host would be considered the most relevant to human disease, it comes with a lot of challenges. Such studies require human participants who, for example, are likely to be co-infected with other diseases and/or who will have already been chronically infected for an unknown period of time. These would add variability in the study and thus make it more difficult to directly decipher the effects of the intestinal worm of interest. Hence, we model human intestinal helminthiasis by studying close relatives of human intestinal worms that infect animals including rodents, livestock and canines (Boes & Helwigh, 2000; Bethony *et al*, 2005; Montaño *et al*, 2021a). Employing such models has disadvantages too, as we do not fully mimic what happens in nature (Colombo & Grencis, 2020). This is particularly evident when bolus (large, single dose of larvae) infections are used for infecting the animal models, and the absence of co-infection. Nevertheless, model infections have allowed us to study how the worm parasites successfully modulate the immune system of their host(s) (immunomodulation) thus providing us with the knowledge and understanding to not only work towards developing drugs against helminths, but also to utilise molecular mechanisms exhibited by worms for treatment of allergies, inflammation and autoimmune diseases

(Proudfoot, 2004; Wu *et al*, 2017). One such model for persistent gastrointestinal worm infections is the nematode *Heligmosomoides bakeri*. It is worth noting that *H. bakeri is* often referred to as *Heligmosomoides polygyrus* in the literature (Behnke *et al*, 2009). To distinguish *H. bakeri* from *H. polygyrus* that infects wild mice, in this thesis *H. bakeri* is used to denote the parasite species that is passaged in laboratory mice, based on the latest genome naming (Coghlan *et al*, 2018).

Genome-wide phylogenetic analysis organises nematodes to different clades. The human hookworm *N. americanus*, the sheep-infective nematodes *T. circumcincta* and *H. contortus* and the mouse-infective *H. bakeri*, cluster in the nematode clade V (Coghlan *et al*, 2018). *H. bakeri* is used to model the chronic infection and immunomodulation exhibited by the human hookworm *N. americanus* and livestock-infective nematodes in laboratory settings (Reynolds *et al*, 2012). *H. bakeri* naturally infects mice, and can establish long-term infections lasting up to 10 months in susceptible mice (Robinson *et al*, 1989), in comparison to other hookworm models like the parasitic nematode *Nippostrongylus brasiliensis* which is only able to infect rodents (rats and mice) for 2-3 weeks (Day *et al*, 1979). However, the life cycle of *H. bakeri* makes it less pathogenic compared to *N. brasiliensis*, which has a similar life cycle to *N. americanus. N. americanus* penetrates the human skin and enters in the blood circulation to reach the heart and then the lungs to then be coughed up and swallowed to make it to the intestine of the host (Loukas *et al*, 2016). On the other hand, *H. bakeri* completes its life cycle wholly in the host gut. Infective larvae are ingested by mice and reach the small intestine where they burrow through the wall of the small intestine to the submucosa. There the larvae moult twice and in approximately 10 days post-ingestion they emerge in the gut lumen as adult worms (Fig. 1.2) (Reynolds *et al*, 2012). Moreover, *N. americanus* and *N. brasiliensis* feed on mucosal tissue and blood, causing anaemia (Williamson *et al*, 2002; Bouchery *et al*, 2018), whereas *H. bakeri* feeds on host tissue but not on blood (Bansemir & Sukhdeo, 1994). Despite these differences, *H. bakeri* and the human- and animal-infective nematodes induce similar immune responses in the host and exhibit comparable immunoregulatory properties (McSorley & Maizels, 2012). Hence, we can extrapolate findings from studying *H. bakeri* to use as a basis for human and livestock nematode infections. Additionally, the advanced genetic tools that are available in the mouse host make *H. bakeri* an attractive model organism for human and animal gastrointestinal nematode infections.

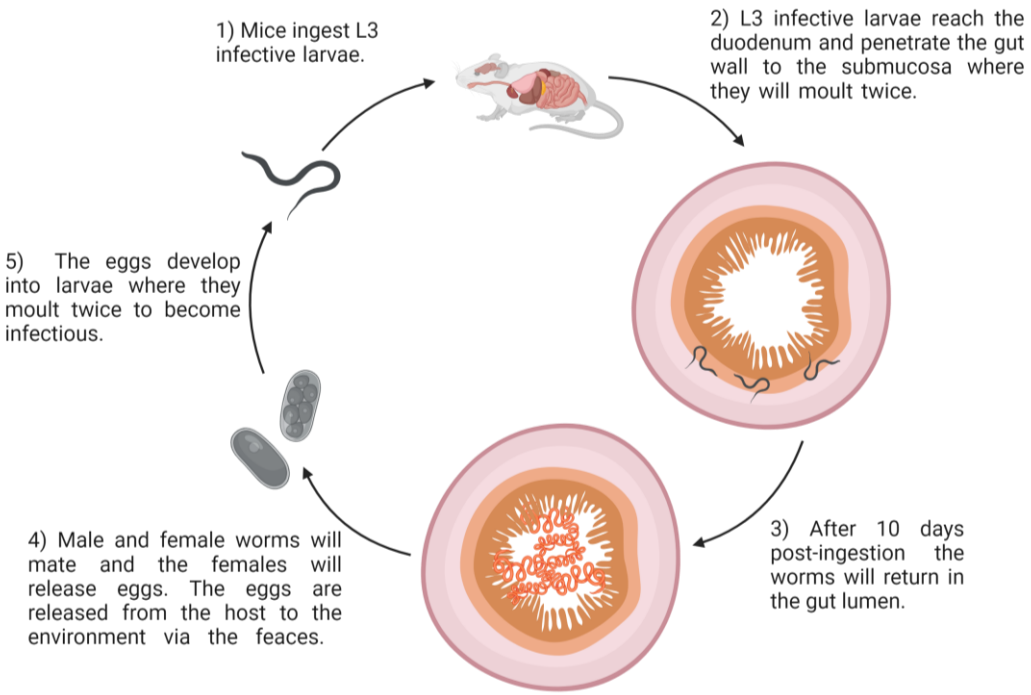


Figure 1.2 | Life cycle of *H. bakeri*. Figure adapted from Reynolds *et al,* (2012), and created with BioRender.com.

1.3 Host immune responses to helminth infections at the barrier

Helminths have co-evolved with mammals for long periods such that evolutionary selection pressures have led to evolution of host disease tolerance (Girgis *et al*, 2013; King & Li, 2018). Gastrointestinal nematode infections induce innate and adaptive type 2 cellular and cytokine responses. Type 2 immunity is protective and also promotes wound healing and tissue repair. Most gastrointestinal nematodes (and generally most helminths) come in contact with and/or breach the epithelial cell barrier (Mckay *et al*, 2017). The intestinal epithelial cells play a role in the initiation and effector functions of the type 2 immunity. The gut epithelial cells sense the worms through pattern recognition receptors and/or the disruption of the epithelial cell layer inflicted by the worms burrowing in the gut wall stimulates the gut epithelial cells to mount the first response to helminths (Babu & Nutman, 2019; Bouchery *et al*, 2019). Epithelial cells release a range of signals. For example, they produce antimicrobial peptides and secrete the chemokine thymic stromal lymphopoietin (TSLP) and cytokines such as IL-33 (released from necrotic epithelial cells) and IL-25 (released from epithelial tuft cells), and alarmin molecules (Babu & Nutman, 2019). These molecules are important for the initiation of type 2 responses, which regulate effector

proteins that disrupt the environment where the parasite lives in the host (i.e. intestinal lumen).

The IL-25, IL-33 and TSLP released by the epithelium stimulate the group 2 innate lymphoid cells (ILC2s). Activated ILC2s co-ordinate the type 2 immunity cytokines to mediate protection against the parasite (Bouchery *et al*, 2019). ILC2s release effector cytokines including IL-4, IL-5, IL-9 and IL-13 (Babu & Nutman, 2019). IL-4 and IL-13 stimulate differentiation and hence expansion of the epithelial tuft cells which in turn promotes further release of IL-25 thus driving a positive feedback loop of type 2 responses (von Moltke *et al*, 2015). The epithelial stem cells also differentiate to goblet cells hence expanding the goblet cell population (a phenomenon called goblet cell hyperplasia) after stimulation with IL-4 and IL-13. Goblet cells secrete mucus and the resistin-like molecule (RELM)-β, which are involved in trapping the parasite, obscuring parasite mobility and feeding, and inhibiting the chemosensory functions of the worms, thus affect parasite fitness enabling its expulsion (Artis *et al*, 2004; Herbert *et al*, 2009; Sharpe *et al*, 2018). The contractility of the smooth-muscle of the gut also increases, resulting in increased peristaltic movement, contributing to what is called the “weep and sweep” response for parasite expulsion (Bouchery *et al*, 2019). The IL-9 released by ILC2s targets mast cells which are involved in wound healing and tissue repair as well as the expulsion of worms (Reitz *et al*, 2018). Additionally, the IL-5 also secreted from ILC2s drives eosinophil responses. Eosinophil responses have proinflammatory and immunoregulatory roles and are involved in defence against mucosal pathogens (Travers & Rothenberg, 2015). Through their actions, eosinophils promote wound healing but are also implicated in killing the parasite (Travers & Rothenberg, 2015).

Moreover, epithelial-derived cytokines such as TSLP secreted from the intestinal epithelium in response to the parasitic infection and/or due to tissue damage, are detected by dendritic cells (antigen-presenting cells) (Sorobetea *et al*, 2018). Activated dendritic cells present worm antigens to naive T-cells instructing them to differentiate into the T-helper cell type 2 (Th2) subset in the presence of IL-4 which is produced by basophils (cells that exert functions similar to antigen-presenting cells) (Hammad & Lambrecht, 2015). Th2 cells secrete cytokines including IL-4, IL-5, IL-9 and IL-13 similarly to ILC2s (Walker & McKenzie, 2017), reinforcing the positive feedback loop of mucus production via goblet cell expansion, IL-25 production via tuft cell expansion, and the muscle contraction for expulsion of worms. However, another

major role for Th2 cells is to prime B-cells to generate IgG1 and IgE antibody isotypes that are parasite specific to aid parasite expulsion (Harris & Gause, 2011; Matsumoto *et al*, 2013). Although less well studied, IgA isotype antibodies generated by B-cells after activation by dendritic cells or T-regulatory cells (Macpherson *et al*, 2012), are associated with lower worm burdens and are thought to inactivate metabolic enzymes, interfering in this way with the growth, metabolism, fecundity and motility of the worms (Gill *et al*, 1993; Clerc *et al*, 2018).

During infection, pathogen-associated molecular patterns and the release of cytokines also leads to activation of macrophages (Rolot & Dewals, 2018). The cytokine IL-4 specifically induces alternatively activated macrophages (AAMs) or M2 cells (Stein *et al*, 1992). AAMs are involved in trapping worms and affect parasite fitness and viability (Babu & Nutman, 2019). AAMs are also involved in wound healing and in regulating inflammation. This is achieved by upregulation of anti-inflammatory cytokines and other molecules like the chitinase protein Ym1, Arginase-1 and the resistin-like molecule (RELM)-α (Sutherland *et al*, 2018). In this way, AAMs play a role in dampening the immune responses by Th2 cytokines to avoid a hyper-immune response (Pesce *et al*, 2009). AAMs and dendritic cells can also induce T-regulatory cells (White *et al*, 2020). T-regulatory cells protect the host from worm-associated pathologies in the intestine while at the same time subversion of this response facilitates the survival of the parasite (D'Elia *et al*, 2009; White *et al*, 2020). Lastly, although type 2 responses are generally involved in host protection against intestinal worms, chronic infections established by some helminth species are associated with the expansion of T-regulatory cells or induction of type 1 immune responses which involve Th1 cells and IFN-γ (Cortés *et al*, 2017). A simplified overview of the host immune responses to intestinal helminths is shown in Figure 1.3.

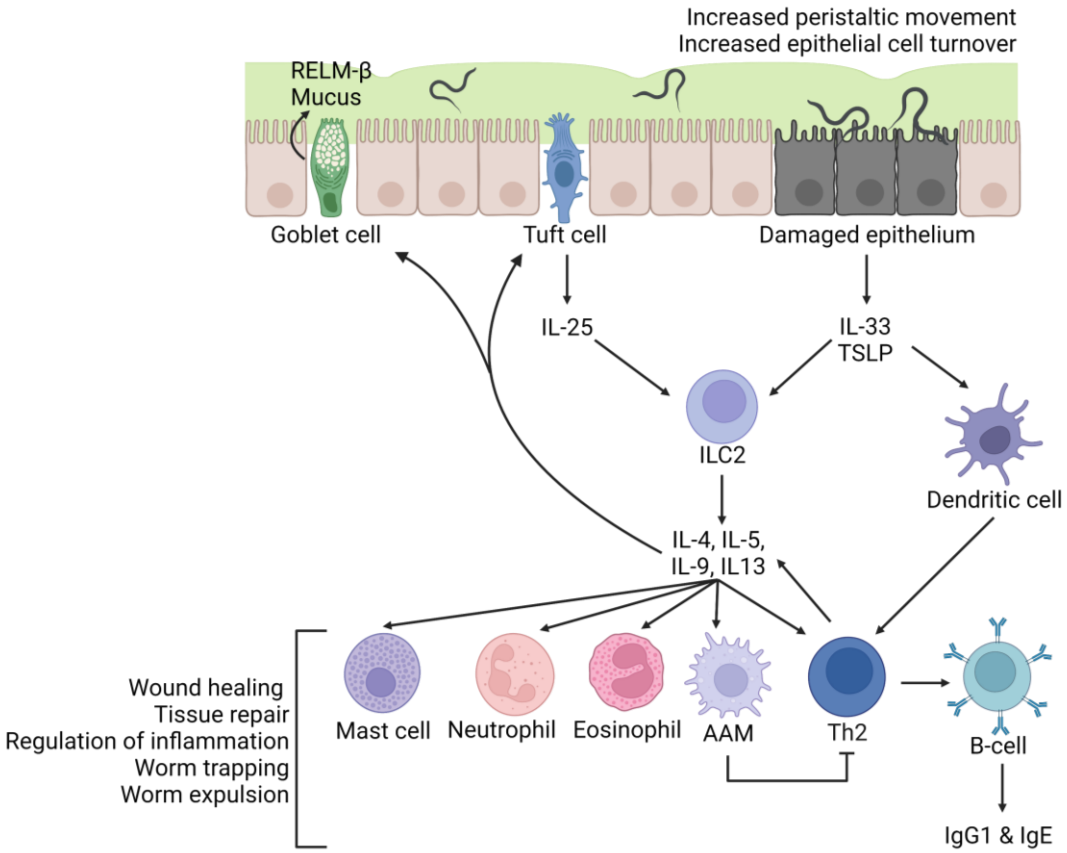


Figure 1.3 | Simplified overview of the host immune response against gastrointestinal helminth infection. Sensing of intestinal worms by the gut epithelium leads to release of IL-25 by tuft cells, and IL-33 and TLSP from damaged cells. These molecules activate ILC2 which release IL-4, IL-5, IL-9 and IL-13. These cytokines, along with dendritic cells, induce differentiation of naïve T-cells to Th2 cells which can in turn produce more IL-4, IL-5, IL-9 and IL-13 cytokines and also induce B-cells to generate parasite-specific IgG1 and IgE antibody isotypes. The IL-4, IL-5, IL-9 and IL-13 cytokines also induce goblet cell and tuft cell hyperplasia. Intestinal epithelial cell turnover and peristaltic movement (muscle contractility) increase, aiding to expel the worms. Mast cells, neutrophils, eosinophils and AAMs are also induced by the IL-4, IL-5, IL-9 and IL-13 cytokines. AAMs specifically regulate inflammation by downregulating Th2 responses. Along with mast cells, neutrophils and eosinophils, AAM promote wound healing and tissue repair, worm trapping and worm expulsion. Figure created using BioRender.com. AAM = alternatively activated macrophages; Ig = immunoglobulin; IL = interleukin; ILC2 = innate lymphoid type-2 cells; RELM = Resistin-like molecule; Th2 = T-helper 2 cells; TSLP = Thymic stromal lymphopoietin.

1.4 Host immune modulation by helminth products

Despite the immune responses triggered by helminths, these worms very successfully parasitise their hosts and establish chronic infections. In contrast to other pathogens like viruses, bacteria, and unicellular protozoa which overtake the host immune system through rapid multiplication, parasitic worms manipulate host immune responses without rapid expansion (Maizels, 2020). Co-existence of helminths with their hosts for millions of years and the evolutionary arms race between them has allowed helminths to evolve a diverse repertoire of excreted/secreted immunogenic and immunomodulatory molecules (Maizels *et al*, 2018).

Helminths achieve modulation of type 2 and type 1 immune responses by excretion/secretion of various bioactive molecules, collectively known as ES (Excretory/Secretory) products, in the host environment where they reside. These effector molecules can impact different phases of the immune response, from the initiation of the response to wound healing and tissue remodelling (Maizels *et al*, 2018). Interestingly, some of the proteins released by helminths are homologous to host molecules. This includes the cysteine protease inhibitor proteins cystatins which are released by several helminth species including the gastrointestinal nematodes *H. bakeri*, *N. brasiliensis*, *A. lumbricoides*, and the filarial nematodes *Litomosoides sigmodontis* and *Onchocerca volvulus* (Dainichi *et al*, 2001; Schönemeyer *et al*, 2001; Pfaff *et al*, 2002; Sun *et al*, 2013; Coronado *et al*, 2017). The helminth cystatins interfere with the processing of antigens in antigen-presenting cells (Dainichi *et al*, 2001) and inhibit proliferation of T-cells (Schönemeyer *et al*, 2001). Helminth cystatins also induce IL-10 production in macrophages (Schönemeyer *et al*, 2001), a cytokine that possesses anti-inflammatory properties and is associated with limiting host immune responses against pathogens (Iyer & Cheng, 2012). Other ES products generated by helminths have developed *de novo*, meaning that they share no obvious homology to host products (Maizels *et al*, 2018). Examples include the proteins *Ancylostoma* secreted protein 2, NaASP-2, generated by *N. americanus* (Tribolet *et al*, 2015) and the Alarmin Release Inhibitor HpARI derived from *H. bakeri* (Osbourn *et al*, 2017). NaASP-2 binds a component of the B-cell receptor complex and supresses the expression of messenger RNAs (mRNAs) that encode proteins involved in the migration of leukocytes and the signalling pathways of B-cells (Tribolet *et al*, 2015). In this way, NaASP-2 modulates the adaptive immune responses of the host (Maizels *et al*, 2018). HpARI downregulates initiation of type 2 responses by binding IL-33 and preventing its release from necrotic cells (Osbourn *et al*, 2017).

The prevalence of helminth infections is negatively associated with allergy-related and autoimmune diseases which are prevalent in developed countries (Pearce *et al*, 2000). This pattern is consistent with the hygiene hypothesis, which proposes that lack of exposure to infections resulting from improved household sanitation and personal hygiene explains the rise of allergies in children (Strachan, 1989). Allergens and autoimmune diseases trigger a type 2 immune response (Akdis *et al*, 2020), which is very similar to the immune responses induced by helminths. As helminths excrete/secrete bioactive molecules with immunomodulatory properties some of which act to suppress type 2 host responses, the scientific community is working towards understanding whether these could be used as therapeutics for allergens and autoimmune diseases (Maizels *et al*, 2018; Maizels, 2020).

Beyond host modulation, it is important to be aware that helminth ES products could be involved in other processes. ES products could mediate inter-species interactions (i.e. communication between species) with species other than just the host as discussed above. For example, ES products could facilitate communication between the parasite and other pathogens that co-infect the host, and/or the microbiome. ES products could also be involved in intra-species communication (i.e. communication within species).

Interestingly, infection with gastrointestinal helminths changes the composition and diversity of the gut and faecal microbiota (Walk *et al*, 2010; Jenkins *et al*, 2017; Rapin *et al*, 2020). Helminths and bacteria interactions can shape host immune responses as shown by lack of attenuation of severe inflammation in *H. bakeri*-infected mice that were treated with antibiotics to reduce intestinal bacteria, compared to untreated *H. bakeri*-infected mice (Zaiss *et al*, 2015). Hence, through the microbiome, *H. bakeri* was able to confer protection against allergic asthma (Zaiss *et al*, 2015). Additionally, *H. bakeri* ES products have antimicrobial activity but lack of gut microbiota in mice is associated with reduced parasite fecundity and fitness (Rausch *et al*, 2018). This suggests that, although our understanding of host-microbiota-helminths is limited, the interactions are complex and the outcome of these interactions could be variable (Rapin *et al*, 2020).

1.5 Helminth Extracellular Vesicles (EVs)

Helminth ES products have mostly been studied for their role in host-pathogen interactions due to the prevalence of parasitic worms and the need of therapeutic interventions, as opposed to interactions with other worms of the same species or

other organisms. As described previously, several studies attest to the importance of the ES molecules in modulation of immune responses (Maizels *et al*, 2018). To understand the biology of these molecules and develop intervention strategies against the worms however, we need to study not only the effects of ES products on the host but also: (1) how these molecules are exported, (2) how they are selected for export, (3) whether they are delivered to their target(s), and if yes (4) how. The effector molecules excreted/secreted by helminths are not only limited to proteins, but helminths ES products also include carbohydrates/glycans (Hokke & van Diepen, 2017), metabolites (Whitman *et al*, 2021), nucleic acids such as mRNA and small RNAs (sRNAs) (Britton *et al*, 2014; Eichenberger *et al*, 2018b), lipids (Hewitson *et al*, 2009) and extracellular vesicles (EVs) (Drurey & Maizels, 2021). The EVs (lipid bilayer-enclosed particles) can contain any of the aforementioned molecules (Drurey & Maizels, 2021).

The gastrointestinal tract is enriched with digestive enzymes including proteases and nucleases which can break down proteins and nucleic acids respectively (Motta *et al*, 2011; Liu *et al*, 2015). The EVs secreted by helminths could provide a safe and efficient transport mechanism to protect the parasite cargo molecules and enable their internalisation into host cells (Britton *et al*, 2020). Studies on the cargo of helminth EVs using proteomic (Ditgen *et al*, 2014; Montaño *et al*, 2021b), (small) RNA (Britton *et al*, 2014; Claycomb *et al*, 2017; Eichenberger *et al*, 2018b) and more recently glycomic analyses (Whitehead *et al*, 2020) have revealed candidate molecules that might be important in the immunomodulation of host cells. Helminth EVs were described almost 50 years ago in the trematode *F. hepatica* (Threadgold, 1963), but they were not understood or recognised as a carrier of cargo. It is only in the last decade that the involvement of EVs in intra- and extracellular signalling has begun to be reported (Marcilla *et al*, 2012) and their functional properties realised (Buck *et al*, 2014). EVs can be released by helminths from various sources such as the excretory/secretory pores, the intestine, the mouth, the anal pore and the surface (the natural body covering) of the worm (Drurey *et al*, 2020). The EVs can be internalised by host cells in a clathrin-dependent endocytosis and clathrin-independent manner via a receptor-mediated endocytosis or by direct fusion of the EV lipid bilayer with the host cell membrane (Zakeri *et al*, 2018). It is not yet clear whether uptake of helminth EVs is cell type-specific or not. However, evidence from the mammalian EV field suggests that various considerations, including the heterogeneity of EVs, can hinder the ability to draw conclusions about EV specificity (Mulcahy *et al*, 2014).

Several research groups report uptake of helminth EVs by dendritic cells, macrophages and epithelial cells (Drurey *et al*, 2020). Our group has shown that the gastrointestinal nematode *H. bakeri* EVs are internalised by mouse intestinal epithelial cells and bone-marrow derived macrophages *in vitro* (Buck *et al*, 2014; Coakley *et al*, 2017). However, further experiments are currently underway to understand if there is target cell specificity by *H. bakeri* EVs in the epithelial barrier which is composed of at least 7 types of cells (enterocytes, enteroendocrine cells, tuft cells, goblet cells, stem cells, Paneth cells, and M cells) (Allaire *et al*, 2018). Internalisation of *H. bakeri* EVs by mouse intestinal epithelial cells leads to suppression of host genes involved in immunity and inflammation, such as the genes coding for the IL-33 receptor and the Dusp1 phosphatase (involved in upregulation of the immunosuppressive cytokine IL-10 (Hammer *et al*, 2006)) (Buck *et al*, 2014). Uptake of *H. bakeri* EVs by bone-marrow derived macrophages supressed molecules involved in type 1 and type 2 immune responses, including Ym1 and RELM-α (Coakley *et al*, 2017). Antibodies raised against these EVs through EV vaccination confer strong protection against infection (Coakley *et al*, 2017), suggesting the EVs and their cargo are important for parasite survival *in vivo*. Furthermore, the gastrointestinal nematodes *Trichuris muris* and *N. brasiliensis* also secrete EVs. Eichenberger and colleagues illustrate that *T. muris and N. brasiliensis* EVs are actively internalised by epithelial cells within small intestinal and colonic organoids respectively (Eichenberger *et al*, 2018a, 2018b). Production of pro-inflammatory cytokines including IL-1β, IL-6, IL-17a, and IFN-γ is suppressed in mice after treatment with *N. brasiliensis* EVs while the levels of the anti-inflammatory cytokine IL-10 was increased (Eichenberger *et al*, 2018a). The human- and animal-infective trematode *Schistosoma japonicum* also releases EVs. The parasite resides in the blood vessel, drains blood from the intestine and can cause liver cirrhosis (scarring). Interestingly, *S. japonicum* EVs are taken up by mouse liver cells *in vitro* (Zhu *et al*, 2016), and primarily by macrophages but also other peripheral blood immune cells (T-cells, B-cells and Natural killer cells) *in vivo* and *in vitro* (Liu *et al*, 2019a). *S. japonicum* EV uptake resulted in increased cell proliferation in macrophages and elevated levels of the inflammatory cytokine TNF-α, which were shown to be important for egg production and parasite survival (Liu *et al*, 2019a).

1.6 Small RNAs (sRNAs)

Small RNAs (sRNAs) are also present in helminth ES products including EVs. sRNAs are a class of non-coding RNA molecules (i.e. do not code for a protein) that are less than 200 nucleotides (nt) in length (Sandberg *et al*, 2013). There are three major

classes of sRNAs: microRNAs (miRNAs), small/short interfering RNAs (siRNAs), and PIWI-interacting RNAs (piRNAs) (Zhang, 2009). sRNAs are approximately 18 to 30 nt long (Zhang, 2009). Along with other sRNAs such as secondary siRNAs and more recently transfer RNA (tRNA)-derived fragments (tRFs), miRNAs, siRNAs and piRNAs function as guides to direct Argonaute (AGO) proteins to targets in order to regulate target gene expression (Gu *et al*, 2009; Youngman & Claycomb, 2014; Kuscu *et al*, 2018). Gene expression can be regulated by AGO:sRNA complexes at the epigenetic level via DNA methylation and heterochromatin formation (known as transcriptional gene silencing) (Holoch & Moazed, 2015), or post-transcriptionally by mRNA degradation and/or translational inhibition/arrest (known as post-transcriptional gene silencing) (Höck & Meister, 2008). The various mechanisms of regulating gene activity are discussed further in Section 1.7.

The first miRNA, lin-4, was discovered by the Ambros and Ruvkun groups who showed that lin-4 can repress expression of the lin-14 gene through direct interactions in the 3' untranslated region (3'UTR) of the lin-14 mRNA, and this is important for regulating developmental timing in the free-living nematode *Caenorhabditis elegans* (Lee *et al*, 1993; Wightman *et al*, 1993). The Baulcombe lab further showed that plants express siRNAs that are around 25 nt in length. These siRNAs have antisense polarity and are complementary to the target mRNAs, and accumulate after transgene- and virus-induced post-transcriptional gene silencing (Hamilton & Baulcombe, 1999). The authors later showed that the siRNAs are implicated in RNA interference (RNAi; a process by which gene expression is repressed) (Hamilton *et al*, 2002). Since then, our understanding of the functional role(s) of sRNAs has greatly expanded. sRNAs are involved in gene regulation mainly by silencing gene expression and are implicated in various processes including development, cell differentiation, proliferation and apoptosis, and immunity/defence. Deregulation of sRNAs and their function(s) is associated with diseases like cardiovascular and neurodegenerative diseases, and cancer (Zhang, 2009; Gebert & Macrae, 2019).

1.6.1 miRNA biogenesis

miRNAs are on average 22 nt in size. In metazoans, canonical miRNAs are transcribed in the cell nucleus by RNA polymerase II which generates the primary miRNAs (pri-miRNAs) (Lee *et al*, 2002, 2004) that can derive from exons or introns of non-coding transcripts or from introns of precursor mRNAs. Pri-miRNAs can be several kilobases long, have a 5' 7-methyl guanylate (m7G) cap and a 3' poly(A) tail

(Cai *et al*, 2004; Lee *et al*, 2004), and fold back on themselves to generate a hairpin precursor (Bartel, 2018). The Microprocessor complex made of the proteins Drosha and DGCR8 (DiGeorge syndrome critical region 8) cleaves the pri-miRNA to release a stem-loop precursor miRNA (pre-miRNA) of around 70 nt (Lee *et al*, 2002; Han *et al*, 2006). This cleavage leaves a monophosphate feature at the 5' ends of the molecule (later referred to in this thesis as 5'P). The pre-miRNA is then exported to the cytoplasm by Exportin 5 and RAN-GTP and processed by Dicer which cleaves near the loop, leaving a mature miRNA duplex (Yi *et al*, 2003). The miRNA duplex is then loaded into an Argonaute (AGO) protein where one of the strands is discarded and degraded (passenger strand) and the other strand (guide strand) remains associated with AGO (Bartel, 2018) (Fig. 1.4A).

Non-canonical miRNAs are generated slightly differently to canonical miRNAs. Non-canonical miRNAs bypass the Microprocessor complex but feed into the later steps of the canonical miRNA pathway (Bartel, 2018). An example of non-canonical miRNAs include mirtrons which originate from introns but are processed by the spliceosome (Ruby *et al*, 2007). The so-called chimeric hairpin RNAs are another example of non-canonical miRNAs. These are generated with or as part of other sRNA types (Bartel, 2018). The miRNA miR-1839 is an example as it is transcribed within the small nucleolar RNA (snoRNA), ACA45 (Ender *et al*, 2008). In this case, the snoRNA is processed first and then the molecule is independent of processing by the Microprocessor complex but requires Dicer (Ender *et al*, 2008).

1.6.2 Biogenesis of siRNAs & secondary siRNAs

siRNAs are also about 22 nt in length and are generated in a similar fashion to miRNAs. However, siRNAs are processed from long, linear, double-stranded RNAs rather than hairpin precursors (Sheu-Gruttadauria & MacRae, 2017). The double-stranded RNA is cleaved by Dicer into the siRNA duplex, leaving a 5' monophosphate at the ends of each strand. Similarly to miRNAs, the duplex is loaded into AGO and the passenger strand is discarded whilst the guide strand is retained with AGO and will serve as the guide to direct AGO to its target to directly regulate gene expression (Bartel, 2018) (Fig. 1.4B).

In plants and nematodes, different classes of sRNAs can act as primers to amplify the signal for gene regulation by generating secondary sRNAs (Pak & Fire, 2007). Based on work in *C. elegans*, siRNAs generated by Dicer (sometimes referred to as primary siRNAs) are used by the RNA-dependent RNA Polymerases (RdRPs) RRF-1 and

EGO-1 to generate secondary siRNAs from an mRNA template (Aoki *et al*, 2007; Maniar & Fire, 2011). Secondary siRNAs are synthesised *de novo*, are single-stranded and possess a triphosphate moiety at their 5' end (referred to as 5'PPP in this thesis) (Pak & Fire, 2007). The secondary siRNAs are loaded into nematode-specific AGO proteins called Secondary AGOs or WAGOs (Worm-specific AGO proteins) to regulate the activity of genes (Ketting & Cochella, 2020) (Fig. 1.4B). In *C. elegans*, secondary siRNAs can also be generated from primary piRNAs. piRNAs are small non-coding RNAs of about 21 nt in size in *C. elegans* (26-31 nt in length in *Drosophila melanogaster* and other vertebrates) (Weick & Miska, 2014). piRNAs are mainly expressed in the germline, and associate with a specific class of AGO proteins, the PIWI subfamily (Girard *et al*, 2006; Vagin *et al*, 2006). piRNAs were initially identified to supress the activity of mobile genetic elements thus they play a role in genome defence in the germline, a function conserved among various organisms (Siomi *et al*, 2011; Britton *et al*, 2020). piRNAs are also expressed in somatic cells and emerging evidence shows that they play an important role in regulating gene expression and are implicated in disease (Wu *et al*, 2020). However, the extent to which they operate in each context remains to be elucidated. Although the biogenesis route of piRNAs differs between organisms, piRNA generation is independent of Dicer and does not involve a double-stranded RNA precursor (Weick & Miska, 2014). In *C. elegans*, piRNAs can associate with the PIWI-AGO protein PRG-1 to then initiate synthesis of the secondary siRNAs by the RdRPs RRF-1 and EGO-1. These can then be incorporated into WAGO proteins such as HRDE-1, a germline secondary AGO protein (Weick & Miska, 2014).

The mechanisms for secondary siRNA biogenesis in plants differs, and several biogenesis pathways exist. For example, phasiRNAs are a class of secondary siRNAs which are synthesised in a phased manner (i.e. head-to-tail arrangement starting precisely at a specific nucleotide) with the involvement of miRNAs, RdRPs, and Dicer (Fei *et al*, 2013). Briefly, formation of sRNA precursors is triggered through cleavage of non-coding RNA or mRNA facilitated by AGO:miRNA complexes (Fei *et al*, 2013). These sRNA precursors are then amplified by an RdRP (RDR6) into double-stranded sRNAs which are then processed by Dicer proteins (Deng *et al*, 2018). Processing by different Dicer proteins leads to generation of secondary siRNAs that are of different lengths (21 or 24 nt) which do not possess a 5'PPP. These can then be loaded into AGO proteins to regulate gene expression (Deng *et al*, 2018).

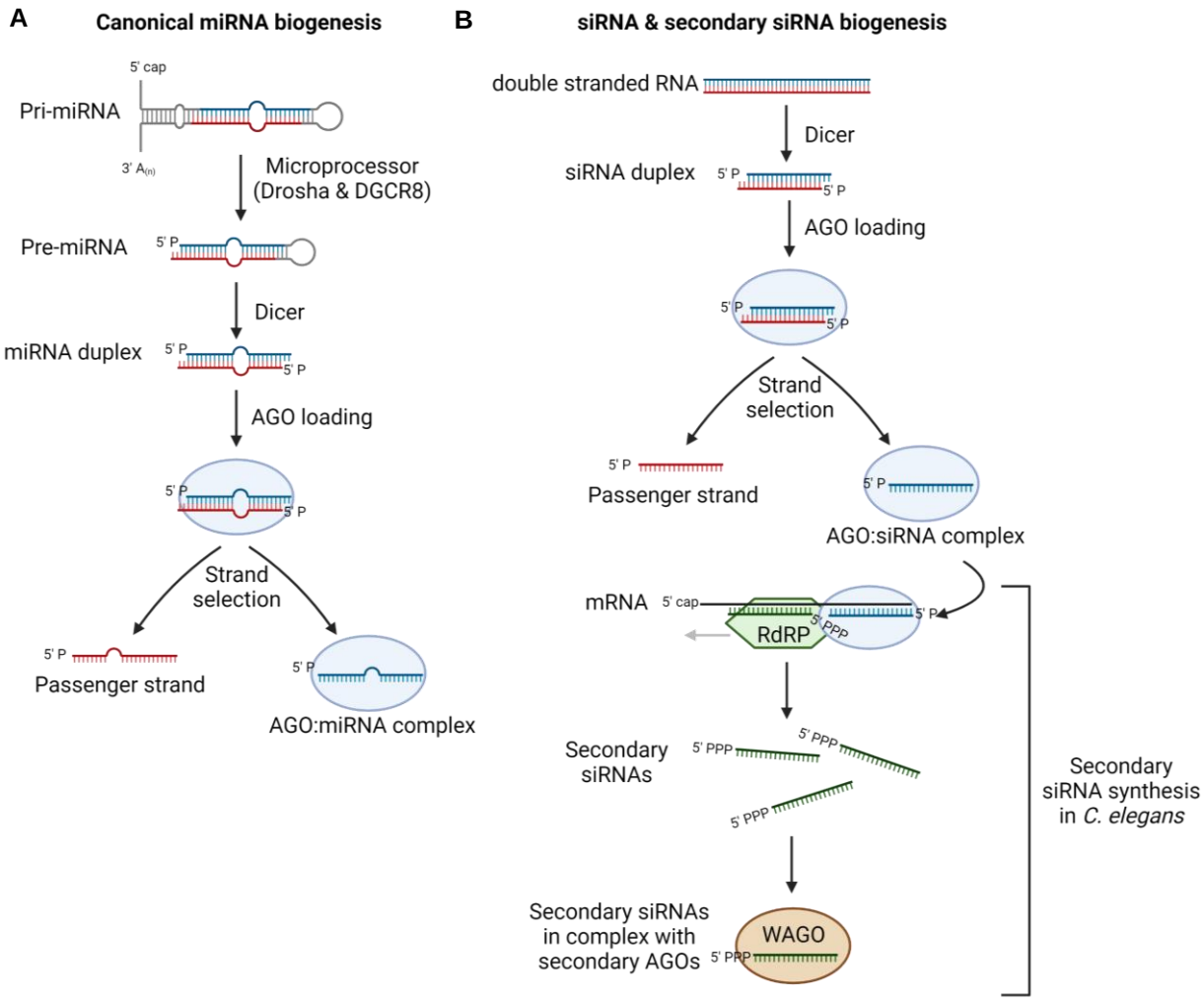


Figure 1.4 | Biogenesis of (a) canonical miRNAs, (b) siRNAs and secondary siRNAs. The figure shows the canonical miRNA biogenesis pathway and one mechanism of biogenesis of secondary siRNAs that occurs in *C. elegans* after generation of primary siRNAs. The figure was created using BioRender.com. DGCR8 = DiGeorge syndrome critical region 8; RdRP = RNA-dependent RNA polymerase; pri-miRNA = primary miRNAs; pre-miRNA = precursor miRNA.

It is worth noting that not all organisms express all the known types of sRNAs. For example, miRNAs are expressed in eukaryotes (mammals, worms, plants, insects and protists) with only a limited number of miRNAs or miRNA-like RNAs found in bacteria and viruses (Carthew & Sontheimer, 2009; Tarver *et al*, 2012; Cardin & Borchert, 2017). On the other hand, siRNAs are found in plants, worms and insects but a very limited number are detected in mammals, although mammalian cells can process exogenous double-stranded RNA to siRNAs (Langer *et al*, 2012). The less well-known class of sRNAs, secondary siRNAs, are deemed to be unique to plants, fungi, and nematodes but are not present in other worms or animals (Sijen *et al*, 2001; Sugiyama *et al*, 2005; Sanan-Mishra *et al*, 2021).

1.7 Argonaute (AGO) proteins: The heart of RNA interference

The discovery of small RNAs as one of the released components of helminths strongly suggests that RNA could be an immunomodulator molecule that can act inside recipient cells (Claycomb *et al*, 2017). This hypothesis is reinforced by studies reporting EV internalisation by host cells *in vitro* and *in vivo* (Drurey & Maizels, 2021). It has thus been of interest to the scientific community to understand the mechanisms of export and internalisation of sRNA molecules as well as their functional mechanism and the effect they might exert on the host. However, sRNAs are powerless on their own and are also susceptible to degradation (Ketting & Cochella, 2020). To regulate gene expression, sRNAs operate in complex with AGO proteins.

1.7.1 AGO proteins and RNA interference

AGO proteins are encoded by all three domains of life (Archaea, Bacteria and Eukaryota) (Swarts *et al*, 2014). Eukaryotic AGOs are characterised by their bilobal structure consisting of four domains: the N-terminal domain, the PAZ (PIWI-Argonaute-Zwille) domain, MID (Middle) domain and the C-terminal PIWI (P-element Induced WImpy testis) domain (Willkomm *et al*, 2015) (Fig 1.5). Prokaryotic AGOs (pAGOs; archaeal and bacterial) can have additional domains, while a large class of pAGOs termed short pAGOs, lack the N-terminal and PAZ domains (Ryazansky *et al*, 2018). AGO proteins associate with sRNAs where the MID and PAZ domains bind the 5'- and the 3'-end of the small nucleic acid guide respectively (Lingel *et al*, 2004; Boland *et al*, 2010). The PIWI domain has an RNase H-like fold and can possess endonuclease/slicer activity (Parker *et al*, 2004). The ability to cleave a target is provided by conserved amino acid residues, referred to as the catalytic triad or tetrad (DDX or DEDX, where 'X' is usually D or H), and requires

magnesium ions (Mg^{2+}) (Rivas *et al*, 2005; MacRae *et al*, 2008; Park & Shin, 2014). However, possessing the catalytic motif does not necessarily equate to slicer activity. It is proposed that post-translational modifications of AGO or the guide nucleic acid itself may contribute to the endonuclease activity (Höck & Meister, 2008; Park *et al*, 2017). For example, Park *et al*, (2017) showed that recombinant human AGO3 can cleave the complementary target of miR-20a but not of let-7, miR-16 and miR-19b.

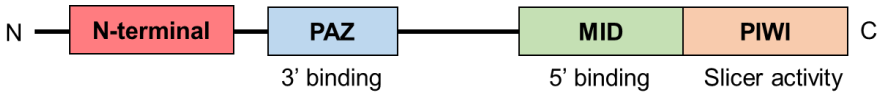


Figure 1.5 | Domain composition of eukaryotic Argonaute proteins. The MID domain binds the 5' end of the guide sRNA, the PAZ domain binds the 3' end of the guide sRNA while the PIWI domain can possess slicer activity. PAZ = PIWI-Argonaute-Zwille; MID = Middle; PIWI = P element Induced WImpy testis.

Phylogenetically, the eukaryotic AGO family is divided into the AGO, PIWI, WAGO (Worm-specific AGO) and *Trypanosoma* AGO subfamilies (Garcia Silva *et al*, 2010; Swarts *et al*, 2014). Eukaryotic AGO proteins associate mostly with sRNAs and target RNA transcripts whereas prokaryotic AGOs can bind short DNA guides and are directed preferentially to DNA loci (Ryazansky *et al*, 2018). The sRNAs and short DNA guides act as specificity factors that direct AGO effector complexes to their homologous target sequences through complementary base-pairing (Willkomm *et al*, 2015). Eukaryotic AGOs will usually bind duplex sRNAs after being processed by Dicer in the cytoplasm. Then the AGO will preferentially load the strand with the thermodynamically less stable 5' end nucleotide pair (Schwarz *et al*, 2003) and the strand will direct the protein to the target.

miRNA guides form a "seed" interaction with the target involving perfect pairing at nucleotides 2-8 from the 5' end of the sRNA guide. Seed interactions promote target repression and mRNA degradation (Gebert & Macrae, 2019). To enhance target specificity and stability, supplementary nucleotide pairing can occur at nucleotides ~13-16 (Bartel, 2018). Extensive pairing of the target with the 3' end of the miRNA however, can promote end tailing and trimming of the miRNA guide leading to its degradation (Gebert & Macrae, 2019). Targets fully complementary to the miRNA will be cleaved if the AGO protein possesses slicer activity (Gebert & Macrae, 2019). Sometimes, however, mismatches between the target and the 5' end of the miRNA or pairing of the miRNA and target with full complementarity leads to

release/unloading of the miRNA from the AGO (De *et al*, 2013). This is thought to be due to destabilisation of the AGO:guide complex (De *et al*, 2013). Plant and cnidaria miRNAs bind their target with near-perfect complementarity (Carthew & Sontheimer, 2009; Dexheimer & Cochella, 2020). siRNAs bind their targets with perfect complementarity hence usually silence genes by endonucleolytic cleavage of the target (Carthew & Sontheimer, 2009). piRNAs can target mRNAs in various ways including with perfect complementary similarly to siRNAs, but also with a seed pairing similarly to miRNAs (Svendsen & Montgomery, 2018). AGO:guide:target interactions can lead to gene downregulation/silencing or upregulation/activation.

The general term for gene silencing dictated by sRNAs is called RNA interference (RNAi). The importance of RNAi and its putative applications in science as a molecular tool for research and for the development of therapies has been recognised by the award of a Nobel prize in 2006 to Fire and Mello, who discovered RNAi in *C. elegans* (Fire *et al*, 1998). RNAi involves complexes formed by AGO, the sRNA guide and other proteins. RNAi complexes include the RISC (RNA-Induced Silencing Complex) which acts in the cytoplasm, and the nuclear-equivalent form of RISC, the RITS (RNA-Induced Transcriptional Silencing) complex (Pratt & MacRae, 2009). In bilaterian animals, the miRNA-RISC typically targets the 3'UTR of the target mRNA to silence its expression (Bartel, 2009; Dexheimer & Cochella, 2020). 3'UTR regions seem to be the most common region of the mRNAs to be targeted by miRNAs and 3'UTRs in general have long been associated with post-transcriptional gene regulation (Mazumder *et al*, 2003). Interestingly, a recent report from the Corey lab shows that miRNA association with the 3'UTR is a poor predictor of gene repression and calls for experimental validation of the interaction of AGO:miRNA complexes on 3'UTRs (Chu *et al*, 2020). Other regions of the mRNA can also be targeted by RISCs to downregulate gene expression including the coding sequence (CDS) (Tay *et al*, 2008; Fang & Rajewsky, 2011; Hausser *et al*, 2013) and the 5'UTR (Lytle *et al*, 2007; Zhou & Rigoutsos, 2014).

3' UTR and CDS regions are the most studied target locations, probably because the functional outcome of AGO:guide:3'UTR/CDS target region interaction is best understood. However, analysis of other potential target site locations in genes could expand our understanding and shed further insight into the mechanisms behind gene regulation. For example, the concept of sRNAs targeting intronic sites is more accepted in plants (Meng *et al*, 2013; Wang *et al*, 2015) compared to the mammalian

field. Perhaps this is due to the fact that directing AGO:sRNA complexes to intronic target sites would happen in the nucleus and thus attribute AGOs with nuclear functions. The idea that mammalian AGO proteins also operate in the nucleus has been challenged. This is because some studies reporting mammalian AGO localisation and function in the nucleus used an antibody (anti-AGO2 monoclonal 11A9) that was later demonstrated to non-specifically detect the nuclear chromatin remodelling complex SWI/SNF (Van Eijl *et al*, 2017). Nevertheless, various groups have used other molecular tools to examine the putative nuclear function of mammalian AGOs and report that indeed AGOs can target intronic and other regulatory loci too (Huang & Li, 2014). For example, intronic regions have been identified as potential targets of the human AGO1 which has been shown to regulate splicing of the human fibronectin 1 gene by binding intronic loci that are near an alternative exon (Alló *et al*, 2009). Targeting intronic regions promotes establishment of heterochromatin (silencing mark) via histone methylation which is hypothesised to be propagated and slow down elongation by RNA polymerase II, increasing alternative exon inclusion (Alló *et al*, 2009). Human AGO1 has also been shown to bind enhancers and promoters and its depletion results in dysregulation of alternative splicing (Alló *et al*, 2014). miRNA-AGO2 complexes have also been reported to act in the nucleus of human and murine embryonic stem cells and of human cancer cells and target introns (Sarshad *et al*, 2018; Chu *et al*, 2021). Sarshad *et al*, (2018) speculate that targeting intronic regions in precursor mRNAs happens co-transcriptionally and causes gene silencing, while Chu *et al*, (2021) hypothesise that AGO:miRNA complexes could affect alternative splicing by blocking the recognition site of splicing factors. These observations and hypotheses could lead to the discovery of novel mechanisms of gene regulation by AGO complexes in mammals but require further investigation.

The AGO complexes regulate gene expression in various ways. This includes regulation at the epigenetic/transcriptional level or post-transcriptionally (Fig. 1.6). Even though AGO proteins are the key players of the complexes involved in transcriptional and post-transcriptional gene silencing, the mechanism of action of each complex is also defined by the AGO post-translational modifications, protein interactors of AGOs, the guide nucleic acid and the target locus.

1.7.2 Epigenetic/Transcriptional gene regulation

Epigenetic regulation of gene expression is mediated by changes in the configuration of chromatin. Heterochromatin, the closed chromatin configuration, is a transcriptional repressive mark which is established through DNA methylation or histone modifications (Saksouk *et al*, 2015). RNAi can influence heterochromatin formation (Holoch & Moazed, 2015). This has been shown in unicellular eukaryotes (yeast and ciliates), fungi, plants, and animal cells including *C. elegans*, the fruit fly *Drosophila melanogaster* and humans (Holoch & Moazed, 2015).

RNA-directed DNA methylation is deemed unique to plants (Erdmann & Picard, 2020). This biological process was first observed by Wassenegger *et al*, (1994) who showed that after viroids integrated into the genome of tobacco plants, the viroid sequences would gain DNA methylation but not the tobacco genome. It was later shown that methylation was triggered by double-stranded RNA molecules generated during viroid replication which were processed to approximately 23 nt sRNAs (Mette *et al*, 2000). RNA-directed DNA methylation in plants mainly occurs with siRNAs as the guide of the AGO complex, but some miRNAs have also been reported to play a role in this phenomenon (Holoch & Moazed, 2015; Teotia *et al*, 2017). For example, 21-22 nt siRNAs loaded into *Arabidopsis thaliana* AGO6 guide the AGO protein to establish DNA methylation marks on transcriptionally active transposable elements (McCue *et al*, 2015). In rice (*Oryza sativa*), 24 nt long miRNAs such as miR1863 are loaded into AGO4 clade proteins (AGO4a, AGO4b, and AGO16) and direct DNA methylation around the miRNA locus and target binding sites (Wu *et al*, 2010). The mechanism of RNA-directed DNA methylation via miRNAs however, remains unclear (Teotia *et al*, 2017).

Knockout of AGO1, DCR1 (Dicer) and RDP1 (RNA-dependent RNA polymerase) in fission yeast (*Schizosaccharomyces pombe*) showed that these proteins are important for the establishment of H3K9 methylation and processing of non-coding RNAs transcribed from centromeric repeat regions into siRNAs (Reinhart & Bartel, 2002; Volpe *et al*, 2002). Building on these findings, Verdel *et al*, (2004) identified components of the RITS complex in fission yeast, physically linking heterochromatin formation and RNAi after detecting the heterochromatin-associated protein Chp1 co-association with AGO1 and Tas3, a protein required for H3K9 methylation. Moreover, siRNAs are generated from nascent transcripts produced by centromeric repeats involving the complex RDRC (RNA-directed RNA polymerase complex). The

RDRC physically interacts with RITS. This interaction is siRNA-dependent and is mediated by Dicer and the H3 histone methyltransferase, CLR4 (Motamedi *et al*, 2004) Collectively, these findings further highlight that heterochromatin formation is RNAi-dependent and highlights the functional importance of sRNAs as well as all the proteins involved in this pathway.

The relation of RNAi and heterochromatin has been illustrated in some multicellular organisms too (Holoch & Moazed, 2015). Mutations in RNAi components in *D. melanogaster* reduced H3K9 methylation on heterochromatic tandem arrays and transgenes and also affects localisation of the heterochromatin proteins HP1 and HP2 (Pal-Bhadra *et al*, 2004). Moreover, a similar mechanism is seen in *C. elegans* where heterochromatin formation is directed by sRNAs, the AGO protein NRDE-3, NRDE-2 (nuclear RNAi effector) and HPL-2 (HP1-like protein) (Fields & Kennedy, 2019). In mammals, AGO:guide complexes can also affect heterochromatin formation indirectly. For example, cancer research has revealed that some miRNAs are involved in promoting DNA methylation by modulating DNA-methyltransferases and other proteins involved in methylation (Wang *et al*, 2017a). Enforced expression of the miRNA miR-29 family (miR-29a, 29b and 29c) in lung cancer cell lines leads to downregulation of the DNA methyltransferases DNMT3A and DNMT3B (Fabbri *et al*, 2007). In turn, this restores the DNA methylation patterns, allowing expression of tumour-suppressive genes (Fabbri *et al*, 2007). Moreover, the human AGO protein of the PIWI subfamily PIWIL4 in a complex with piRNAs can directly affect heterochromatin formation by recruiting heterochromatin formation factors such as the histone deacetylase HDAC4 to suppress transcription of the exogenous human immunodeficiency virus type 1, HIV-1, in human T-cells (He *et al*, 2020). These findings not only enhance our understanding of the regulation of gene expression but also could be exploited in therapeutic interventions. In this case, silencing of PIWIL4 could be employed as a strategy to permit HIV-1 transcription and hence reverse HIV-1 latency so that HIV-1 reservoirs can be eradicated by maintaining antiretroviral therapies that are already available (Ait-Ammar *et al*, 2020; He *et al*, 2020).

1.7.3 Post-transcriptional gene regulation

Post-transcriptional gene regulation generally occurs in the cytoplasm and involves silencing of targets by translational repression, mRNA deadenylation or mRNA cleavage (Höck & Meister, 2008). AGOs interact with proteins from the GW182 family and this interaction has been associated with formation of the high molecular

weight-RISC (HMW-RISC) complex which is required for suppression of partially complementary targets (Liu *et al*, 2005a; Jakymiw *et al*, 2005). GW182 proteins recruit RNA-processing factors such as the poly(A)-binding protein PABP, and deadenylase complexes such as carbon catabolite repression 4 (CCR4)-negative on TATA-less (NOT) complex, to inhibit translation initiation and trigger transcript de-adenylation and destabilisation (Huntzinger *et al*, 2013). Translational initiation can also be inhibited after the loaded RISC complex causes dissociation of the RNA helicase eIF4A (eukaryotic initiation factor 4A) which is part of the eIF4F complex (Fukaya *et al*, 2014). In this way the assembly of the eIF4F complex responsible for initiation of the mRNA translation is blocked (Fukaya *et al*, 2014). Evidence from studies in *C. elegans* and human cells *in vitro* suggest that translation can also be inhibited by AGO complexes after translation initiation during active mRNA translation by repressing translation elongation (Olsen & Ambros, 1999; Seggerson *et al*, 2002; Nottrott *et al*, 2006; Petersen *et al*, 2006; Friend *et al*, 2012). However, Ricci *et al*, (2013) demonstrate that miRNAs do not repress the elongation stage but rather exclusively affect the early steps of translation. Moreover, cleavage of the mRNA transcript by AGO is another way of silencing gene expression post transcription, as mentioned previously. This occurs if the AGO protein is catalytically active and the guide sRNA is fully complementary to the mRNA target (Liu *et al*, 2004; Meister *et al*, 2004).

1.7.4 RNA activation as another mechanism of gene regulation

A less well studied and understood mechanism for regulation of gene activity which was described in 2006, is RNA activation (RNAa) (Li *et al*, 2006). This is a transcriptional activation mechanism where sRNAs act in a non-canonical way with AGO proteins, leading to upregulation of the target. RNAa has been linked to epigenetic changes, transcriptional initiation stimulation and enhanced transcriptional elongation (Portnoy *et al*, 2016). For example, the expression of E-cadherin and cold-shock domain-containing protein C2 (CSDC2) in human cells was induced following transfection with pre-miR-373 and miR-373, which target the promoter of the genes (Place *et al*, 2008). Transfection with the miRNA mimics led to an increase in the RNA polymerase II detected at the promoters of the E-cadherin and CSDC2 genes (Place *et al*, 2008). By targeting the TATA-box motif in promoters, siRNAs and miRNAs can upregulate gene expression in mammalian cells (Fan *et al*, 2014). RNAa has been described to occur in HIV-1 too, where a viral miRNA binds the TATA-box region to upregulate the activity of the HIV-promoter in turn inducing viral replication (Zhang *et al*, 2014). The *C. elegans* WAGO protein CSR-1 has been attributed with

the ability to induce gene expression (a mechanism also referred to as gene licensing) in the germline (Seth *et al*, 2013; Wedeles *et al*, 2013). Although not yet clear, it has been proposed that CSR-1 could license gene expression by enhancing deposition of active histone modification marks such as H3K36me3 (Wedeles *et al*, 2014; Wedeles, 2018). Additionally, gene expression can also be upregulated by AGO:sRNA complexes by translational activation. Examples of sRNAs involved in RNAa include the mouse miRNA miR10a which associates to the 5'UTR of ribosomal protein transcripts and increases mRNA translation (Ørom *et al*, 2008), and the human miRNA miR-483 which binds to the 5'UTR of the insulin-like growth factor 2 mRNA enhancing translation by increasing association of the mRNA molecule to the RNA-helicase DHX9 (Liu *et al*, 2013).

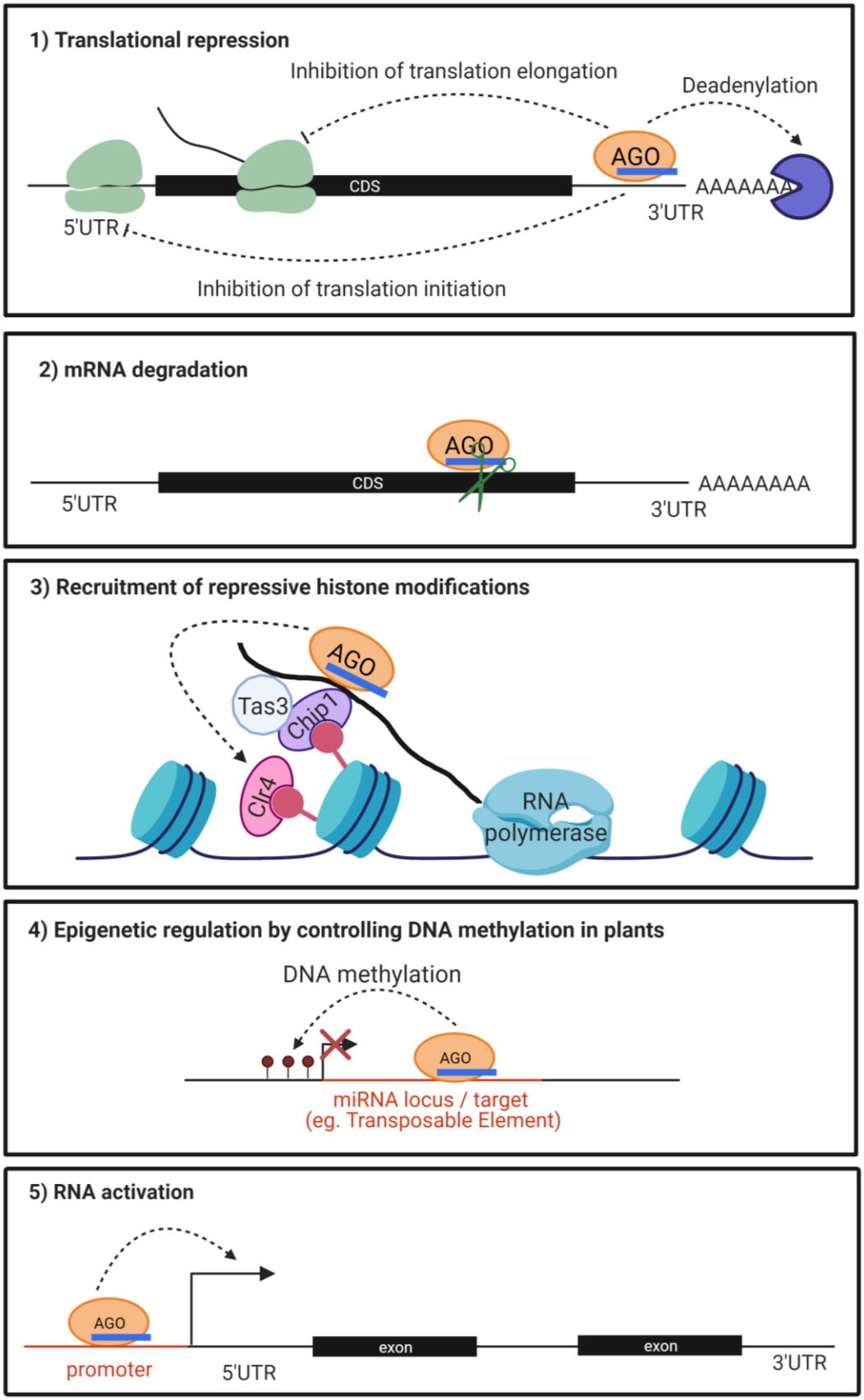


Figure legend on next page.

Figure 1.6 | Mechanisms of gene regulation by Argonaute proteins. AGOs can repress translation (panel 1) by inhibiting translational initiation, elongation and causing mRNA destabilisation by triggering deadenylation. Some AGOs possess slicer activity (shown as scissors) which leads to cleavage and degradation of the transcript (panel 2). Studies in fission yeast revealed that AGOs can silence genes at the epigenetic level by recruiting repressive histone marks (panel 3). Another way to regulate gene expression at the epigenetic level as mediated by siRNAs and sometimes miRNAs is by controlling DNA methylation (panel 4). The example shown in panel 4 is based on studies in plants. AGO proteins can upregulate gene expression by binding the promoter region of genes in a process referred to as RNA activation (panel 5). Figure created using BioRender.com.

1.7.5 Worm-specific AGO (WAGO) proteins

Nematodes encode a worm-specific AGO (WAGO) subfamily that is distinct to the AGO and PIWI clades (Buck & Blaxter, 2013). WAGOs are also known as Secondary AGOs as they bind secondary siRNAs generated by primary siRNAs or piRNAs. The free-living nematode *C. elegans* possesses 27 AGO-like genes (Yigit *et al*, 2006), 17 of which are members of the WAGO clade (Youngman & Claycomb, 2014). However, only 19 of the AGO-like genes encode functional proteins (Charlesworth *et al*, 2021), and 12 of these belong to the WAGO subfamily (Gu *et al*, 2009). In *C. elegans* WAGO proteins associate with secondary siRNAs that are termed “22G” RNAs because they are generally 22 nt in size and start with Guanosine (G), and are 5’PPP (Gu *et al*, 2009). WAGO proteins can be cytoplasmic (eg. WAGO-1/2/3/4/5, SAGO-1, SAGO-2, and PPW-1) or nuclear (eg. WAGO-10/11, HRDE-1, and NRDE-3) and can function in somatic (eg. NRDE-3) or germline cells (eg. HRDE-1) (Billi *et al*, 2014). Although the specific mechanism(s) by which WAGOs regulate gene expression are still being discovered, WAGOs are involved in transcriptional and post-transcriptional gene regulation (Ketting & Cochella, 2020). Apart from the WAGO protein CSR-1, WAGOs do not contain the catalytic motif, and therefore are predicted to not possess slicer activity (Yigit *et al*, 2006; Aoki *et al*, 2007). Some *C. elegans* WAGO proteins including CSR-1, WAGO-1/4, and PPW-1, have also been reported to be secreted in EVs (Nikonorova *et al*, 2021). As *C. elegans* is not a parasitic worm, it is hypothesised that EVs and their cargo (including WAGOs and their associated sRNAs) are involved in communication with other *C. elegans* worms (intraspecies communication), as shown by transfer of fluorescently-labelled male-derived EVs and their cargo to the hermaphrodite uterus during copulation (Nikonorova *et al*, 2021). Interestingly, some

evidence suggests that WAGO proteins could be involved in parasitism. For example, our lab has reported that an extracellular WAGO protein, named “exWAGO”, is secreted by the mouse-infective gastrointestinal nematode *H. bakeri* in EVs (Buck *et al*, 2014; Chow *et al*, 2019). As *H. bakeri* EVs are internalised by mouse intestinal epithelial cells and bone-marrow-derived macrophages (Buck *et al*, 2014; Coakley *et al*, 2017), it is hypothesised that exWAGO could manipulate host gene expression as set out to investigate in this thesis, facilitating parasitism. The idea that WAGO proteins are involved in parasitism is also hypothesised by Hunt *et al*, (2018) who report upregulation of WAGO-like proteins in the parasitic life stage of four *Strongyloides* species compared to the genetically identical free-living stage worms. Further studies and evidence are required to understand whether and how the WAGO proteins and their associated sRNAs might facilitate helminth parasitism.

1.8 sRNAs in helminth ES products

Helminth-derived sRNAs have been detected in host serum, plasma and urine, which can therefore serve as potential biomarkers for prognosis/diagnosis of helminth infections in humans and animals (Cai *et al*, 2016), including the filarial nematodes *Dirofilaria immitis*, *Onchocerca volvulus* (Tritten *et al*, 2014a; Quintana *et al*, 2015), *Loa loa*, *Onchocerca ochengi* (Tritten *et al*, 2014b; Quintana *et al*, 2015), *Litomosoides sigmodontis* (Quintana *et al*, 2019) and the trematodes *Schistosoma japonicum* (Cheng *et al*, 2013) and *Schistosoma mansoni* (Hoy *et al*, 2014). To this day, only sRNAs from tissue-dwelling parasitic helminths but not from gastrointestinal nematodes have been detected in the circulation of the host (Britton *et al*, 2020).

Furthermore, helminth sRNAs are found both encapsulated in EVs (Buck *et al*, 2014; Eichenberger *et al*, 2018a, 2018b; Liu *et al*, 2019a; Taylor *et al*, 2020; Tran *et al*, 2021) but also unencapsulated (Buck *et al*, 2014; Taylor *et al*, 2020). Buck *et al*, (2014) reported the predominant presence of miRNAs in EVs secreted by the gastrointestinal nematode *H. bakeri* and also other sRNAs including full length Y-RNAs (function to regulate various cellular processes including RNA stability, DNA replication, and cellular stress responses (Kowalski & Krude, 2015; Valkov & Das, 2020)), rRNAs and tRNAs. miRNAs are also found outside of vesicles although to a lesser extent compared to vesicular miRNAs, along with rRNAs, tRNAs and fragments of Y-RNAs rather than full length (Buck *et al*, 2014). It is important to note however, that the sRNA libraries generated by Buck *et al*, (2014) capture only 5’P and 5’OH molecules and it was later reported that the most predominant sRNA population inside

EVs is in fact 5'PPP secondary siRNAs (Chow *et al*, 2019). The EVs of the gastrointestinal nematodes *N. brasiliensis* and *T. muris* (Eichenberger *et al*, 2018a, 2018b), and the parasitic flatworms *S. japonicum* (Liu *et al*, 2019a) and *F. hepatica* (Tran *et al*, 2021) also contain miRNAs. The nematode *Trichinella spiralis* secretes miRNAs inside EVs at the adult stage but at the muscle larval stage it secretes miRNAs that are unencapsulated (Taylor *et al*, 2020). All these studies show that some of the helminth miRNAs that are secreted share a part of their sequence known as the "seed" region (nucleotides 2-8 from the 5' end of the miRNA) with the host miRNAs. These findings led to the hypothesis that helminth miRNAs could affect the expression of host genes normally regulated by host miRNAs. Examples of host genes predicted to be affected by helminth miRNAs include genes involved in inflammation (Buck *et al*, 2014; Eichenberger *et al*, 2018a, 2018b) and immunity (Buck *et al*, 2014; Eichenberger *et al*, 2018a, 2018b; Liu *et al*, 2019a; Luo *et al*, 2018).

Indeed, some of the predicted host gene targets are silenced by helminth miRNAs, as validated *in vitro*. For example, the expression of *Dusp1*, a gene encoding a phosphatase involved in upregulation of the immunosuppressive cytokine IL-10 (Hammer *et al*, 2006), is repressed by miRNA mimics of *H. bakeri* miRNAs let-7, miR-200 and miR-425 in a reporter assay (Buck *et al*, 2014). Moreover, using a reporter assay again, transfection of a miRNA mimic of the vesicular *S. japonicum* miRNA miR-125b led to suppression of the *Pros1* and *F11r* genes expressing proteins involved in inhibition of inflammatory responses and regulating formation of tight junctions in epithelial cells respectively (Liu *et al*, 2019a). However, it is important to understand the mechanism by which these parasite sRNAs would silence host genes.

It is assumed that the parasite miRNAs will achieve manipulation of host genes by associating with host AGO proteins since sRNAs require to be bound by AGOs to function. Following *in vitro* treatment of a murine macrophage cell line (RAW264.7 cells) with EVs derived from *S. japonicum*, Liu *et al*, (2019) immunopurified the mouse host AGO2 (mAGO2) protein and detected *S. japonicum*-derived miRNAs by Reverse-Transcription quantitative PCR (RT-qPCR). Immunoprecipitation of the mouse AGO2 (mAGO2) protein from peritoneal macrophages from mice infected with *F. hepatica* and uninfected mice as negative control reveals that mAGO2 is bound to parasite miRNAs (Tran *et al*, 2021). These examples illustrate that helminth sRNAs can associate with host AGO proteins, and therefore this could allow parasite sRNAs to manipulate host gene expression. Hijacking of the host AGO machinery complex

has been previously described in other organisms including plants, fungi, and bacteria (Cai *et al*, 2019). This phenomenon is called cross-species/kingdom RNAi.

1.9 Cross-species RNA interference

Cross-species RNAi involves the export of sRNAs from one organism and internalisation of these sRNAs by an organism of a different species. Exported RNA (i.e. extracellular RNA/exRNA) has been implicated in biomarker discovery, cancer therapeutics and has recently been recognised as an important mediator of cell-to-cell crosstalk (Li *et al*, 2018; Sadik *et al*, 2018). Several reports describe a functional role for exRNA and some implicate exRNA in the regulation of the immune system (De Candia *et al*, 2016). For example, by generating *Dicer* knockout T-regulatory cells (abolishing in this way generation of miRNAs) and transfecting the miRNA mimic let-7d into Th1 cells, Okoye *et al*, (2014) showed that let-7d is transported from T-regulatory cells to Th1 cells via EVs leading to repression of Th1 cell proliferation and expression of *IFN-γ*. exRNA can regulate gene expression via cross-species RNAi, as internalised exRNA can hijack the AGO protein of the recipient cell to induce gene silencing (Fig. 1.7).

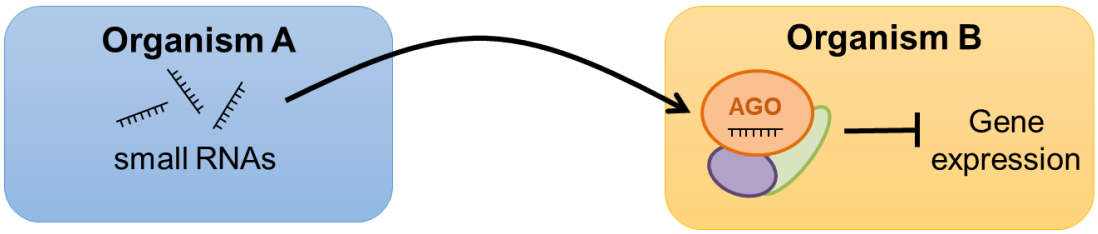


Figure 1.7 | One form of cross-species RNAi. Small RNAs from one organism are exported and internalised by an organism of a different species. These small RNAs can hijack the RISC of the recipient organism by associating with the AGO protein. This can lead to dysregulation of gene expression inside the host.

One of the first examples of cross-species RNAi was reported by the Jin lab, illustrating that sRNAs secreted by *Botrytis cinerea*, a fungal pathogen of the plant *Arabidopsis thaliana*, hijack the host RNAi pathway by binding to *Arabidopsis* Argonaute 1 (AGO1) (Weiberg *et al*, 2013). This leads to silencing of host genes involved in immunity, enabling *B. cinerea* to establish infection (Weiberg *et al*, 2013). Since then, several studies demonstrate that parasite-derived sRNAs can control the host AGO and alter gene activity to benefit the parasite (Cai *et al*, 2019). Cross-species RNAi does not only function in a parasite-to-host direction, but also in

a host-to-parasite direction (Cai *et al*, 2019). In this case, miRNAs, siRNAs and phasiRNAs secreted by the host *Arabidopsis* are delivered into *B. cinerea* cells mediating suppression of genes implicated in fungal pathogenicity (Cai *et al*, 2018).

Intriguingly, organisms do not only employ their sRNA molecules to manipulate gene expression in the recipient organism, but can also deliver their RNAi machinery in the recipient. This was first described by Garcia-Silva *et al*, (2014), who reported that the parasites *Trypanosoma cruzi* which lack canonical sRNA pathways secrete an AGO protein, TcPIWI-tryp, in EVs. These EVs can be transferred to susceptible green monkey (Vero) cells and to other *T. cruzi in vitro*. However, the function of TcPIWI-tryp and its putative function in host-pathogen and parasite-parasite communication remains elusive (Garcia-Silva *et al*, 2014). Moreover, Wang *et al*, (2017b) identified that EVs secreted from malaria-infected human red blood cells (RBCs) carry AGO2 loaded with miRNAs. The AGO2-miRNA complexes are delivered to the malaria-causing pathogen *Plasmodium falciparum* within infected RBCs as shown by immunofluorescence experiments and a pull-down of the human AGO2 from *P. falciparum* parasites followed by sRNA sequencing (Wang *et al*, 2017b). The AGO2-miRNA miR451/140 complexes target the malaria antigen PfEMP1 (Wang *et al*, 2017b), which causes rosetting of infected RBCs with uninfected RBCs (Juillerat *et al*, 2011). It is worth noting though that the results by Wang *et al*, (2017b) are contradicted by Hentzschel *et al*, (2020) who demonstrated that immunofluorescence detection of human AGO2 in parasites is explained by antibody cross-reactivity with parasite molecules. Along with other data presented, the hypothesis that human AGO2 is transported from the host to the parasite is not supported (Hentzschel *et al*, 2020).

1.10 *H. bakeri* secretes an AGO protein – links to cross-species RNAi?

Our group has identified an AGO protein released by the gastrointestinal nematode *H. bakeri* in the excretory/secretory (HES) products (Buck *et al*, 2014; Chow *et al*, 2019). This AGO protein is 101.7 kDa, belongs to the worm-specific AGO (WAGO) clade, and we named it “exWAGO” (ex; extracellular) (Buck *et al*, 2014; Chow *et al*, 2019). exWAGO is the most highly expressed AGO within *H. bakeri* out of the 9 AGO genes it is thought to express, and the only AGO protein found in HES based on mass spectrometry analysis (Buck *et al*, 2014; Chow *et al*, 2019). Phylogenetic, RNA sequencing and RT-qPCR analyses demonstrate that the orthologues of exWAGO are highly conserved and abundantly expressed in Clade V gastrointestinal

nematodes, such as the human hookworm *N. americanus*, and the sheep parasites *T. circumcincta* and *H. contortus* (Chow *et al*, 2019). The percentage of amino acid identity between exWAGO and its orthologues from Clade V parasitic nematodes is approximately 70-80% (Table 1.1). The exWAGO orthologues of the Clade V free-living nematode *C. elegans* are the WAGO proteins SAGO-1, SAGO-2 and PPW-1 (Chow *et al*, 2019), with amino acid identity around 30% (Table 1.1). SAGO-1, SAGO-2 and PPW-1 were the first to be discovered to act as Secondary AGO proteins binding 5'PPP sRNAs based on environmental RNAi (Yigit *et al*, 2006). The function(s) of the *C. elegans* exWAGO orthologues remain largely unknown but data from the Claycomb lab suggests that SAGO-2 and PPW-1 might play a role in regulating genes involved in immune responses (Seroussi, unpublished). Parasitic nematodes from Clade III also express exWAGO homologues. The percentage of amino acid identity between exWAGO and its homologues from Clade III parasitic nematodes is around 30-40% (Table 1.1).

exWAGO is found in both the vesicular and the non-vesicular fraction of the total excretory/secretory products of *H. bakeri* (HES) (Buck *et al*, 2014). Liquid Chromatography tandem Mass Spectrometry (LC-MS/MS) and western blot analyses by protein weight suggests exWAGO is enriched in *H. bakeri* EVs compared to the non-vesicular fraction (Buck *et al*, 2014; Chow *et al*, 2019, Supp. Fig. 1). Unpublished data from our lab generated by Dr. Franklin Chow, however, indicate that when analysed by volume, exWAGO is enriched in the non-vesicular fraction of HES (Chow, unpublished). These results suggest that exWAGO is mainly found unencapsulated but that it is not the most abundant protein found in the non-vesicular fraction. What dictates the encapsulation of exWAGO in EVs or how the two extracellular (vesicular and non-vesicular) forms of exWAGO differ remains undetermined.

The presence of exWAGO and extracellular sRNAs in the ES products of *H. bakeri* (Buck *et al*, 2014; Chow *et al*, 2019) suggests exWAGO might be associated with these sRNAs and this raises the question of whether exWAGO is involved in driving the export of sRNAs (Chow *et al*, 2019). The fact that *H. bakeri* EVs can be internalised by mouse intestinal cells and bone-marrow derived macrophages *in vitro* (Buck *et al*, 2014; Coakley *et al*, 2017) suggests that exWAGO and the sRNA cargo may modulate host gene expression through cross-species RNAi (Buck *et al*, 2014; Coakley *et al*, 2015). We hypothesise that exWAGO might be involved in dysregulation of host gene expression directly, perhaps in combination with parasite

sRNAs hijacking the host AGO proteins. exWAGO could also facilitate communication between *H. bakeri* worms (intraspecies communication) and/or between worms or parasites of other species, and/or the gut microbiome too (interspecies communication). Hence, we want to understand whether exWAGO is involved in gene silencing and particularly whether it can manipulate the activity of mouse host genes.

Table 1.1 | The percentage amino acid identity of the exWAGO homologues in parasitic nematodes of clade III and V, and *C. elegans*. The amino acid identity was found using the alignment tool of the protein BLAST analysis webtool (https://blast.ncbi.nlm.nih.gov/). The protein sequences can be found in Appendix A.

Clade	Species	Gene name	% amino acid identity	Host(s)
V	*Caenorhabditis elegans*	SAGO-1	28	n/a
V	*Caenorhabditis elegans*	SAGO-2	30	n/a
V	*Caenorhabditis elegans*	PPW-1	30	n/a
V	*Ancylostoma caninum*	ACAN_exon_1	79	Dogs, cats and humans
V	*Ancylostoma ceylonicum*	ACEY_exon_1	79	Humans
V	*Ancylostoma duodenale*	ancylostoma_duodenale.ANCDUO_12582	74	Humans
V	*Angiostrongylus cantonensis*	ACANT_exon	70	Rats and humans
V	*Angiostrongylus costaricensis*	ACOST_exon_2	69	Rats and humans
V	*Dictyocaulus viviparus*	DVIP_exon_fixed	69	Cattle
V	*Haemonchus contortus*	H.contortus_HCOI02120600.t1	76	Sheep and goats
V	*Necator americanus*	N.americanus_NAME_model	77	Humans
V	*Nippostrongylus brasiliensis*	N. brasiliensis.NBR_exon	78	Rats
V	*Oesophagostomum dentatum*	ODENT_exon	77	Ruminants, pigs, monkeys and humans
V	*Teladorsagia circumcincta*	Tcirc.Tcirc_exon_trim	81	Sheep and goats
III	*Acanthocheilonema viteae*	acanthocheilonema_viteae.nAv.1.0.1.t07849RA	34	Rodents
III	*Anisakis simplex*	anisakis_simplex.ASIM_0001785301mRNA1	29	Marine fish, shellfish and humans
III	*Ascaris lumbricoides*	ascaris_lumbricoides.ALUE_0001385501mRNA1	37	Humans
III	*Ascaris suum*	ascaris_suum.ASU_04951	37	Pigs
III	*Brugia malayi*	Brugia_malayi.Bm4557	35	Humans
III	*Dirofilaria immitis*	Dirofilaria_immitis.nDi.2.2.2.t06656	34	Dogs, cats and ferrets
III	*Enterobius vermicularis*	enterobius_vermicularis.EVEC_0000434401mRNA1	33	Humans
III	*Litomosoides sigmodontis*	Litomosoides_sigmodontis.nLs.2.1.2.t06959RA	35	Rodents
III	*Loa loa*	Loa_loa.EN70_11666	34	Humans
III	*Onchocerca flexuosa*	onchocerca_flexuosa.OFLC_0000331401mRNA1	34	Deer
III	*Onchocerca ochengi*	onchocerca_ochengi.nOo.2.0.1.t05983RA	34	Cattle
III	*Onchocerca volvulus*	Onchocerca_volvulus.OVOC1650	33	Humans
III	*Symphacia muris*	syphacia_muris.SMUV_0000813301mRNA1	30	Rats
III	*Thelazia callipaeda*	thelazia_callipaeda.TCLT_0000080401mRNA1	36	Humans, dogs and cats
III	*Wuchereria bancrofti*	wuchereria_bancrofti.WBA_0000646201-mRNA-1	32	Humans

1.11 PhD Aim & Objectives

The main aim of this project is to elucidate the function of exWAGO. As *H. bakeri* is a genetically non-tractable organism (Lendner *et al*, 2008), there are limitations in our ability to test the function of exWAGO inside the parasite. Hence, this work mainly focuses on functionally characterising the exWAGO that is found in HES and in the host cells/tissues. In particular, we wish to understand whether exWAGO directly mediates the silencing of host genes and the mechanism by which it does this. As a component of this aim, we further wish to understand whether exWAGO represents a new vaccine candidate. The specific objectives are to:

Objective 1: Identify the small guide RNAs associated with exWAGO

This will allow us to determine the sRNA sequences to which exWAGO binds and will indicate whether exWAGO has specificity for particular types of guides.

Objective 2: Identify and validate the targets exWAGO is directed to

This will allow us to identify candidate genes whose expression might be downregulated by exWAGO and build a model for the recognition criteria of targets.

Objective 3: Examine the potential cleavage activity of exWAGO and determine the host proteins it interacts with

Testing whether exWAGO possesses cleavage activity and identification of host proteins with which exWAGO associates will help inform us as to whether exWAGO acts in a canonical way (i.e. involved in gene silencing via mRNA cleavage/destabilisation) or direct us as to how else it might mediate gene regulation.

By performing various functional assays *in vitro* & *in vivo* along with the data gathered in Objectives 1, 2 and 3, we will build a model for the mode of action of exWAGO.

Objective 4: Evaluate whether exWAGO can be a vaccine candidate

This will allow us to determine if exWAGO could be a vaccine against *H. bakeri* infection and whether it could be used as a vaccine candidate for other parasitic helminths secreting exWAGO homologues. This objective will also illustrate the importance of exWAGO in infection.

Chapter 2: Material and Methods

2.1 Cell culture

Immortalised intestinal epithelial (MODE-K) cells from the small intestine of C3H/He female mice were cultured at 37°C with 5% CO_2 in DMEM medium supplemented with fetal bovine serum (10%), L-glutamine (1%), penicillin-streptomycin (1%), sodium pyruvate (1%) and non-essential amino acids (1%) (Vidal *et al*, 1993). MODE-K cells were kindly provided by Dominique Kaiserlian (French Institute of Health and Medical Research).

ExWAGO-expressing MODE-K cells were previously generated by Dr Franklin Chow and Mr. Jonathan Wild. Briefly, FLAG-tagged exWAGO generated by overlap extension PCR was cloned into pCDNA 3.1 (+) plasmid using HindIII and Xbal restriction enzyme digestion. The plasmid was transformed in DH5α *Escherichia coli* cells and single colonies were analysed by colony PCR. The plasmid was purified from successfully transformed colonies and then transfected into MODE-K cells. Cells with successfully transfected plasmids were selected using blasticidin. High expressing single clones were selected and the highest exWAGO-expressing clone was used here (based on western blot analysis of cell lysates). ExWAGO-expressing MODE-K cells were cultured as described above with 5 µg/ml blasticidin.

2.2 Animals used

All the mice used in this project were bred by the in-house facilities at The University of Edinburgh. Experimental procedures were executed under a UK Home Office licence as approved by qualified veterinarians and were carried out by qualified personnel in accordance with the UK Home Office guidelines.

2.3 *H. bakeri* life cycle, collection of HES & mouse gut tissue

CBA × C57BL/6 F1 male mice (referred to as F1 in the thesis) were infected with 400 *H. bakeri* L3 stage larvae by oral gavage and adult worms were cultivated as described in Johnston *et al*, (2015). Briefly, 14 days post-infection with L3 larvae mice were culled and adult worms were harvested from the mouse gut. The adult worms were extensively washed to remove any mouse tissue and kept in culture medium (RPMI1640 supplemented with 1.2% glucose, 5 U/ml penicillin, 5 µg/ml streptomycin, 2 mM L-glutamine and 1% gentamycin) where they release their excretory/secretory products known as “HES” (Johnston *et al*, 2015). HES-containing culture medium was collected as follows: The culture medium 1 day after worms were harvested was

discarded and replaced. Thereafter, the conditioned media (HES) was collected every 3-4 days for a total of two weeks. The HES collected was spun (1,200 rpm, 5 min, room temperature) to remove eggs, filtered (0.22 µm), and then stored at -20°C until required.

Gut tissue from *H. bakeri*-infected or uninfected F1 mice was used for immunoprecipitations. The gut was removed from the mice, straightened and the proximal ~6 cm of the duodenum (top part of small intestine) were cut into ~2 cm pieces (remained unflushed), snap frozen in liquid nitrogen and then stored at -80°C until required. The time between culling mice and harvesting the gut tissue was kept to a minimum (~5-10 mins). The whole procedure was performed as quickly as possible to limit RNA and protein degradation in the tissue.

2.4 EV purification and collection of EV-depleted HES

In these experiments HES was collected from day 1-4 post-harvest (termed “day 4”) and day 4-8 post-harvest (termed “day 8”). HES collected 24 hours post-harvest (termed “day 0” is discarded). The HES (~50-60 ml per batch of HES; batch is defined as the HES collected per mouse harvest) is then concentrated to <12.5 ml using a 5 kDa Vivaspin (4,000 rpm, 4°C) (GE Healthcare, 28-9323-59). Concentrated HES was then subject to ultracentrifugation in Beckman polyallomer centrifuge tubes (Beckman, 331374) and spun at 100,000 rcf, 70 min, 4°C using a Beckman ultracentrifuge with a SW40 swing out rotor. The supernatant which is called “EV-depleted HES”, was collected and stored at -80°C until required. The EV pellet was washed twice with 12.5 ml of cold PBS and spun as before. PBS washes were discarded, and the EV pellet was resuspended in 120-150 µl PBS. The concentration of EVs was measured using the Qubit Protein Assay kit (Thermo, Q33211). The EV-depleted HES that was collected was concentrated to a final volume of 0.5-1.0 ml after buffer exchange in > 40 ml of PBS using a 5 kDa Vivaspin (4,000 rpm, 4°C). The protein concentration was measured using the Qubit Protein Assay kit and the concentrated material was stored at -80°C until required.

2.5 Recombinant mAGO2 and exWAGO proteins used

The recombinant mouse AGO2 (mAGO2) used in gel shift assays (Section 2.11) was purchased from Sino Biological (catalogue number: 50683-M07B). This mAGO2 recombinant protein is 99 kDa and carries a His-tag at the N-terminus.

The recombinant exWAGO used for immunisation of mice (Section 2.20), for gel shift assays (Section 2.11) and for western blot analyses, was produced for the Buck Lab by Sino Biological. The exWAGO recombinant protein is 107.3 kDa and was designed to have 3x FLAG tags, a PreScission cleavage site, followed by a 6x His tag at the N-terminus. The protein sequence of the recombinant exWAGO is shown below with the FLAG tags marked in (dark)yellow, the PreScission Cleavage marked in blue, His tag marked in (dark)green and the exWAGO sequence marked in grey:

MDYKDDDDKDYKDDDDKDYKDDDDKALLEVLFQGPASGHHHHHHSGGGGSMD
QLKTGMGQLSVGAVALPEKRSPGGIGNKVDFVTNLTELSLKPNVPYYKYDIRMYIV
YKGNDALEHLKELTKQTKDDFPEQERKSAAVAVYKHLCKTYKDVFLPDGALLYDR
AAVLFSAQRQLKLDGEEKQFMLPASVVSSAGPDATGIRVVIKKVKDQFQVTSNDLS
KAVNVRDMERDKGILEVLNLAVSQKGYMETSQFVTYGSGVHYLFDHRALGFRDNE
LPELMDGKYMGIGLTKSVKVLEGDSGKGNSAFVVTDVTKGAFHVDEQNLMEKISQ
MSIFFDQRTGQSSFNAKNAMQPFNQKAILQQIKGLYVRTTYGKKKTFPIGNLAAAA
NALKFQTADGAQCTVEQYFKKHYNIQLKYPGMFTVSERHNPHTYYPVELLTVAPS
QRVTLQQQTPDQVASMIKASATLPQTRLHQTKIMKDALDITPRNHNLATAGISVAN
GFTAVSGRVLPSPRIAYGGNQILRPVDNCKWNGDRSVFLEPAKLTNWAVCVTLTQ
QDARRLQIKEYISRVEMRCRNRGMQVDPVAEVFTLKHQTFDGLKEWYASQKQKN
RRYLMFITSDGIKQHDSIKLLEVEYQIVSQEIKGSKVDAVVTKNQNQTLDNVVAKIN
MKLGGVNYNVMLGVKNDDKAFSWLNDKDRMFVGFEISNPPALSKVEIERGASYK
MPSVLGWGANCAGNHQQYIGDYVYIQPRQSDMMGAKLSELIVDILKRFRAATTIAP
RHIVLYFSGISEGQFSLVTDTYMRAVNTGIASLSPNYKPSVTAVAVSKDHNERIYKT
NISGNRATEQNIPPGTVIDTKIVSPVINEFYLNSHSAFQGTAKTPKYSLLADNSKIPLD
VIEGMTHGLCYLHEIVTSTVSVPVPLIVADRCAKRGHNVYIANSNQGEHSVNTIDEA
NAKLVNDGDLKKVRYNA

2.6 Antibodies used

The antibodies used in this thesis for immunoprecipitation, western blot analyses and immunofluorescence are detailed in Table 2.1.

Table 2.1 | List of antibodies used for immunoprecipitation, western blotting, and immunofluorescence analyses.

Used for immunoprecipitation			
Antibody information	**Manufacturer**		**Catalogue no.**
Rat anti-exWAGO antiserum 1 & 3	Produced in-house		
Naïve rat serum	Produced in-house		
Mouse anti-mAGO2	Provided by Dónal O'Carroll, unpublished		
Mouse IgG	Sigma Aldrich		I5381
Mouse anti-Dab2/p96	BD Transduction Laboratories		610464
Rabbit anti-Upab2l	ThermoFisher Scientific		PA5-36998
Used for western blot analysis			
Antibody information	**Dilution & Buffer**	**Manufacturer**	**Catalogue no.**
Mouse anti-FLAG	1:4,000 in 3% milk/TBST	Sigma Aldrich	F3165
Rabbit anti-exWAGO	1:2,000 in 5% BSA/TBST	Eurogentec for Buck lab (Chow *et al* , 2019)	
Rabbit anti-SFPQ	1:1,500 in 3% milk/TBST	Abcam	ab38148
Mouse anti-mAGO2	1:4,000 in 3% milk/TBST	Provided by Dónal O'Carroll, unpublished	
Mouse anti-Dab2/p96	1:2,000 in 3% milk/TBST	BD Transduction Laboratories	610464
Rabbit anti-Upab2l	1:1,000 in 3% milk/TBST	ThermoFisher Scientific	PA5-36998
Goat anti-rabbit IgG (H + L) Dylight 800	1:10,000 in appropriate buffer	ThermoFisher Scientific	SA5-35571
Goat anti-mouse IgG (H + L) AlexaFluor 680	1:10,000 in appropriate buffer	Invitrogen	A-21058
Used for immunofluorescence			
Antibody information	**Dilution**	**Manufacturer**	**Catalogue no.**
Epoxy cleaned rat anti-exWAGO antiserum 2	1:50	Produced in-house	
Rabbit anti-SFPQ	1:100	Abcam	ab38148
Rat anti-exWAGO antiserum 3	1:500	Produced in-house	
Naïve rat serum	1:500	Produced in-house	
Goat anti-rabbit IgG (H + L) AlexaFluor 546	1:2,000	Invitrogen	A-11010
Goat anti-rat IgG (H + L) AlexaFluor 488	1:2,000	Invitrogen	A-11006
Goat anti-rat IgG (H + L) AlexaFluor 647	1:500	Invitrogen	A-21247

2.7 Immunoprecipitation (IP)

To immunopurify exWAGO several starting materials were used including: (1) MODE-K cells expressing recombinant FLAG-tagged exWAGO, (2) *H. bakeri*-infected mouse gut tissue, (3) adult worms, (4) EVs, and (5) EV-depleted HES. The parental MODE-K cell line, naïve rat serum or gut from naïve F1 mice were used as negative controls in IPs to eliminate non-specific protein interactions. mAGO2, Dab2 and Ubap2l were immunopurified only from mouse gut tissue. The specific samples used are detailed in each chapter, depending on the aim of the experiment, for example to identify the sRNA guides, proteins or mRNA targets with which exWAGO associates.

2.7.1 Preparation of lysate

The general IP protocol used for the aforementioned starting materials is the same, with the exception of the FLAG-tagged exWAGO-expressing MODE-K cell line. The major difference in the general IP protocol is the way each material was lysed and the buffer used for lysis of adult worms. For exWAGO IPs from MODE-K cells, several lysis buffers were used depending on the beads used. For IPs with Protein G beads (Thermo, 10003D) cells were lysed in Protein G lysis buffer (50 mM Tris.HCl pH 7.5, 300 mM NaCl, 1% IGEPAL CA-630, 5 mM EDTA, 10% glycerol; filtered with 0.22 µm

Millex-GP filter) containing protease and phosphatase inhibitors (1 tablet per 10 ml) (Roche,11873580001 and 04906845001) for 20 min on ice. For IPs with anti-FLAG M2 beads (Sigma, M8823) cells were lysed in anti-FLAG M2 lysis buffer (50 mM Tris.HCl pH 7.5, 300 mM NaCl, 1% Triton X-100, 5 mM EDTA, 10% glycerol; filtered with 0.22 µm Millex-GP filter) containing protease and phosphatase inhibitors (1 tablet per 10 ml) (Roche,11873580001 and 04906845001) for 20 min on ice or the cells were lysed using the van Nues lysis buffer (50 mM Tris.HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 0.1% NP40, 0.5% Triton X-100; filtered with 0.22 µm Millex-GP filter) containing protease and phosphatase inhibitors (1 tablet per 10 ml) (Roche,11873580001 and 04906845001) for 20 min on ice. Gut pieces were crushed using a pestle and mortar under liquid nitrogen and lysed in Protein G lysis buffer containing protease and phosphatase inhibitors (1 tablet per 5 ml) (Roche,11873580001 and 04906845001) and 200 U/ml RNasin (Promega; N2515). Cell/Tissue debris were removed by centrifugation (16.1K rcf, 20 min, 4°C). To generate adult worm lysate, adult worms were washed in PBS and then lysed by the Tissue Lyser II (Qiagen) in pre-cooled cartridges for 2 min at 30 Hz twice, in worm lysis buffer (150 mM NaCl, 10 mM Tris.HCl, 0.5 mM EDTA, 0.5% NP40 filtered using a 0.22 µm Millex-GP filter) containing protease and phosphatase inhibitors (1 tablet per 5 ml) (Roche,11873580001 and 04906845001) and 200 U/ml RNasin (Promega; N2515). Unlysed material was removed by centrifugation (16.1K rcf, 10 min, 4°C) and the pellet was discarded. The supernatant (lysate) generated from adult worms was used for protein quantification using the Qubit Protein Assay kit (Thermo, Q33211). The samples were stored in -80°C until required. For exWAGO IPs from EVs or EV-depleted HES the material was lysed in TBS/0.05% Triton X-100 with a protease inhibitor (1 tablet per 10 ml; Roche,11873580001). For the exWAGO IP from EV-depleted HES where sRNAs were then extracted and subjected to RT-PCR analysis as published in Chow *et al*, (2019) (Section 2.9), detergent was omitted from the sample to avoid lysis of any EVs that were not removed during ultracentrifugation.

2.7.2 Antibody conjugation

To immunopurify exWAGO, mAGO2, Dab2 or Ubap2l, the appropriate antibody (Section 2.6) was conjugated to magnetic beads. For IPs using adult worm lysate, EVs, EV-depleted HES, gut tissue, and exWAGO-expressing MODE-K or parental cells, Protein G beads (Thermo, 10003D) were used. Protein L beads (Thermo, 88850) were also used for the vesicular exWAGO IP that was then subjected to RT-qPCR analysis as published in Chow *et al*, (2019). For FLAG IPs using

exWAGO-expressing MODE-K or parental cells anti-FLAG M2 magnetic beads (Sigma, M8823) were used. The type and amount of beads as well as the amount of antibody used per IP are detailed in the Results section.

The Protein G, Protein L, and anti-FLAG M2 beads were washed five times with Binding Wash Buffer (PBS, 0.02% Tween 20; filtered with 0.22 µm Millex-GP filter). The antibodies were conjugated on the beads by incubation of the antibody in Binding Wash Buffer (2 h, 4°C, rotating wheel). Unconjugated antibody was then removed, and the beads were equilibrated with the appropriate lysis buffer three times. For IPs using anti-FLAG M2 beads no antibody was required to be conjugated on the beads as the beads were already conjugated with anti-FLAG antibodies. When executing the van Nues protocol with anti-FLAG M2 beads, the beads were equilibrated in the van Nues lysis buffer three times.

2.7.3 Sample addition, immunoprecipitation, and elution

The (cell/tissue/worm/EV/EV-depleted HES) lysate was incubated 45 min or 4 h or overnight (4°C, rotating wheel) as stated in the Results section. After incubation with the sample, the unbound fraction was removed and stored -80°C until use. With exception of the van Nues protocol, the beads were washed with Low Salt buffer (50 mM Tris.HCl pH 7.5, 300 mM NaCl, 5 mM $MgCl_2$, 0.5% Triton X-100 and 2.5% glycerol; filtered with 0.22 µm Millex-GP filter), followed by two washes with High Salt buffer (50 mM Tris.HCl pH 7.5, 800 mM NaCl, 10 mM $MgCl_2$, 0.5% Triton X-100 and 2.5% glycerol; filtered with 0.22 µm Millex-GP filter) (5 min, 4°C, rotating wheel). The beads were washed once more with Low Salt buffer, followed by a PNK buffer wash (50 mM Tris.HCl pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 0.5% Triton X-100; filtered with 0.22 µm Millex-GP filter). For mass spectrometry analysis and for validation of Dab2 and Ubap2l as protein interactors, the samples were washed three times with Low Salt buffer only (tubes were tipped over 30x during each wash). For the van Nues protocol, the anti-FLAG M2 beads were washed three times (tubes were tipped over 30x during each wash) with the van Nues lysis buffer. All the wash buffers were used cold and stored at 4°C.

For western blot analysis, proteins were eluted from the beads in 4X NuPAGE LDS sample buffer (Thermo, NP0008) with 0.1 M DTT (10 min, 70°C, 1,000 rpm). For RNA analysis, the beads were incubated with Qiazol (700 µl, 5 min, room temperature) (Qiagen, Q79306). The eluate was stored at -80°C until use. For mass spectrometry analysis, the beads were washed once with Ultrapure water, and the beads were then

frozen at -80°C until required. The specific details of each IP are described in the Results chapters.

2.8 Western Blot analysis

Protein samples were reduced in 4X NuPAGE LDS sample buffer (Thermo, NP0008) with 0.1 M DTT (10 min, 70°C, 1,000 rpm) and separated on 4-12% Bis-Tris NuPAGE SDS gels (Thermo) in 1X NuPAGE MOPS SDS running buffer (Thermo, NP000102) (180V for 85 min or 120V for 150 min). 1-4 µl of the Precision Plus Protein All Blue Prestained Protein Standards ladder (Bio-Rad, 1610373) was used as a size marker. The separated proteins were then transferred on an Immobilon-FL PVDF membrane (Millipore, IPFL00010) pre-activated in 100% methanol. Wet transfer was performed in 1X NuPAGE Transfer buffer (Thermo, NP00061) with 10% methanol (100V for 105 min). The membrane was then blocked in 3% milk in TBS/T (1xTBS with 0.1% Tween 20) or 5% Bovine Serum Albumin (BSA) in TBS/T (1 h 45 min, on shaker) depending on the antibody used (Table 2.1). The blocked membrane was incubated with the primary antibody (overnight, 4°C, on shaker). Following four TBS/T washes (15 min, on shaker) the membrane was incubated with a fluorescently-labelled secondary antibody (1h, in dark, on shaker). After further four TBS/T washes (15 min, on shaker) and a TBS wash (5 min, on shaker), the membrane was scanned at 700 and/or 800 nm using the Odyssey CLx infrared imaging system (LI-COR).

2.9 Reverse Transcription-quantitative PCR (RT-qPCR)

For the detection of *H. bakeri* sRNAs by RT-qPCR, RNA was extracted using the miRNeasy Serum/Plasma kit (Qiagen, 217184) according to the manufacturer's instructions. RNA was eluted from the beads in 700 µl of Qiazol (Qiagen, Q79306). Prior to the extraction the samples were spiked with RT4 synthetic RNA at a final concentration of 0.1 pM and the RNA was eluted in 14 µl RNase-free water. During development and optimisation of the protocol for exWAGO IP from mouse gut tissue, the RNA extraction efficiency of the miRNeasy Serum/Plasma kit (Qiagen, 217184) and the Direct-zol RNA MiniPrep kit (Zymo Research, R2050) was examined. RNA was eluted from the beads in 700 µl of Qiazol (Qiagen, Q79306) and after the RNA was extracted, it was eluted in 25 µl RNase-free water. The Direct-zol RNA MiniPrep kit (Zymo Research, R2050) was used according to the manufacturer's instructions (including the DNase treatment) and the RNA was eluted in 25 µl RNase-free water. After RNA was extracted, the samples were spiked with synthetic cel-miR-39-3p at a final concentration of 0.1 pM at the stage of cDNA. Reverse Transcription and

quantitative PCR were performed as described in Chow *et al* (2019). The RT-qPCR primers used are detailed in Table 2.2.

Table 2.2 | List of primers used in the RT-qPCR

nc16320	GATGACCAACCGGCTGTGGAAGC
nc57384	GTAGTTGGGGTGGTTGTAGG
nc23553	GAACGACTGCTTCTATGCCACCCGA
Y-RNA-3p	CGACAAAAGCTCGACCGGCGC
miR-100	AACCCGTAGATCCGAACTTGTGT
RT4	CTTGCGCAGATAGTCGACACGA

2.10 sRNA libraries

2.10.1 Sample preparation

In Chapter 3, we identified the sRNA populations that are bound by exWAGO inside adult worms, outside of adult worms either inside or outside EVs (vesicular and non-vesicular exWAGO respectively) and we also identified the population of sRNAs secreted by the parasite in total HES.

To identify the sRNAs bound by the vesicular and non-vesicular form of exWAGO, 17 µg of total protein from EVs and 170 µg of total protein from EV-depleted HES was used per replicate, respectively. This is based on data where the same protein amount of EVs and EV-depleted HES were analysed for exWAGO by western blot and then quantified (Chow *et al*, 2019, supplementary data). The data show that there is approximately 10X more exWAGO in EVs compared to the EV-depleted HES when the same amount of total protein is analysed. ExWAGO immunoprecipitations were performed as described in Section 2.7. Briefly, 50 µl of Protein G magnetic beads (Thermo, 10003D) were washed with TBS with 0.05% Tween 20. The EVs and EV-depleted HES were lysed in TBS with 0.05% Triton X-100 with 1 tablet of cOmplete protease inhibitor (Roche, 04693124001) for every 10 ml of buffer and 200 U/ml RNase Inhibitors (Promega, N2515). 10 µl of rat exWAGO-antiserum 3 were used per IP and samples incubated with the beads on a spinning wheel for 45 min at 4°C. The beads were washed with the Low Salt, High Salt and PNK buffers as described in Section 2.7. The RNA was eluted directly in 700 µl of Qiazol (Qiagen, Q79306). To identify the sRNAs enriched in the exWAGO IPs relative to the Unbound fraction of the IP, I also extracted RNA from 200 µl of the Unbound fraction and that was mixed with 1 ml of Qiazol.

To identify the sRNAs bound by the intra-parasite form of exWAGO, adult *H. bakeri* worms that had been one day in culture medium after extraction from F1 mice were used. Adult worms were washed in PBS and then lysed by the Tissue Lyser II (Qiagen) in pre-cooled cartridges for 2 min at 30 Hz twice, in worm lysis buffer (150 mM NaCl, 10 mM Tris.HCl, 0.5 mM EDTA, 0.5% NP40 filtered using a 0.22 µm Millex-GP filter) containing protease and phosphatase inhibitors (1 tablet per 5 ml) (Roche,11873580001 and 04906845001) and 200 U/ml RNasin (Promega, N2515). The lysate was then centrifuged (16.1K rcf, 10 min, 4°C) and the pellet was discarded. The supernatant (lysate) was used for protein quantification using the Qubit Protein Assay kit (Thermo, Q33211). The samples were stored at -80°C until required. For the IPs performed to then generate sRNA libraries, 150 µg of total protein of adult worm lysate was used per sample. 85% of the IP was eluted directly into 700 µl of Qiazol (Qiagen, Q79306).

For the sRNA libraries generated from HES, HES was concentrated to 0.5-1.0 ml using a 5 kDa Vivaspin (4,000 rpm, 4°C) (Fisher, 10646375) and buffer exchanged in PBS (> 20 ml) twice to a final volume of 0.5-1.0 ml. The protein concentration was quantified using the Qubit Protein Assay kit (Thermo, Q33211). The samples were stored at -80°C until required. For the sRNA libraries, 68 µg of total protein of HES was used per sample. This equated to 102.7 µl of HES which was mixed with 550 µl of Qiazol (Qiagen, Q79306).

All samples used for the generation of sRNA libraries were spiked with 7 µl of 10 pM RT4 spike (5’ CTTGCGCAGATAGTCGACACGA 3’) in Qiazol (Qiagen, Q79306) as an internal control for RNA recovery. The RNA was then extracted using the miRNA Serum/Plasma kit (Qiagen, 21784) according to the manufacturer’s instructions and was eluted in 14 µl RNase-free water.

2.10.2 Terminator 5’-phosphate-dependent exonuclease and 5’ polyphosphatase treatments

Prior to generation of the sRNA libraries, the purified RNA underwent various enzymatic treatments. The terminator 5’-phosphate-dependent exonuclease treatment (Lucigen, TER51020) allows degradation of 5’P RNAs hence allowing us to deplete to a great extent the non-5’PPP molecules. The 5’ polyphosphatase treatment (Lucigen, RP8092H) converts 5’PPP molecules to 5’P, allowing these molecules to be ligated to the 5’ adapter. Both treatments were executed according to the manufacturer’s instructions. For samples that were terminator-treated, 11.8 µl of the

extracted RNA was used with the terminator 10X reaction buffer A provided with the kit to allow maximum digestion of 5'P RNAs. The reaction was terminated by addition of Qiazol (Qiagen, Q79306) and the RNA was extracted using the miRNA Serum/Plasma kit (Qiagen, 21784) according to the manufacturer's instructions and was eluted in 22 µl RNase-free water. The terminator-treated RNA and non-terminator-treated samples were polyphosphatase treated (11 µl of the purified RNA) as described in Section 2.11.4, and the reaction was terminated by ethanol precipitation.

2.10.3 Ethanol precipitation

The RNA was ethanol precipitated (overnight, -80°C) by adding 2.5X sample volume of ice cold 100% ethanol, 1/10th sample volume of 3M sodium acetate and 0.5 µl GlycoBlue Coprecipitant (Thermo, AM9515). The samples were then centrifuged (16.1K rcf, 4°C, 30 min) and the supernatant discarded. The pellet was washed twice with 1 ml of 70% ethanol (16.1K rcf, 4°C, 15 min), air-dried on ice, and resuspended in 2.5 µl of nuclease-free water and left on ice (10 min), followed by addition of 2.5 µl of Buffer 1 (TriLink, L-3206) incubation on ice (10 min). The RNA was stored at -80°C.

2.10.4 Library preparation

The sRNA libraries were prepared using the CleanTag Small RNA Library Preparation Kit (TriLink, L-3206) according to the manufacturer's instructions, however half reaction volumes were used. Briefly, the 3' adapter (used at 1:12 dilution) was ligated on the 3' end of the RNA template, then the 5' adapter (used at a 1:12 dilution) was ligated on the 5' end of the RNA template. The tagged libraries were then reverse transcribed and amplified using PCR for 20 cycles following the manufacturer's instructions using the reagents provided in the CleanTag Small RNA Library Preparation Kit.

2.10.5 Quality control & size purification of the libraries

The profile of each sRNA library prior to sequencing was assessed using either Novex 6% TBE acrylamide gels (Thermo) or the High Sensitivity DNA Bioanalyser chip as described in Sections 2.13.13 and 2.13.14. To obtain a similar sequencing depth in our libraries, 50 ng per sRNA library/sample was pooled. If the concentration of the library was too low, as much material as possible was used. The pooled libraries were size selected using gel purification (Section 2.13.13) to remove adapter dimers (where the adapters are ligated to one another and there is no insert). EV, EV-depleted HES,

and HES libraries were size selected between 140-180 bp focusing on the sRNAs (Pool 1). The adult worm libraries were size selected between 140-220 bp (Pool 2), as according to QC analysis there was a population of RNA at 200 bp and we were interested in identifying what species of RNA this is.

2.10.6 Further quality control & sequencing of libraries

The concentration of the pooled library was measured using the Qubit dsDNA HS Assay Kit (Thermo, Q32851). Pool 1 had a final volume of 25.5 µl and had a concentration of 17.2 ng/µl, and Pool 2 had a final volume of 16.5 µl and had a concentration of 10.8 ng/µl after size selection. The library quality was also examined on the 2100 Bioanalyzer Instrument (Agilent). 1 ng of the sample was tested using the Agilent High Sensitivity DNA kit (Agilent, 5067-4626). A peak at 143 nt indicates the presence of guide RNAs. A peak at 120-129 nt would indicate adapter dimers. 11.52 ng of libraries (combined Pool 1 and 2) were sent for sequencing. The libraries were sequenced on the Illumina NextSeq 2000 platform using a single-end 100 base pair run by the Edinburgh Clinical Research Facility.

2.10.7 Bioinformatic analysis of sRNA libraries

Processing of the sRNA reads to remove adapters and map to the genome was performed by Dr Cei Abreu-Goodger. Briefly, the raw sequencing data were checked using FastQC (v0.11.9) for quality issues (Andrews, 2010). No sequencing problems were found. The 3' adapter sequence (TGGAATTCTCGGGTGCCAAGG) was trimmed using reaper (Davis *et al*, 2013), and reads smaller than 18 nt in size after trimming were removed. Reads between 20-25 nt were selected and collapsed into unique reads with their counts using tally (Davis *et al*, 2013). Count matrices were prepared across all samples, as well as plots showing the length distribution and first-nucleotide preference, using *ad hoc* R scripts (R Core Team, 2021). Reads were mapped to the nematode genome (version GCA_900096555.1) with up to 1 nt mismatch using bowtie1 (Langmead *et al*, 2009). The reads mapping to the *H. bakeri* genome were then categorised to annotated regions of the *H. bakeri* genome as described in (Chow *et al*, 2019). Then, I analysed the curated dataset using the web tool Degust (v4.1.1) for differential expression analysis and data visualisation (Powell, 2019, DOI 10.5281/zenodo.3258932). Differential expression analysis on Degust was executed using the Voom/Lima method, and data were analysed using a false discovery rate cut-off of 0.01. The Degust file generated can be found at

https://degust.erc.monash.edu/degust/compare.html?code=129df539f9a57f6794dd202c4ecbe76b#/.

2.11 Electrophoretic Mobility Shit Assays (EMSAs)

2.11.1 Synthesis of 5'PPP guide RNA using *in vitro* transcription

All 5'PPP sRNAs used in this project were synthesised using *in vitro* transcription and all DNA oligos were obtained from IDT. The T7 promoter sequence DNA oligo (TAATACGACTCACTATA) was first annealed to a DNA template oligo comprising the complementary sequence of the guide RNA of interest (in this case of nc16320) followed by the complementary sequence of the T7 promoter (GCTTCCACAGCCGGTTGGTCATCTATAGTGAGTCGTATTA). The following reaction was set up for the annealing of the template and the T7 promoter oligos (95°C for 5 min, slow cool down to room temperature): 0.75 µl of 100 µM T7 promoter oligo was mixed with 0.75 µl of 100 µM template (1.8 µg) and 4.8 µl RNase-free water. The mixture of 6.3 µl of the annealed oligos was then put into a 20 µl (final volume) mix for *in vitro* transcription at 42°C for 3 h, using the AmpliScribe T7-Flash Transcription kit (Lucigen, ASF3507). The DNA template was removed using a DNaseI treatment. The *in vitro* transcription reaction and the DNaseI treatment were executed according to the manufacturer's instructions.

2.11.2 Size purification of oligos

In vitro transcribed or commercially bought RNA oligos (IDT) were gel purified after size separation. RNA samples were prepared to a 1:1 ratio using the Gel loading buffer II (Ambion, AM8546G) and secondary structure was denatured (95°C for 5 min). The RNA was loaded onto a pre-run Novex 15% TBE-Urea Gel (Thermo) in 1X Novex TBE Running buffer (Thermo, LC6675) (80V for 10 min followed by 125V for 115 min). Synthetic sRNAs (20 ng) (IDT) were used as size markers. The RNA was then visualised using 0.01% SYBR gold staining in 1XTBE Running buffer (5 min, in dark, on shaker), and the gel was scanned on the D-DiGit Gel Scanner (Licor). The gel band corresponding to the RNA of interest was excised using a scalpel and blade under a blue LED transilluminator (IO RODEO). The gel bands were shredded by passing them through a pierced 0.5 ml tube to a 1.5 ml DNA LoBind tube (Eppendorf, 0030108051) by centrifugation (3.3K rcf, 2 min or longer as required, room temperature). 300 µl of RNase-free water was added, followed by four freeze-thaw cycles (15 min on dry ice followed by 15 min at room temperature per cycle). The samples were incubated to release the RNA from the gel in the water (overnight, 4°C,

spinning wheel). The gel was removed from the RNA samples using Spin-X centrifuge tube filters (Corning, 8161) (3.3K rcf, 5 min, 4°C) and the RNA was then ethanol precipitated.

2.11.3 Ethanol precipitation

The size purified RNA was ethanol precipitated (overnight, -80°C) by adding 2.5X sample volume of ice cold 100% ethanol, 1/10th sample volume of 3M sodium acetate and 0.5 µl GlycoBlue Coprecipitant (Thermo, AM9515). The samples were then centrifuged (16.1K rcf, 30 min, 4°C,) and the supernatant discarded. The pellet was washed twice with 1 ml of 70% ethanol (16.1K rcf, 15 min, 4°C,), air-dried on ice, and resuspended in RNase-free water. The RNA concentration was measured by NanoDrop.

2.11.4 Generation of 5'P guide RNA

5'P guide RNA was generated by 5' Polyphosphatase treatment of the size purified 5'PPP RNA synthesised by *in vitro* transcription (Sections 2.11.1-2.11.3). Briefly, <5 µg of RNA was mixed with 1 µl of RNA 5' Polyphosphatase (Epicentre, RP8092H), 3 µl of 10X reaction buffer, 0.5 µl of RNasin Ribonuclease Inhibitor (Promega, N2515), and water to make a 30 µl reaction. Following incubation (37°C, 30 min), the RNA was ethanol precipitated as described in Section 2.11.3.

2.11.5 Radioactive labelling of RNA oligos

Size purified 5'PPP RNA was radioactively labelled using [5'-^{32}P] pCp (cytidine 3', 5' bisphosphate (PerkinElmer, BLU019A250UC) and the T4 RNA Ligase 1 (NEB, MO204L) in a 20 µl reaction (4°C, overnight), as follows: 2 µl of 10X reaction buffer, 2 µl of 10 mM ATP, 2 µl of 100% DMSO, 100 ng RNA, 1 µl [5'-^{32}P] pCp, 1 µl T4 RNA Ligase 1 and water as required.

Size purified 5'P RNA was radiolabelled using [γ-^{32}P] ATP (PerkinElmer, NEG502Z250UC) and the T4 Polynucleotide Kinase (Thermo, EK0031) in a total reaction volume of 20 µl (37°C, 1 h), as follows: 2 µl of 10X reaction buffer A, 100 ng RNA, 1 µl [γ-^{32}P] ATP, 1 µl T4 Polynucleotide Kinase and water as required.

The radiolabelling reactions were stopped by addition of 30 µl of 0.1 mM EDTA and unincorporated radioactivity was removed using a MicroSpin G-25 column (GE, 27-5325-01), as per manufacturer's instructions. From previous work in the Buck lab, it is assumed that approximately 50% of the RNA is recovered after the spin column (i.e. 50 ng of RNA). The radiolabelled RNA was stored at -20°C until use.

2.11.6 Gel shift assay

Radiolabelled guide RNA was loaded onto an Argonaute protein (either recombinant exWAGO or recombinant mAGO2) in a 10 µl reaction. The RNA (5 nM) was added to the binding buffer (25 mM HEPES-KOH pH 7.5, 50 mM KCl, 1.5 mM $MgCl_2$, 5 mM DTT, 0.05% IGEPAL CA-630, 2% glycerol and 0.406 units of RNasin Ribonuclease Inhibitors (Promega, N2515)), followed by addition of the Argonaute protein (0-500 nM). The reaction was set up on ice until all reagents were added. The reaction was then incubated (37°C, 1 h) to allow the binding reaction to reach equilibrium.

Then, 1.5 µl of loading dye (50% glycerol, 0.1% bromophenol blue, 0.1% xylene) was added to the reaction. The samples were loaded and separated on a pre-run native 5.9% TBE gel (5.9% acrylamide:bisacrylamide (37.5:1), 2.5% glycerol, 0.5X Novex TBE Running buffer, 1.5 mM $MgCl_2$, 0.1% ammonium persulphate and 0.1% TEMED) in 0.5X Novex TBE Running buffer with 1.5 mM $MgCl_2$ (100V for 55 min). The gels were then exposed on a Phosphor cassette (4°C, approximately 2 h 30) which was scanned on the Typhoon TRIO Variable Mode Imager (GE) under storage phosphor and best sensitivity modes and with 200 µm as the pixel size or on the Typhoon FLA 7000 Imaging System (GE) using the phosphor stage at 650 nm with L5 latitude mode and with 50 µm as the pixel size. Images were processed using ImageQuantTL v7.0 software and exported as .bmp images for quantification analysis.

2.11.7 EMSA data analysis

The exposed images were analysed using the Fiji Image J Software (Schindelin *et al*, 2012) using inverted images so that bands are white. A rectangle was drawn around each band, and the integrated density (IntDen) of the band was calculated by the software. The area/size of the rectangle remained constant when analysing a gel. The integrated density of the free guide RNA, the background per lane, and the Argonaute:guide RNA complex (shifted band) were obtained and the fraction of bound guide was calculated as follows:

$$\frac{\text{Complex IntDen} - \text{Background IntDen}}{(\text{Complex IntDen} - \text{Background IntDen}) + (\text{Free IntDen} - \text{Background IntDen})}$$

The data were fitted using a non-linear regression using saturation binding – specific binding with Hill slope equation on GraphPad Prism v.8.4.1, to calculate the Kd and Bmax values.

2.12 Slicer assay

To test the ability of exWAGO to cleave targets *in vitro*, I used the guide RNAs generated in Section 2.11. I also generated fully complementary targets of these guides using *in vitro* transcription. The same method was used to synthesise, size purify and radiolabel the targets as described in Section 2.11. The DNA templates used and the RNA targets generated are shown in Table 2.3.

Table 2.3 | List of DNA templates used to generate RNA targets fully complimentary to the guide sRNAs tested.

Oligo details		Oligo sequence
T7 promoter		TAATACGACTCACTATA
Generates target complementary to nc16320	Used	GAGTTTGAATGAGTCCCTCTGTTTATGGATGATGACCAA CCGGCTGTGGAAGCAATGGTCTATAGTGAGTCGTATTA
	Generated	GACCAUUGCUUCCACAGCCGGUUGGUCAUCAUCCAUAA ACAGAGGGACUCAUUCAAACUC
Generates target complementary to miR100	Used	GAGTTTGAATGAGTCCCTCTGTTTATGGATAACCCGTAG ATCCGAACTTGTGAATGGTCTATAGTGAGTCGTATTA
	Generated	GACCAUUCACAAGUUCGGAUCUACGGGUUAUCCAUAAA CAGAGGGACUCAUUCAAACUC

Non-radiolabelled ‘cold’ guide RNA (25 nM) was first loaded onto recombinant Argonaute protein (either exWAGO or mAGO2) (25 nM) in an 8 µl reaction in binding buffer containing BSA (1%) as described in Section 2.11.6. Then the radiolabelled ‘hot’ target was prepared by denaturing the molecule to resolve any secondary structures (95°C for 1 min, followed by snap-cooling on ice). After loading exWAGO with ‘cold’ guide, the linear ‘hot’ target (5 nM) was added to the reaction, with a final volume of 10 µl. The samples were mixed, briefly spun, and incubated (37°C, 1 h) to allow cleavage to occur. After incubation, the samples were briefly spun and the reaction was quenched by addition of 10 µl 2X Urea dye (8 M urea, 5 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). The samples were denatured (95°C, 1 min) and spun briefly. The 20 µl reaction was resolved on a pre-run 15% (0.5x) TBE-Urea SequaGel (0.75 mm) gel (80V for 10 min, followed by 125V for 108 min) in 0.5X Novex TBE Running buffer. The gels were exposed on a Phosphor cassette and imaged as described in Section 2.11.6. The exposed images were analysed using the Fiji Image J Software (Schindelin *et al*, 2012). The integrated density of the target RNA, the background per lane, and the cleaved target RNA were obtained as described in Section 2.11.7, and the fraction of target cleaved was calculated as follows:

$$\frac{\text{Cleaved Target IntDen} - \text{Background IntDen}}{(\text{Cleaved Target IntDen} - \text{Background IntDen}) + (\text{Target} - \text{Background IntDen})}$$

2.13 Modified CLASH protocol, library preparation and bioinformatic analyses

2.13.1 Immunoprecipitation of exWAGO or mAGO2 from gut samples

In vivo gut material was obtained from the duodenum of three F1 male mice 14 days post challenge with 400 L3 stage *H. bakeri* and from three uninfected F1 male mice. Immunoprecipitations were executed as described in Section 2.7. 100 µl of Protein G magnetic beads (Thermo, 10003D) and 20 µl of rat exWAGO-antiserum or rat naïve serum or mouse anti-mAGO2 antibody were used per sample. Sample immunoprecipitations were performed for 45 min at 4°C. All buffers used (Protein G lysis, Low Salt, High Salt and PNK buffers) contained 5 mM of β-mercaptoethanol. Prior to eluting the protein and the RNA, several enzymatic reactions were carried out, as described below. After each enzymatic treatment the beads were washed as follows: one wash with Low Salt buffer followed by two washes with High Salt buffer (5 min, 4°C, rotating wheel), followed by one wash with Low Salt buffer and finally one wash with PNK buffer wash.

2.13.2 'On bead' RNase digestion

Following sample immunoprecipitation and bead washes, the beads were resuspended in RNace-IT/PNK buffer (per sample: 500 µl PNK buffer with 1 µl of 1:40 RNace-IT ribonuclease cocktail) (Agilent, 400720) and incubated (20°C, 7 min, 1,050 rpm). The beads were placed on the magnet and the reaction mix was removed. This treatment shortens the mRNA targets to allow intermolecular ligation with the guide at later stages and leaves the target mRNA with a 5'OH and a 3'P end. The beads were washed as before.

2.13.3 'On bead' 5' end phosphorylation

The 5'OH group of the target mRNA is converted to 5'P to allow RNA-RNA ligation. The beads were resuspended in 80 µl of reaction mix per sample (described below) and incubated (20°C, 2 h 30 min, 1,000 rpm intermittent shaking). The beads were then washed as before.

Reagent	**Volume (µl)**	**Provider**	**Catalogue no.**
T4 PNK (3' phosphatase minus)	4.0	NEB	M0236L
10X PNK buffer	8.0	NEB	included in M0236L
100 mM ATP	0.8	Promega	E6011
RNasin ribonuclease inhibitor	2.0	Promega	N2115
Nuclease-free water	65.2	-	-

2.13.4 'On bead' intermolecular guide RNA-target RNA ligation

To generate guide-target chimeras, the beads were resuspended in 160 µl of reaction mix per sample (described below) and incubated (16°C, overnight, 1000 rpm intermittent shaking). The beads were then washed as before.

Reagent	Volume (µl)	Provider	Catalogue no.
T4 RNA ligase I (ssRNA ligase)	4.0	NEB	M0204L
10X ligase buffer	16.0	NEB	included in M0204L
100 mM ATP	1.6	Promega	E6011
RNasin ribonuclease inhibitor	4.0	Promega	N2115
Nuclease-free water	134.4	-	-

2.13.5 'On bead' 3' end dephosphorylation

The free 3'P group generated after RNase digestion with RNace-IT is converted to 3'OH end to allow ligation of the 3' adapter. The beads were resuspended in 80 µl of reaction mix per sample (described below) and incubated (20°C, 45 min, 1,000 rpm intermittent shaking). The beads were then washed as before.

Reagent	Volume (µl)	Provider	Catalogue no.
TSAP Thermosensitive Alkaline Phosphatase	8.0	NEB	M9910
MULTI-CORE 10X TSAP buffer	8.0	NEB	included in M9910
RNasin ribonuclease inhibitor	2.0	Promega	N2115
Nuclease-free water	62.0	-	-

2.13.6 'On bead' 3' end adapter ligation

To ligate the 3' adapter, the beads were resuspended in 80 µl of reaction mix per sample (described below) and incubated (16°C, 6 h or overnight, 1,000 rpm intermittent shaking). The beads were then washed as before.

Reagent	Volume (µl)	Provider	Catalogue no.
3' IR800-labelled adapter (100 µM)	0.8	TriLink	custom made
T4 RNA Ligase 2, truncated K227Q	4.0	NEB	M0351L
10X ligase buffer	8.0	NEB	included in M0351L
PEG 8000 (50%)	16.0	NEB	included in M0351L
RNasin ribonuclease inhibitor	2.0	Promega	N2115
Nuclease-free water	49.2	-	-

2.13.7 Elution of exWAGO/mAGO2-RNA complexes

After the final wash, 9/10th of the beads were eluted in 700 µl Qiazol Lysis Reagent (Qiagen, 79306) for RNA analyses and the remaining 1/10th of the beads were eluted in 30 µl of 4X NuPAGE LDS sample buffer (Thermo, NP0008) with 0.1 M DTT (10 min, 70°C, 1,000 rpm) for protein analyses. The eluate was removed from the magnetic beads and the samples were stored at -80°C.

2.13.8 Western blot analysis

To check that the IPs were successful, the protein samples were analysed using western blotting (see Section 2.8).

2.13.9 RNA extraction

The RNA from the samples eluted in Qiazol Lysis Reagent (Qiagen, Q79306) was extracted using the miRNA Serum/Plasma kit (Qiagen, 21784) according to the manufacturer's instructions. RNAs were eluted in 14 µl RNase-free water.

2.13.10 5' Polyphosphatase treatment

To allow ligation of the 5' adapter to the 5' end of the guide RNA, the 5'end of the guide RNA needs to be a monophosphate (5'P) rather than a triphosphate (5'PPP). All the RNA was used in a 30 µl reaction (described below) and incubated (37°C, 30 min).

Reagent	**Volume (µl)**	**Provider**	**Catalogue no.**
RNA	14.0	-	-
RNA 5' Polyphosphatase	0.5	Epicentre	RP8092H
10X reaction buffer	3.0	Epicentre	included in RP8092H
RNasin ribonuclease inhibitor	0.5	Promega	N2515
Nuclease-free water	12.0	-	-

2.13.11 Ethanol precipitation

The RNA was ethanol precipitated as described in Section 2.10.3. The air-dried pellet was resuspended in 2.5 µl of nuclease-free water and left on ice (10 min), followed by addition of 2.5 µl of Buffer 1 (TriLink, L-3206) incubation on ice (10 min). 0.2 µl of the RNA made to 1 µl using nuclease-free water were used to measure the RNA concentration using the Qubit RNA HS Assay kit (Thermo, Q32852) according to the manufacturer's instructions. The RNA was kept at -80°C.

2.13.12 Library preparation

The libraries were then prepared using the CleanTag Small RNA Library Preparation Kit (TriLink, L-3206). The manufacturer's protocol was followed from Step 2: 5' Adapter Ligation to 3' Tagged RNA Template onwards, but half reaction volumes were performed. Briefly, the 5' adapter (used at a 1:12 dilution) was ligated on the 5' end of the RNA template. The tagged library was then reverse transcribed and amplified using PCR for 19 cycles.

2.13.13 Quality control & size purification of the libraries

The libraries (40 µl) were mixed with 10 µl of 6X Gel loading dye (NEB, B7025S) and 10 µl of nuclease-free water. Equal volume (6 µl) of the libraries from all the samples were pooled together and resolved on a 12-well 6% Novex TBE gel (Thermo, EC62652BOX) in 1X Novex TBE Running buffer (Thermo, LC6675) (100V for 80 min). 100 ng of the digested plasmid pBR322 DNA-MspI digest (NEB, N3032S) was run on the gel as a size marker. The RNA was then visualised using 0.01% SYBR gold staining in 1XTBE Running buffer (15 min, in dark, on shaker), and the gel was scanned on the D-DiGit Gel Scanner (Licor). The gel band corresponding to the tagged library (approximately 140-217 bp) was excised using a scalpel and blade under a blue LED transilluminator (IO RODEO). The gel bands were shredded by passing them through a pierced 0.5 ml tube to a 1.5 ml DNA LoBind tube (Eppendorf, 0030108051) by centrifugation (3.3K rcf, 2 min or as required, room temperature). 300 µl of nuclease-free water was added. The samples were incubated to release the RNA from the gel in the water (overnight, 4°C, spinning wheel). The gel was removed from the RNA samples using Spin-X centrifuge tube filters (Corning, 8161) (3.3K rcf, 5 min, 4°C) and the RNA was then precipitated overnight as described in Section 2.10.3. Each reaction was resuspended in 4 µl of nuclease-free water and the RNA was pooled together (final volume was 28 µl).

2.13.14 Further quality control & sequencing of libraries

The concentration of the pooled library was measured using the Qubit dsDNA HS Assay Kit (Thermo, Q32851). The library quality was also examined on the 2100 Bioanalyzer Instrument (Agilent). 1 ng of the sample was tested using the Agilent High Sensitivity DNA kit (Agilent, 5067-4626). A peak at 143 nt and a shoulder indicates the presence of guide RNAs and targets or chimeric reads. A peak at 120-129 nt indicates adapter dimers. 2.47 ng of the pooled library was sent for sequencing. The libraries were sequenced on the Illumina NovaSeq 6000 system using a 100 bp single

end run by the Edinburgh Genomics facility. The number of reads detected are shown in Table 2.4:

Table 2.4 | The number of reads sequenced per sample. Inf = infected gut tissue, Un = uninfected gut tissue; yes/no = refers to RNase treatment; A-C = replicates.

Sample Name	Barcode Sequence	Total no. of Reads	No. of Reads after adapter removal
Inf_A_yes_WAGO	CCGTCC	32,031,094	28,372,258
Inf_A_no_WAGO	GTAGAG	17,898,020	16,100,243
Inf_A_mAGO2	GTCCGC	34,128,068	32,177,941
Inf_B_yes_WAGO	GTGAAA	23,072,196	21,358,864
Inf_B_no_WAGO	GTGGCC	25,507,352	23,703,388
Inf_B_mAGO2	GTTTCG	35,557,501	34,467,928
Inf_C_yes_WAGO	CGTACG	25,491,195	22,655,480
Inf_C_no_WAGO	GAGTGG	30,441,529	27,203,148
Inf_C_mAGO2	GGTAGC	32,170,997	30,717,921
Inf_naïve_serum	ACTGAT	22,060,767	21,311,523
Un_A_WAGO	ATGAGC	353,177	332,935
Un_A_mAGO2	ATTCCT	42,094,856	41,686,373
Un_B_WAGO	CAAAAG	368,409	343,924
Un_B_mAGO2	CAACTA	33,987,590	33,655,746
Un_C_WAGO	CACCGG	452,753	415,369
Un_C_mAGO2	CACGAT	43,710,897	42,650,622

2.14 Bioinformatic analyses of CLASH data

2.14.1 Identification of exWAGO guide sRNAs

Bioinformatic analyses of exWAGO guide sRNAs from samples generated by CLASH was performed by Jose Roberto Bermudez-Barrientos and also involved the input of Dr Cei Abreu-Goodger. The raw sequencing data were checked using FastQC (v0.11.8) for quality issues (Andrews, 2010). No sequencing problems were found. The 3' adapter sequence (TGGAATTCTCGGGTGCCAAGG) was trimmed using reaper (Davis *et al*, 2013), and reads smaller than 18 nt in size after trimming were removed. All reads were mapped simultaneously to the nematode (version GCA_900096555.1) and mouse (version mm10) genome using ShortStack (Johnson *et al*, 2016) to define clusters of locations on both genomes, allowing up to 1 nucleotide mismatch. In addition, reads were mapped to both genomes with bowtie1 to be able to detect reads that exclusively mapped to one or the other genome (with up to 1 mismatch). Reads mapping to the *H. bakeri* genome were further categorised

to annotated regions of the *H. bakeri* genome as described in Chow *et al*, (2019). To compare if the exWAGO immunoprecipitated from *in vivo* gut samples represents a particular form of the protein (i.e. vesicular, non-vesicular, intra-parasite), reads that mapped exclusively to the worm genome were considered and data were compared using a Multi-Dimensional Scaling plot generated with the edgeR package (Robinson *et al*, 2010).

2.14.2 Identification of exWAGO host targets not in a chimeric read

Bioinformatic analysis was performed by Dr Sujai Kumar to identify putative mouse targets of exWAGO that are not in a chimeric read. The raw sequencing data were checked using FastQC (v0.11.8) for quality issues (Andrews, 2010). No sequencing problems were found. The 3' adapter sequence was trimmed using Cutadapt (v2.7) (Martin, 2011). Reads smaller than 18 nt in size and reads that matched perfectly to the nematode genome (version GCA_900096555.1) were removed. Reads mapping to RefSeq mouse rRNAs were identified using bowtie (v1.2.2) (Langmead *et al*, 2009) allowing up to 1 nucleotide mismatch and were also removed. The remaining sequences were then mapped to the mouse genome (version mm10) with up to 1 nucleotide mismatch. The data were analysed using ShortStack (Johnson *et al*, 2016). Clusters that overlapped known mouse miRNA coordinates (GENCODE Mouse Annotations version M23) (Frankish *et al*, 2019) were identified using BEDTtools (v2.29.0) (Quinlan & Hall, 2010) and removed, thus removing host guide RNAs. The data files with read counts were uploaded on the web tool Degust (v4.1.1) for differential expression analysis and data visualisation (Powell, 2019, DOI 10.5281/zenodo.3258932). Differential expression analysis on Degust was executed using the Voom/Lima method, and data were analysed using a false discovery rate cut-off of 0.01 and the fold change was set to average. The Degust file can be found at http://degust.erc.monash.edu/degust/compare.html?code=3ea6602f802df716e7c3e352ddd3c580.

2.15 Gene Ontology enrichment analysis

Gene ontology enrichment analysis of biological processes was performed using the online tool found at http://geneontology.org/ (Ashburner *et al*, 2000; Mlecnik *et al*, 2018). The statistical analysis tool PANTHER Overrepresentation Test (released 2021-02-24) with the GO Ontology database DOI:10.5281/zenodo.5080993 (released

2021-07-02) against the *Mus musculus* reference list was employed. The test type applied was Fisher's Exact with False Discovery Rate correction (FDR < 0.05).

2.16 Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS)

2.16.1 Sample preparation for LC-MS/MS

Immunopurified exWAGO complexes or appropriate negative controls were analysed by LC-MS/MS to identify the proteins that interact with exWAGO. Following the last wash step of the exWAGO immunoprecipitation protocol, the beads were washed once with nuclease-free water and stored at -80°C until further processed by the Proteomics and Metabolomics Facility at the Roslin Institute (Edinburgh) or the EdinOmics facility (University of Edinburgh).

The following protocols have been provided by Dr Dominic Thekkedath Kurian (Proteomics and Metabolomics Facility at the Roslin Institute) and Ms Lisa Imrie (EdinOmics facility).

2.16.2 Qualitative LC-MS/MS by Proteomics and Metabolomics Facility of the Roslin Institute

Qualitative mass spectrometry was used to identify the protein interactions partners of exWAGO in the MODE-K cell line expressing FLAG-tagged exWAGO and in *H. bakeri*-infected mouse gut tissue. Washed beads were resuspended in 10% Triflouroethanol (pH 8.0) (1 h, gentle shaking). The samples were reduced (5mM dithiothreitol) and alkylated (10mM iodoacetamide). The proteins were digested with sequencing grade modified trypsin (Promega) (to avoid trypsin auto-proteolysis) (Yang *et al*, 2015). The mixture containing beads was spun (800 rcf) and supernatant containing the digested peptides was collected. The digested peptides were cleaned up using Stagetips following standard protocols.

Nanoflow LC-MS/MS was performed on a micrOTOF-II mass spectrometer (Bruker, Germany) coupled to a RSLCnano LC system (Thermo). The digested samples were delivered to a trap column (Acclaim PepMap100, 5 µm, 100 Å, 100 µm i.d. × 2cm) at a flow rate of 20 µL/min in 100% solvent A (0.1% formic acid in LC-MS grade water). After 4 min of loading and washing, peptides were transferred to an analytical column (Acclaim PepMap100, 3 µm, 100 Å, 75 µm i.d. × 25 cm) and separated at a flow rate of 300 nL/min using a 60-min gradient from 7% to 35% solvent B (solvent B, 0.1% formic acid in acetonitrile). The eluted peptides from the liquid chromatography were electrosprayed directly on to the mass spectrometer for MS and MS/MS analysis in a

data-dependent mode of acquisition. The m/z values of tryptic peptides were measured using an MS scan (300-2000 m/z), followed by MS/MS scans of the six most intense ions. Rolling collision energy for fragmentation was selected based on the precursor ion mass and a dynamic exclusion was applied for 30 sec.

Raw spectral data were processed using the DataAnalysis (Bruker) software. The resulting peak lists were searched using Mascot 2.4 server (Matrix Science, London, UK) against the Uniprot mouse sequence database containing 53,217 entries. Mass tolerance on peptide precursor ions was fixed at 25 ppm and on fragment ions at 0.06 Da. The peptide charge was set to 2+ and 3+. Carbamidomethylation of cysteine was selected as a fixed modification and oxidation of methionine and de-amidation were chosen as variable modifications. False discovery rate was limited to < 1% for peptide IDs after searching decoy databases.

The .csv files with the proteins detected were processed by the author in Python to produce one file for direct comparison between the samples and the controls. The Python script was written by Dr Mate Ravasz and can be found at https://github.com/Ravasz/ed/blob/master/src/pythoncode/mascot_output_parser.py. Further processing of the data was performed by the author using Microsoft Excel.

2.16.3 Dimethyl-labelling quantitative LC-MS/MS by Proteomics and Metabolomics Facility of the Roslin Institute

Quantitative proteomic analysis using dimethyl-labelling was performed to identify the protein interactor partners of exWAGO using *H. bakeri*-infected mouse gut tissue. The washed beads were resuspended in ammonium bicarbonate (50 mM) and digested with sequencing grade modified trypsin (37°C, 2 h) (Promega), following reduction with dithiothreitol (10 mM) and alkylation with iodoacetamide (20 mM). The resulting peptide mixture was centrifuged (2,000 rcf) and supernatant was collected. The digested peptides were cleaned up using Pierce C18 spin columns and labelled with stable isotopic dimethylation following standard protocols (Boersema *et al*, 2009). Peptides from pulled down samples were resuspended in TEAB (100 mM) and reductive dimethylation reaction was performed by adding formaldehyde isotopes CH_2O or CD_2O to a final concentration of 0.16% (v/v), to generate ‘heavy’ and ‘light’ labelled peptides for experimental and negative control samples respectively. The labelling reaction was completed by addition of sodium cyanoborohydride (24 mM) to both samples and extra cyanoborohydride was quenched by adding ammonia to a

final concentration of 0.16% (v/v) and the ‘light’ (negative control) and ‘heavy’ (experimental sample) labelled peptides were mixed for each sample.

LC-MS/MS was performed as described in Section 2.16.2. Raw spectral data were processed by DataAnalysis (Bruker) software and the resulting peak lists were searched using Mascot 2.4 server (Matrix Science) against Uniprot mouse sequence database containing 53,217 entries. Mass tolerance on peptide precursor ions was fixed at 25 ppm and on fragment ions at 0.5 Da. The peptide charge was set to 2+ and 3+. Carbamidomethylation of cysteine was used as a fixed modification and oxidation of methionine and light and heavy dimethylation of N-terminus and lysine were chosen as variable modifications. False discovery rate was limited to < 1% for peptide IDs after searching decoy databases. Dimethyl quantification was performed by using WARPLC plugin on Proteinscape 3.1 software (Bruker) to integrate extracted ion chromatogram of every precursor. Peptide ratios were normalised based on setting the overall peptide median ratio at one, which corrects for unequal protein sampling and a coefficient of variability of peptide ratios were also determined for each quantified protein. A table of the proteins detected was compiled by the facility.

2.16.4 Label-free quantitative LC-MS/MS by EdinOmics facility

Label-free quantitative proteomic analysis was performed to identify the protein interactor partners of exWAGO inside *H. bakeri* adult worms. Samples were subjected to an on-bead tryptic digest. In short, beads were resuspended in 40 µl 6 M Urea/25 mM Ammonium Bicarbonate before reduction/alkylation. 1 µg of trypsin was added to all samples and incubated overnight at 37°C. Samples were cleaned up the following day using Bond Elut C18 RP tips (Agilent) and peptides were dried down in a speed vac. Samples were reconstituted in 10 µl MS-loading buffer (0.05%TFA in water) and then a 1 in 10 dilution of this (in 0.05% TFA) was filtered using Millex filter. 5 µl was injected for analysis.

Nano-ESI-HPLC-MS/MS analysis was performed using an online system of a nano-HPLC (Dionex Ultimate 3000 RSLC, Thermo-Fisher) coupled to a QExactive mass spectrometer (Thermo-Fisher) with a 300 µm x 5 mm pre-column (Acclaim Pepmap, 5 µm particle size) joined with a 75 µm x 50 cm column (EASY-Spray, 3 µm particle size). The nano-pump was run using solvent A (2% acetonitrile in water and 0.1% formic acid) and solvent B (80% acetonitrile-20% water and 0.1% formic acid) and peptides were separated using a multi-step gradient of 2–98% buffer B at a flow rate of 300 nl/min over 90 min.

Progenesis (version 4 Nonlinear Dynamics, UK) was used for LC-MS label-free quantitation. Filtering was carried out so that only MS/MS peaks with a charge of 2+, 3+ or 4+ were taken into account for the total number of 'features' (signal at one particular retention time and m/z) and only the five most intense spectra per 'feature' were included. MS/MS spectra was searched using MASCOT Version 2.4 (Matrix Science Ltd, UK) against the *H. polygyrus* database found on WormBase ParaSite (BioProject PRJEB15396) with maximum missed-cut value set to 2. The following parameters were used in all searches: i) variable methionine oxidation, ii) fixed cysteine carbamidomethylation, iii) precursor mass tolerance of 10 ppm, iv) MS/MS tolerance of 0.05 Da, v) significance threshold (p) below 0.05 and vi) final peptide score of 20. From the Progenesis exported results sheet, differentially expressed proteins were considered significant if the p-value was less than 0.05 and if the number of peptides used in quantitation per protein was equal to or more than 2. A table of the proteins detected was compiled by the facility.

2.17 Silver stain

To assess the protein profile of immunoprecipitations prior to mass spectrometry analysis, samples were analysed by silver stain. Samples were reduced and separated on 4-12% Bis-Tris NuPAGE SDS gels (Thermo) as described in Section 2.8. 1-2 µl of the Precision Plus Protein All Blue Prestained Protein Standards ladder (Bio-Rad, 1610373) was used as a size marker. The gels were fixed in fixing solution (10% glacial acetic acid, 40% ethanol) (3 h, rocker, room temperature) and sensitised in sensitisation solution (0.2% sodium thiosulphate, 6.8% w/v sodium acetate, 30% ethanol) (30 min, rocker, room temperature). Then the gels were washed with distilled water three times (5 min, rocker, room temperature) and incubated in silver nitrate solution (0.25% w/v silver nitrate) (20 min, rocker, room temperature, dark). Following, the gels were washed with distilled water twice (1 min, rocker, room temperature) and the protein profile was developed in developing solution (2.5% w/v sodium carbonate, 0.0074% formaldehyde) (2-15 min, rocker, room temperature). The reaction was stopped by rinsing the gels with distilled water and incubating them in stop solution (1.46% w/v EDTA disodium salt). The gels were imaged using the ChemiDoc Gel Imaging system (BioRad).

2.18 Immunofluorescence

To examine the localisation of exWAGO in MODE-K cells and validate exWAGO as a putative protein interactor of SFPQ, we used immunofluorescence (Chapter 5, Fig.

5.9C). 1.8-2.0x10^5 MODE-K cells (parental or exWAGO-expressing) were seeded on sterilised coverslips (Fisherbrand, 12-541A) and left to attach overnight. The following day, the cells were washed twice with PBS and then fixed in 4% paraformaldehyde/PBS, permeabilised (0.25% Triton X-100/PBS, 10 min) and blocked in 5% BSA/PBST (0.1% Tween 20) with 22.52 mg/ml glycine (1 h, gentle rocking, room temperature). Cells were then incubated overnight with primary antibody in 5% BSA/PBST. Following three PBS washes, the cells were incubated with the secondary antibodies in 5% BSA/PBST (1 h, dark, room temperature). The coverslips were mounted on glass slides in Vectashield mounting medium with DAPI (Vector Laboratories, H-1200). Images were acquired using the Nikon Ti-E inverted microscope with Yokogawn CSU X-1 spinning disk head and the Andor 888 Ultra EMCCD camera (ω/13 µm pixels). The laser lines 405, 488 and 561 were used (emission filter: 447/38, 523/50 and 620/68 respectively). Image J 1.52n was used for the analysis of the images (Schneider *et al*, 2012).

2.19 Uptake assays

2.19.1 Uptake of Cy-5-labelled EV-depleted HES proteins

To examine if non-vesicular molecules secreted by *H. bakeri* are internalised by epithelial cells *in vitro*, proteins in EV-depleted HES or PBS were labelled using a Cy-5 NHS ester dye (AAT Bioquest, 151) according to the manufacturer's instructions. Briefly, the dye was resuspended in DMSO to 1 mg/ml and then EV-depleted HES or equivalent volume of PBS was labelled at a 15:1 (dye to protein) molar ratio. Unconjugated dye was removed using the Zeba 7K MWCO spin columns (Thermo, 89882) as per manufacturer's instructions. The labelled EV-depleted HES and PBS were stored at -80°C in the dark, until required. 1.5x10^5 MODE-K cells were seeded in a 24-well plate with 400 µl MODE-K media. The following day, the cells were treated with Cy-5-labelled EV-depleted HES at a final concentration of 10 µg/ml or 50 µg/ml or with an equivalent volume of Cy-5-labelled PBS as a negative control for 4 h or 24 h. The cells were washed twice with PBS and fresh media was added. Images were then taken on the EVOS M7000 Imaging system (Invitrogen by Thermo Fisher Scientific).

2.19.2 Uptake of non-vesicular exWAGO

To examine whether the non-vesicular exWAGO is taken up by epithelial cell *in vitro*, 8x10^4 MODE-K cells were seeded in a 24-well plate with 400 µl MODE-K media. The following day, the cells were treated with EV-depleted HES at a final concentration of

50 µg/ml or with an equivalent volume of PBS as a negative control for 4 h or 24 h. Then, the cells were washed twice with PBS, fixed using 4% paraformaldehyde (200 µl, 30 min) and then were left in immunofluorescence buffer (PBS with 10% BSA and 0.1% sodium azide) at 4°C covered in parafilm until usage. Dr Tom Fenton then permeabilised the cells in 0.2% Triton X-100 in PBS (200 µl, 10 min) and washed the cells with PBS. The cells were blocked with 5% goat serum (200 µl, 30 min, room temperature) and washed three times with PBS. Then the primary antibody was added (rat exWAGO anti-serum or naïve rat serum, 1:500) at 4°C overnight, the cells were washed three times with PBS and incubated with the secondary antibody (goat anti-rat AF647, 1:500) 1h in the dark at room temperature. Images were then taken on the EVOS M7000 Imaging system.

2.20 Immunisation of mice

To examine if exWAGO could be a vaccine candidate, female C57BL/6 mice of 7-12 weeks old were immunised intraperitoneally with recombinant FLAG-tagged exWAGO or PBS or HES in Imject alum adjuvant (Thermo, 77161) at 1:1 volume ratio (total intraperitoneal injection volume was 200 µl). Recombinant exWAGO and PBS were administered subcutaneously (200 µl injection) in QuilA adjuvant (15 µg of QuilA per mouse) (Brenntag Biosector, 8047-15-2). Mice were primed (10 µg) and boosted (2 µg), twice 28- and 35-days post priming. 42 days post priming, the mice were challenged with 200 *H. bakeri* L3 stage larvae by oral gavage. Mice were sacrificed 14, 28 and/or 35 days post-challenge. Mice were culled by CO_2 overdose or by intraperitoneal administration with dolethal (30 mg) (Vetoquinol). The blood was acquired via tail bleed or for euthanised mice it was collected using cardiac puncture. Following culling, the small and large intestine were carefully removed and stretched. The small intestine was separated from the large intestine. The small intestine was subsequently divided in three equal parts (named top, middle, bottom section). Each part was cut open and the number of adult worms were counted. Stools from the large intestine were obtained for faecal egg counts on the harvest days as well as on day 21 post-challenge. The immunisation timeline and design are detailed in Chapter 6 (Fig. 6.1, Table 6.1).

2.21 Faecal egg counts

The number of eggs, a measure of worm infectivity, was quantified as described in Camberis *et al*, (2003). Briefly, moist stools were collected, weighed, and eggs were counted using the McMaster Egg Counting Chamber (CellPath, RSW-0611240),

following resuspension in 2 ml of dH_2O and flotation in 2 ml of saturated sodium chloride. The eggs per gram of faeces was calculated as follows:

$$\frac{\text{number of eggs} \times 26.66 \text{ (area of counting chamber)}}{\text{stool weight (g)}}$$

2.22 Collection of serum

Blood collected from the mice was incubated (1-2 h, room temperature) and centrifuged (5,000 rcf, 5 min, room temperature). The supernatant containing the serum was centrifuged again (5,000 rcf, 5 min, room temperature) and the supernatant was stored at -70°C until required.

2.23 ELISA

ExWAGO-specific antibody responses from the sera of vaccinated mice were measured using an Enzyme-Linked Immunosorbent Assay. Flat-bottomed 96-well plates were coated with 0.05 µg of recombinant exWAGO (Sino Biological) in coating buffer (18 mM sodium carbonate, 45 mM sodium bicarbonate, pH 9.6) and incubated (overnight, 4°C). The plates were then blocked (4% BSA in TBS, room temperature, 2 h) and washed three times (TBS/T; 0.1% Tween 20 in TBS). Serum from each mouse was serially diluted in dilution buffer (1% BSA in TBS/T), incubated (overnight, 4°C) and the plates were then washed four times as before. Following addition of the detection antibody (37°C, 3 h, in dark) the plates were washed four times as before and twice with ddH_2O. The reaction was developed using TMB substrate buffer at room temperate (SeraCare, 5120-0047). The reaction was stopped (10% phosphoric acid H_3PO_4) and the optical density was read at 450 nm on the Varioskan Flash (Thermo Scientific). Samples were analysed in duplicates and wells without serum were used to determine the background optical density. Serum from uninfected (naïve) mice was also used as a negative control. Table 2.5 details the detection antibodies used for ELISAs.

Table 2.5 | Antibodies used to measure exWAGO-specific IgG1 and IgA antibodies by ELISA

Detection Antibody	Dilution	TMB incubation time (min)	Manufacturer (catalogue no.)
goat anti-mouse IgG1, Human ads-HRP	1:2,000	8	SouthernBiotech (1070-05)
goat anti-mouse IgA-HRP	1:2,000	11.5	SouthernBiotech (1040-05)

2.24 Measuring ATP content of worms

The amount of ATP per adult worm was measured as a means of assessing worm fitness after vaccination. Three worms per sex per mouse per vaccination group were collected where possible. If inadequate worms were collected from each mouse, worms were supplemented from other mice from the same vaccination group. The worms were visually inspected to ensure they were alive based on mobility and that they were not injured. Each worm was placed in a 2 ml safe-lock Eppendorf tube with a steel bead (5 mm) and 110 µl of PBS. Then 110 µl of CellTiter-Glo reagent (Promega, G7570) were added, and the worms were homogenised in pre-cooled cartridges using the Tissue Lyser II (Qiagen) by bead beating (2 min at 30 Hz, twice). The homogenate was incubated (10 min) to stabilise the luminesce signal and then it was centrifuged (1,000 rcf, 3 min, room temperature). A standard curve from 0.1 nM to 10 µM was also prepared fresh using recombinant ATP (Promega, E601B) by mixing 110 µl of the prepared dilutions with 110 µl of CellTiter-Glo reagent. Then 100 µl per technical duplicate were added on 96-well opaque-walled white plates (Greiner Bio-one, 655075). The plate was incubated (10 min, room temperature) on an orbital shaker and the luminesce was measured on the Varioskan Flash (Thermo Fisher Scientific) using the SkanIt Software 2.4.3 RE (luminometric measurement, 1,000 ms measurement time, auto dynamic range, normal optics). The following negative controls were also used: PBS only, PBS with worm homogenate only, and PBS only with CellTiter-Glo reagent.

Chapter 3: Identification of the guide sRNAs associated with exWAGO and analysis of its binding properties

Introduction

Buck *et al*, (2014) first reported the presence of exWAGO in the excretory/secretory products of *H. bakeri* adult worms (HES). HES can be divided in two moieties: (1) the extracellular vesicles (EVs) fraction, and (2) the non-vesicular fraction (EV-depleted HES). Proteomic and western blot analyses indicate that exWAGO is enriched in the EVs secreted by adult worms compared to the non-vesicular fraction when analysed by protein weight (Buck *et al*, 2014; Chow *et al*, 2019). However, unpublished data from our lab shows that exWAGO is enriched in the non-vesicular extracellular fraction when analysed by volume of material (i.e. when we fractionate HES and analyse equal volumes of the two fractions). These results indicate that exWAGO is mainly found outside of EVs but that it is not the most abundant protein found in the non-vesicular fraction. The fact that exWAGO is present in two extracellular environments implies that exWAGO exists in two different extracellular forms, referred to in this thesis as the vesicular and non-vesicular forms. This raises several questions. What dictates the export route of the vesicular and non-vesicular exWAGO forms? Are they internalised by the same type of host cells? Do they bind the same guide small RNAs (sRNAs) and target the same host genes? If the vesicular and non-vesicular exWAGO bind different guides, this would imply that they might target different host genes or affect the same genes but to a different extent.

AGO proteins bind sRNAs which act as guides directing the AGO protein to target genes through Watson-Crick base pairing. AGOs can bind different classes of sRNAs, including miRNAs, siRNAs and secondary small/short interfering RNAs (secondary siRNAs), as discussed previously (Chapter 1). Which class(es) of sRNA(s) does exWAGO bind? Phylogenetic analysis suggests that the *C. elegans* orthologues of exWAGO are SAGO-1, SAGO-2 and PPW-1 (Chow *et al*, 2019). SAGO-1, SAGO-2 and PPW-1, along with some other *C. elegans* AGO proteins, belong to the Worm-specific Argonaute (WAGO) clade of Argonautes which are also referred to as Secondary AGOs as they bind secondary siRNAs (Yigit *et al*, 2006; Gu *et al*, 2009). The major class of secondary siRNAs in *C. elegans* are termed 22G RNAs. 22G are 22 nt in length, have a Guanosine (G) as their first nucleotide and a 5'triphosphate (5'PPP) feature. Secondary siRNAs function to amplify the silencing signal (Yigit *et al*, 2006; Aoki *et al*, 2007; Pak & Fire, 2007; Gu *et al*, 2009) which can compensate

for inadequate amounts of primary siRNAs required for systemic RNAi and allows transgenerational inheritance of the silencing signal (Ketting & Cochella, 2020). Based on this information, we hypothesise that exWAGO might also act as a Secondary AGO and bind sRNAs with similar features to those of 22G RNAs.

Initial characterisation of the *H. bakeri* EVs secreted by adult worms revealed that the EVs carry miRNAs and Y-RNAs (Buck *et al*, 2014). More recent work by Chow *et al*, (2019) who employed a polyphosphatase (polyP) enzymatic treatment on the total sRNA population purified from EVs showed that the most abundant vesicular sRNAs are in fact secondary siRNAs. The polyP treatment converts the 5'PPP terminus to 5' monophosphate (5'P) thus permitting successful ligation of the 5' adapter required for sequencing. Hence, this allows sequencing of not only 5'P molecules, such as miRNAs and Y-RNAs, but also molecules that have a 5'PPP feature too, such as the secondary siRNAs. Further investigations by Chow *et al*, (2019) showed that the most abundant sRNAs found both in the *H. bakeri* EVs and the *H. bakeri* adult worms have three main properties: (1) they have a 5'PPP feature, (2) they are 23±1 nt in length, and (3) they start with a Guanosine (G) (Chow *et al*, 2019). The 5'PPP feature suggests that these sRNAs are products of RNA-dependent RNA polymerases and classifies them as secondary siRNAs. Inside *H. bakeri* adult worms, most sRNAs map to regions of the genome annotated as antisense to mRNA and antisense to retroelements, hence are hypothesised to be involved in endogenous gene regulation and control of retroelements (Chow *et al*, 2019). Interestingly, the sequences detected to be enriched in the *H. bakeri* EVs compared to the sRNAs in the adult worms map to novel repeats and transposons (Chow *et al*, 2019).

The identification of secondary siRNAs inside *H. bakeri* adult worms and *H. bakeri* EVs, and the presence of exWAGO which is homologous to AGO proteins of the WAGO clade in *C. elegans*, generated the hypothesis that exWAGO binds 5'PPP secondary siRNAs. At the time of starting this thesis, it was known that exWAGO is present in the vesicles, but not known whether it directly associates with vesicular RNAs. To understand if exWAGO functions in gene silencing when inside the host, we first focused on determining the RNAs that it binds.

In this chapter, I identify the sRNAs that are loaded in exWAGO at the different environments it exists, with a focus on the secreted forms of exWAGO. The overall aim was to define the sRNA cargo exWAGO associates with in different environments, in order to understand if it possesses specificity and whether the vesicular and

non-vesicular forms of exWAGO might target the same genes. The work is broken down into generation and analysis of small RNA sequencing data (Section 3.1) and validation of the selectivity of exWAGO *in vitro* using gel shift assays (Section 3.2).

Results

3.1 Defining the small RNA cargo of exWAGO

3.1.1 Generating sRNA libraries

To characterise the sRNA guides bound by exWAGO in the different environments it exists in, including (1) inside the parasite, (2) outside the parasite inside EVs (vesicular form) and (3) outside the parasite but not encapsulated inside vesicles (non-vesicular form), I immunoprecipitated exWAGO:sRNA complexes from adult worms, isolated EVs and EV-depleted HES. The success of the IPs was confirmed using Western blot analysis (Fig. 3.1).

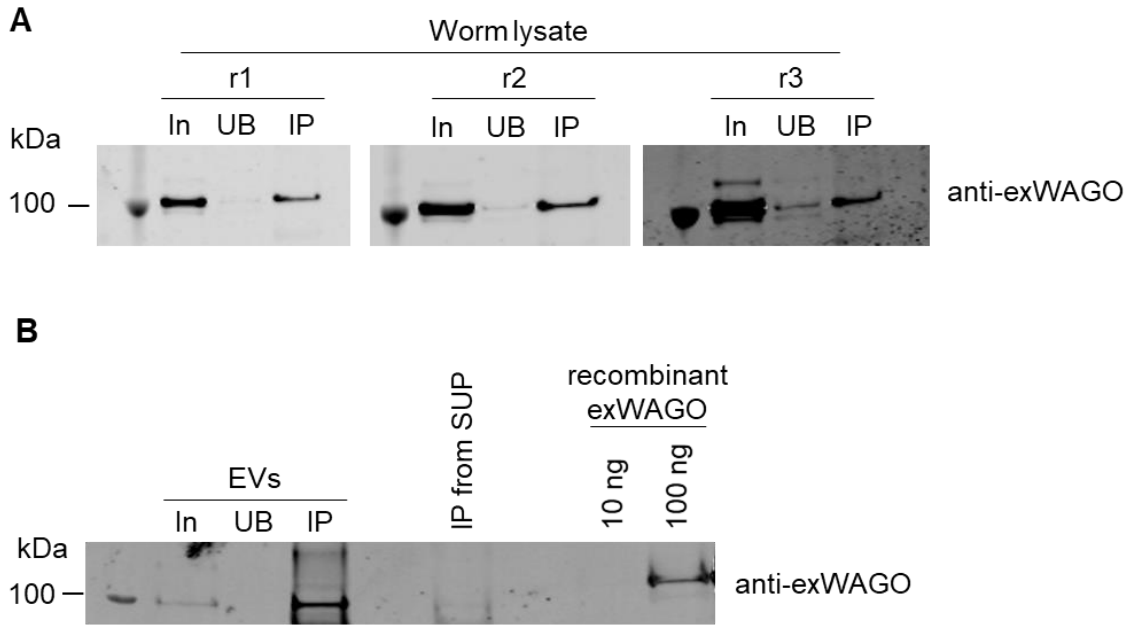


Figure 3.1 | Western blot analysis of exWAGO IPs from adult worm lysate, EVs and EV-depleted HES. A) exWAGO was pulled down from adult worms. Per IP, 150 µg of adult worm lysate protein was used with 75 µl of Protein G beads conjugated with 25 µl of rat anti-exWAGO antibody. Percentage of fraction analysed: 1% input (In); 1% unbound (UB); 5% eluate (IP). r1-3 = biological replicates. B) exWAGO IP from EVs (17 µg) or EV-depleted HES (SUP) (150 µg). For the IPs from EVs, 50 µl of Protein G beads conjugated with 10 µl of rat anti-exWAGO antibody were used per sample. A representative IP from EVs is shown here on the western blot. Percentage of fraction analysed: 3% unbound (UB); 66.7% eluate (IP). In = input (1.2 µg). For the IPs from SUPs (170 µg), 50 µl of Protein G beads conjugated with 10 µl of rat anti-exWAGO antibody were used per sample. A representative IP from SUP (150 µg) is shown here on the western blot. Percentage of fraction analysed: 6.67% eluate (IP). The western blots were probed with rabbit anti-exWAGO antibody (1:2,000) followed by goat anti-rabbit IgG Dylight 800 (1:10,000).

The sRNAs bound to exWAGO were purified and subjected to enzymatic treatments as detailed in Table 3.1. As previous data published by Chow *et al*, (2019) indicates that *H. bakeri* adult worms generate secondary siRNAs and as exWAGO is hypothesised to bind secondary siRNAs similarly to other WAGO proteins, I subjected all the purified sRNAs to polyP treatment to capture both 5'P and 5'PPP sRNAs (Table 3.1). For the samples generated following exWAGO IPs from adult worms, I also kept half of the purified sRNAs from the same samples untreated (Table 3.1). This means that only 5'P molecules will be captured during sequencing when no treatment is applied. Comparison of polyP-treated and untreated sRNAs will allow us to define the exact sRNA population exWAGO binds inside adult worms.

Moreover, the sRNAs from the total excretory/secretory products of the parasite (HES) were also purified to allow us to put into context what sRNAs the extracellular forms of exWAGO preferentially bind. The sRNAs underwent various enzymatic treatments prior to generation of the libraries (Table 3.1). One third of the purified sRNAs were treated with polyP enzyme as mentioned above to capture both 5'P and 5'PPP molecules (Table 3.1). The other third of the sRNAs remained untreated to capture only 5'P molecules, while the remaining sRNAs were treated first with terminator 5'-phosphate-dependent exonuclease treatment followed by polyP treatment (this treatment will be referred to as terminator treatment) (Table 3.1). The terminator enzyme degrades 5'P molecules leaving 5'PPP sRNAs intact. The polyP treatment then allows the 5'PPP guides to be sequenced. With these treatments, we expect to identify the types of different sRNAs that are secreted by the parasite and whether they have a 5'PPP modification. It is important to remember however that enzymatic treatments may not be 100% efficient. Thus, for the terminator treatment for example, we do not expect it to eliminate all 5'P molecules but rather reduce the number of 5'P molecules to a great extent, and this will be reflected by an enrichment in 5'PPP molecules.

Table 3.1 | Summary of the sRNA libraries generated. PolyP = RNA 5' polyphosphatase enzyme; terminator = terminator 5'P-dependent exonuclease.

Starting material		RNA treatment	No. of biological replicates	Index used
Worm Lysate	IP	polyP	3	25-27
		n/a	3	28-30
EVs	IP	polyP	3	31-33
	Unbound		2	35-36
EV-depleted HES (Sup)	IP	polyP	3	37-39
	Unbound		3	40-42
HES	Total	n/a	2	10-11
	Total	polyP	2	12-13
	Total	terminator followed by polyP	2	14-15

Following the enzymatic treatments, the purified sRNAs were then used to generate sRNA libraries for sequencing. The profiles of each sRNA library prior to sequencing were assessed using either TBE gels or the High Sensitivity DNA Bioanalyser chip (Fig. 3.2A&B). Figure 3.2A indicates that all the sRNA libraries have a band just below 150 bp which is deemed to correspond to sRNAs (lanes 1-3, 5-18), except for one of the EV unbound samples (EV UB 1, lane 4). It is thought that addition of one of the adapters was not added due to human error, based on the presence of a band around 60 bp thought to correspond to adapter monomer and the absence of a band at 120 bp thought to correspond to adapter dimers. Adapter dimers are considered a by-product formed during the generation of small RNA sequencing libraries. Adapter dimers form when the 5' adapter is ligated to the 3' adapter that had not been previously conjugated to an RNA molecule, because the adapter was in excess compared to the RNA material present for example (Shore *et al*, 2016). Removal of adapter dimers is important as these can overpower the sequencing signal and reduce the sequencing efficiency and quality of the sRNA library (Shore *et al*, 2016). Figure 3.2B clearly shows an enrichment in the amount of sRNAs and the detection of some other sRNA species detected only when the libraries exWAGO IP from adult worms were treated with polyP (lanes 2-4 compared to lanes 5-7). As polyP treatment allows us to capture 5'PPP, these results suggest that exWAGO binds 5'PPP molecules inside adult worms. Moreover, a sample generated previously in our lab containing adapter dimers only is run on the TBE gels and the High Sensitivity DNA Bioanalyser chip to help us identify the band corresponding to the adapter dimer by-products in our libraries (Fig. 3.2A, lane 19 & 3.2B, lane 1). The data show that all the libraries (except for the EV UB 1) have some adapter dimers, which will later be removed.

To obtain a similar coverage of sequencing depth in our libraries, I pooled 50 ng of each sRNA library per sample, based on information obtained from the High Sensitivity DNA Bioanalyser chip analysis prior to removal of adapter dimers. Where the concentration of the library was too low, I used as much material as possible (minimum amount used was 14.6 ng). To remove adapter dimers from our libraries we size selected the bands in which we expect to find sRNAs using gel purification. EV, EV-depleted HES, and HES libraries were size selected between 140-180 bp focusing on the sRNAs. The adult worm libraries were size selected between 140-220 bp, as according to quality control analysis gel of these libraries (Fig. 3.2B), there is a population of RNA at 200 bp (i.e. 80 bp when the adapters are omitted). Thus, we

included this population in our sequencing to identify what species of RNA this is, although the data presented in this thesis focuses on sRNAs of size 20-25 nt. After pooling of the libraries and removal of the adapter dimers, the quality of the final library was assessed using the High Sensitivity DNA Bioanalyser chip prior to sequencing (Fig. 3.2C, lanes 1-2).

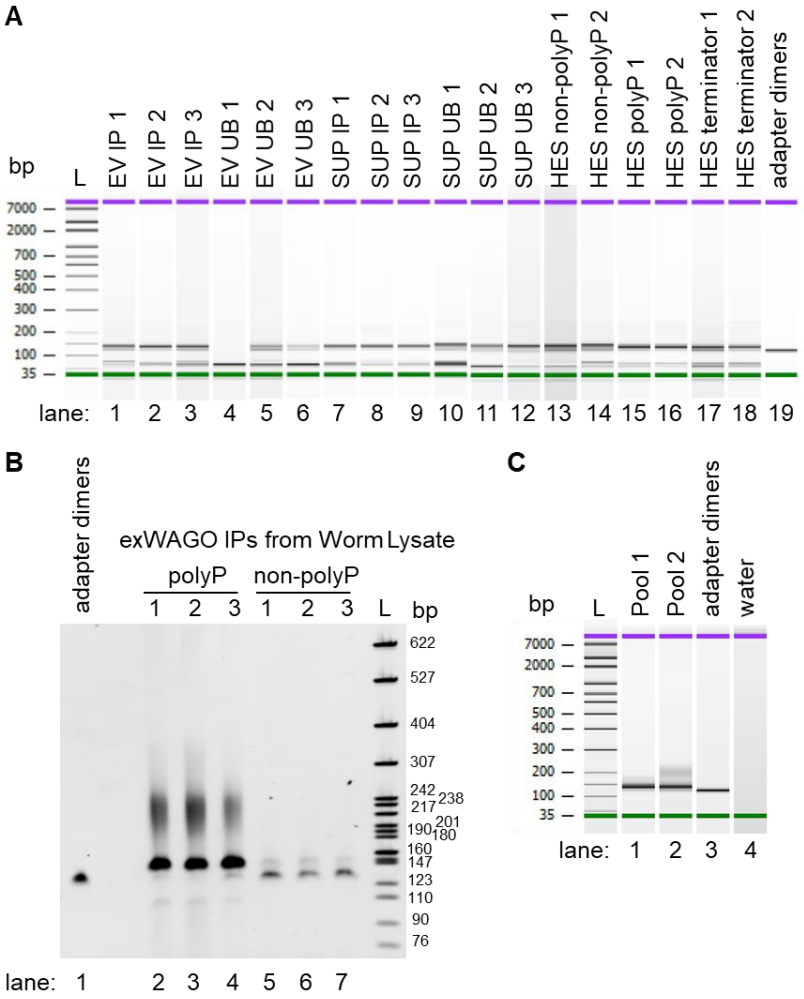


Figure 3.2 | Analysis of the profile of the sRNA libraries generated prior and following size selection using the Bioanalyzer system or TBE gels. A) Bioanalyzer High Sensitivity DNA analysis of the libraries (0.25%) generated using RNA from HES, and from exWAGO IP from EVs and EV-depleted HES (SUP), prior to size selection. UB = unbound, IP = eluate, 1-3 = biological replicates. B) TBE gel of the libraries (5%) generated from RNA material following exWAGO IP from adult worm lysate prior to size selection. 1-3 = biological replicates. C) Bioanalyzer High Sensitivity DNA analysis of the libraries (approximately 1 ng/µl) after size selection. Pool 1 is made from IPs or UB material from EVs and SUP, and total HES. Pool 2 is comprised of the IPs from adult worms. L = ladder, polyP = polyphosphatase-treated, non-polyP = non-polyphosphatase treated, terminator = terminator 5'P-dependent exonuclease -treated.

The libraries were sequenced using a single-end 100 base pair sequencing run on the Illumina NextSeq 2000 platform by the Edinburgh Clinical Research Facility. Sequencing yielded in 9-13 million reads per library of good quality (Table 3.2).

Table 3.2 | The number of reads with high quality score sequenced per library. Quality score greater than 30 means that the probability for calling a base incorrectly is less than 1 in 1000, while the accuracy of base calling is greater than 99.9%. PolyP = polyphosphatase treated; terminator = terminator 5'P-dependent exonuclease.

Starting material	RNA treatment	Index used	No. of reads with quality score > 30
Worm Lysate (IP)	polyP	25	12,508,583
		26	12,735,863
		27	10,125,225
	n/a	28	10,739,917
		29	9,718,302
		30	6,644,007
EVs (IP)	polyP	31	12,031,822
		32	11,434,700
		33	9,454,498
EVs (Unbound)	polyP	35	13,055,961
		36	11,585,632
EV-depleted HES (IP)	polyP	37	10,543,154
		38	10,623,895
		39	12,284,855
EV-depleted HES (Unbound)	polyP	40	11,108,583
		41	10,561,469
		42	8,818,204
HES (Total)	n/a	10	11,138,125
		11	12,665,847
	polyP	12	9,397,885
		13	11,287,222
	terminator followed by polyP	14	10,054,212
		15	9,065,802

Processing of the sequencing data was performed by Dr. Cei Abreu-Goodger. Briefly, only reads of 20-25 nt in length following adapter removal and that mapped to the *H. bakeri* genome with up to 1 nt mismatch were kept for further analysis. As we expect minimal contamination from mouse material, the 1 nt mismatch allowed during mapping accounts for potential sequencing errors. Figure 3.3A shows that approximately 80-90% of the reads per library map to the worm genome with up to 1 nt mismatch. However, only 65-70% of the reads from the libraries generated from exWAGO IPs from adult worms where the sRNAs did not undergo enzymatic

treatment map to the worm genome with up to 1 nt mismatch. I then used the curated dataset for further analysis using the Degust interface (Powell, 2019). The Degust file generated can be found at https://degust.erc.monash.edu/degust/compare.html?code=129df539f9a57f6794dd202c4ecbe76b#/.

To explore the relationship of the samples in terms of their sRNA content, the reads from all the samples were analysed by Multidimensional Scaling (MDS) (Fig. 3.3B&C). The libraries that were not treated with any enzymes (Adult_WIP_mp & HES_total_mp) cluster together. This is not surprising as these libraries capture only 5'P molecules which are not expected to be as dominant as the 5'PPP molecules. Interestingly, the polyP- and terminator-treated libraries from total HES (HES_total_pp and HES_total_tpp) and the libraries from exWAGO IPs from adult worms (Adult_WIP_pp) cluster together. This suggests that the sRNA content of these libraries is similar, despite that the terminator treatment should reduce to a great extent 5'P sequences compared to polyP-only treatment which captures both 5'P and 5'PPP molecules. This supports the idea that the extracellular sRNAs and the molecules exWAGO binds are enriched with 5'PPP rather than 5'P sRNAs. Unexpectedly, the EV samples (EV_WIP_pp and EV_unbound_pp) also cluster together suggesting the unbound fraction of the exWAGO IP from EVs contains very similar sRNAs to the exWAGO-bound fraction. In contrast, the bound and unbound fractions of the exWAGO IP from EV-depleted HES (SUP_WIP_pp and SUP_unbound_pp) are clearly separated by the second dimension, implying that the sRNAs found in the non-vesicular fraction that do not associate with exWAGO are different to those bound by exWAGO. All EV-depleted HES samples, however, are placed away from the EV samples, implying that the sRNA content of these two extracellular environments is different. This prompts the hypothesis that the vesicular and non-vesicular exWAGO bind different guides and hence might target different genes.

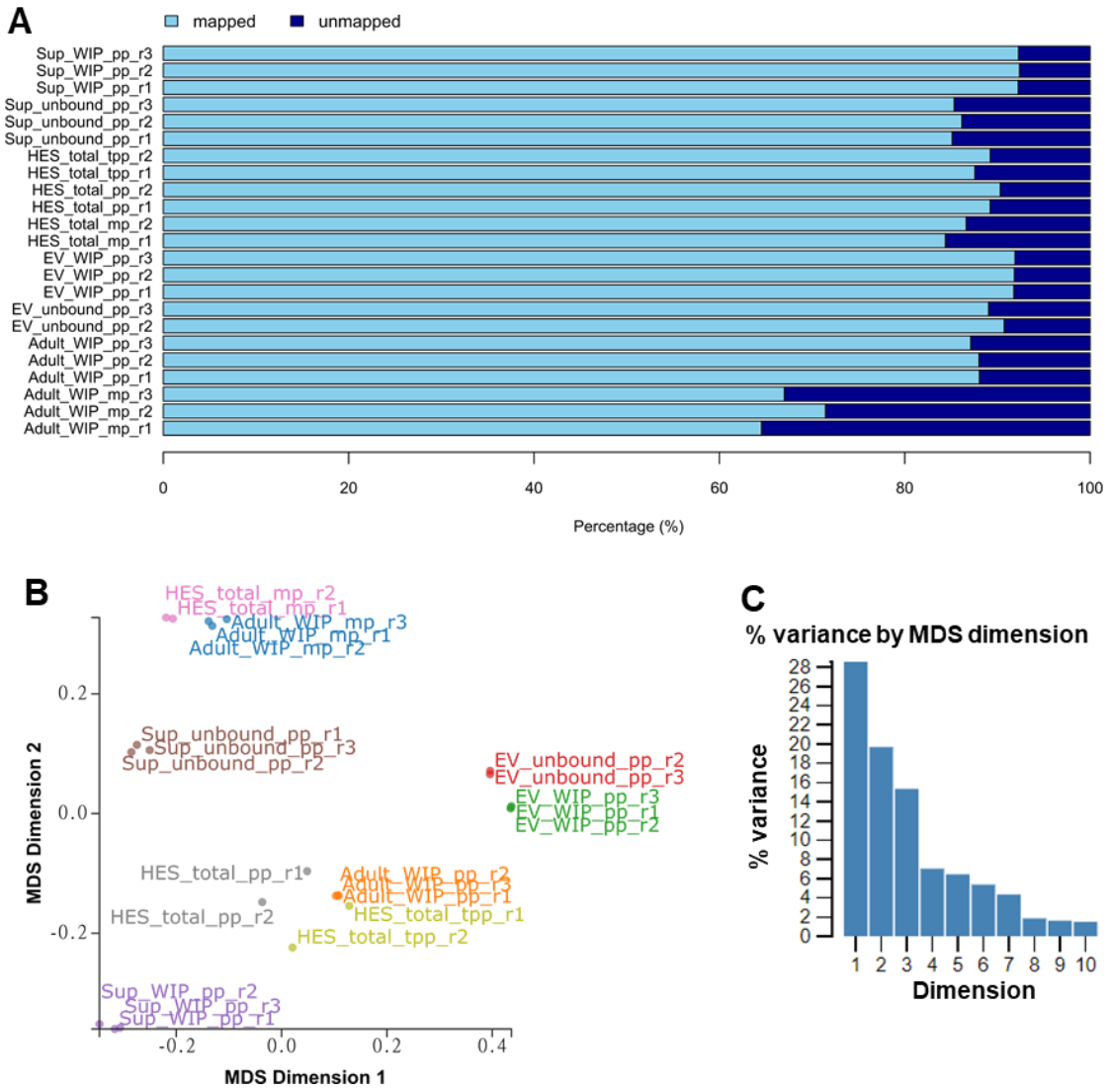


Figure 3.3 | Mapping and clustering of the sRNA data from HES and from exWAGO IPs from adult worms, EVs, EV-depleted HES.

A) The percentage of reads that map to *H. bakeri* genome (up to 1 nt mismatch) or that remain unmapped, when analysing only those reads that are 20-25 nt in length. The figure was generated by Dr. Cei Abreu-Goodger. B) MDS plot of all the sRNA libraries and C) Bar plot indicating the percentage variance by MDS dimension. SUP = EV-depleted HES, Adult = worm lysate, WIP = exWAGO IP, mp = non-polyP treated, pp = polyP-treated, tpp = terminator & polyP-treated, r1-3 = biological replicates 1-3.

3.1.2 Vesicular exWAGO-sRNA cargo

We previously reported that EVs secreted by *H. bakeri* carry 5'P miRNAs and Y-RNAs (Buck *et al*, 2014), but are in fact enriched with 5'PPP secondary siRNAs (Chow *et al*, 2019). These secondary siRNAs are 23±1 nt long, have Guanosine (G) as their preferred starting nucleotide and are preferentially derived from transposons and novel repeats (Chow *et al*, 2019). As exWAGO is secreted in EVs, I asked whether exWAGO directly associates with the sRNAs within the vesicles. exWAGO:guide complexes were immunoprecipitated from isolated EVs and the co-purified sRNA sequences were quantified using RT-qPCR from both the bound and unbound fractions of the IP. Per biological replicate (n = 2), 16.5 µg of EVs were used with 50 µl of Protein L beads and 10 µl anti-exWAGO antiserum. The sRNA sequences from loci nc16320, nc57384 and nc23553 are enriched in the exWAGO IP compared to the unbound fraction by ~9-, 14- and 1.5-fold respectively, while the Y-RNA-3p and miR-100 are depleted from the bound fraction (Fig. 3.4; data published in Chow *et al*, 2019). The data suggest that exWAGO preferentially binds secondary siRNAs rather than miRNAs or Y-RNAs. Additionally, these data may also indicate that exWAGO preferentially binds to siRNAs that carry a 5'PPP modification. These findings suggest that exWAGO may be involved in export and selectivity of sRNAs released by the parasite in vesicles.

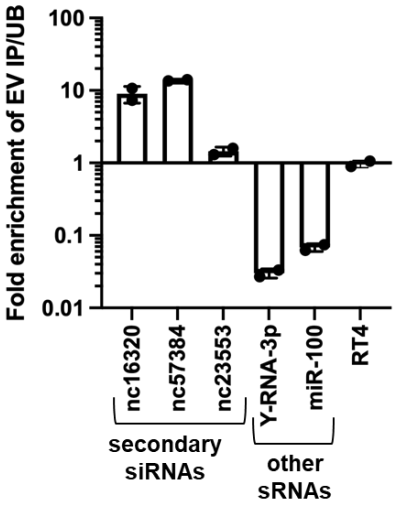


Figure 3.4 | RT-qPCR analysis of sRNAs following exWAGO IP from EVs. The fold enrichment of indicated sRNAs enriched in exWAGO IP from EVs compared to the unbound fraction. RT4 synthetic spike is used as an RNA extraction control. Data represent the average of biological duplicates ± S.D.

To obtain a global overview of the sRNA cargo that exWAGO is loaded with inside EVs, I generated sRNA libraries from the bound and unbound fractions of exWAGO IPs from 17 µg of purified EVs. 50 µl of Protein G beads with 10 µl of anti-exWAGO rat serum were used for each biological replicate (n = 3 for IP samples, n = 2 for the unbound fraction samples). These libraries were treated with polyP enzyme only. Length distribution and first nucleotide analysis indicates that the profile of exWAGO-bound and -unbound sRNAs is almost identical, with most reads starting with G and being 23 nt long (Fig. 3.5A). However, the unbound fraction also contains 22-23 nt reads that start with U (uridine), in contrast to the bound fraction. This length and starting nucleotide bias is often associated with miRNAs which have a 5'P feature (Seitz *et al*, 2011). These observations imply that the unbound fraction probably contains miRNAs which are not bound to exWAGO while both the bound and unbound fractions have sRNAs that start with G and are presumed to have a 5'PPP feature. In turn, this shows that exWAGO possesses molecular selectivity for 5'PPP molecules over 5'P inside EVs, even though a substantial amount of 5'PPP cargo seems to remain also in the unbound. To understand if this is true, we mapped the reads to annotated regions of the *H. bakeri* genome and compared them between the bound and unbound fractions. The analysis indicates that even though the sRNA cargo of the EVs contains miRNAs, Y-RNAs, tRNAs and rRNAs, exWAGO does not associate with these (Fig. 3.5B).

To identify specifically which sequences are preferentially bound by exWAGO inside EVs, I performed differential expression analysis which compares the fold change and average expression of each sRNA between the bound and the unbound fractions. 1,323 sRNAs are significantly enriched in the exWAGO bound fraction while 941 sRNAs are significantly enriched in the unbound fraction (FDR cut off = 0.01) (Fig. 3.5C). Consistent with the qPCR data (Fig 3.4), the Y-RNA-3p and miR-100 are depleted from the bound fraction of vesicular exWAGO. However, the secondary siRNAs from loci nc16320, nc57384 and nc23553 do not show an enrichment for the bound fraction (Table 3.3). The disagreement of the qPCR and sequencing data could be due to differences in normalisation. For example, normalisation for qPCR data is calculated as the fold enrichment of a particular sequence as calculated based of the Ct values obtained for that sequence in the bound and unbound fractions. On the other hand, the expression value calculated per sequence in sRNA sequencing data to then calculate the fold change is a function of the number of reads per sequence relative to the total number of reads sequenced across all samples. Thus, sRNA

sequencing data analysis takes into consideration the total population of sRNAs present in the bound and unbound fractions whereas the qPCR data analysis focuses on the particular sequence. Additionally, the normalisation method used in the analysis by Degust functions poorly if its assumptions are violated (Evans *et al*, 2018). The normalisation method assumes that differential expression across the samples compared is symmetrical and that technical effects (rather than biological effects) do not differ between the differentially expressed and non-differentially expressed sRNAs (Evans *et al*, 2018). In this case, comparing the bound and unbound fractions violates these assumptions, hence caution is required interpreting these results. Another explanation of the inconsistency between the qPCR and sRNA sequencing data would be that there is still a large amount of exWAGO remaining in the unbound fraction, which is not detected by western blot analysis due to sensitivity issues (Fig. 3.1B). This could also explain why we detect a substantial amount of 5'PPP cargo in the unbound fraction from the vesicular exWAGO IP.

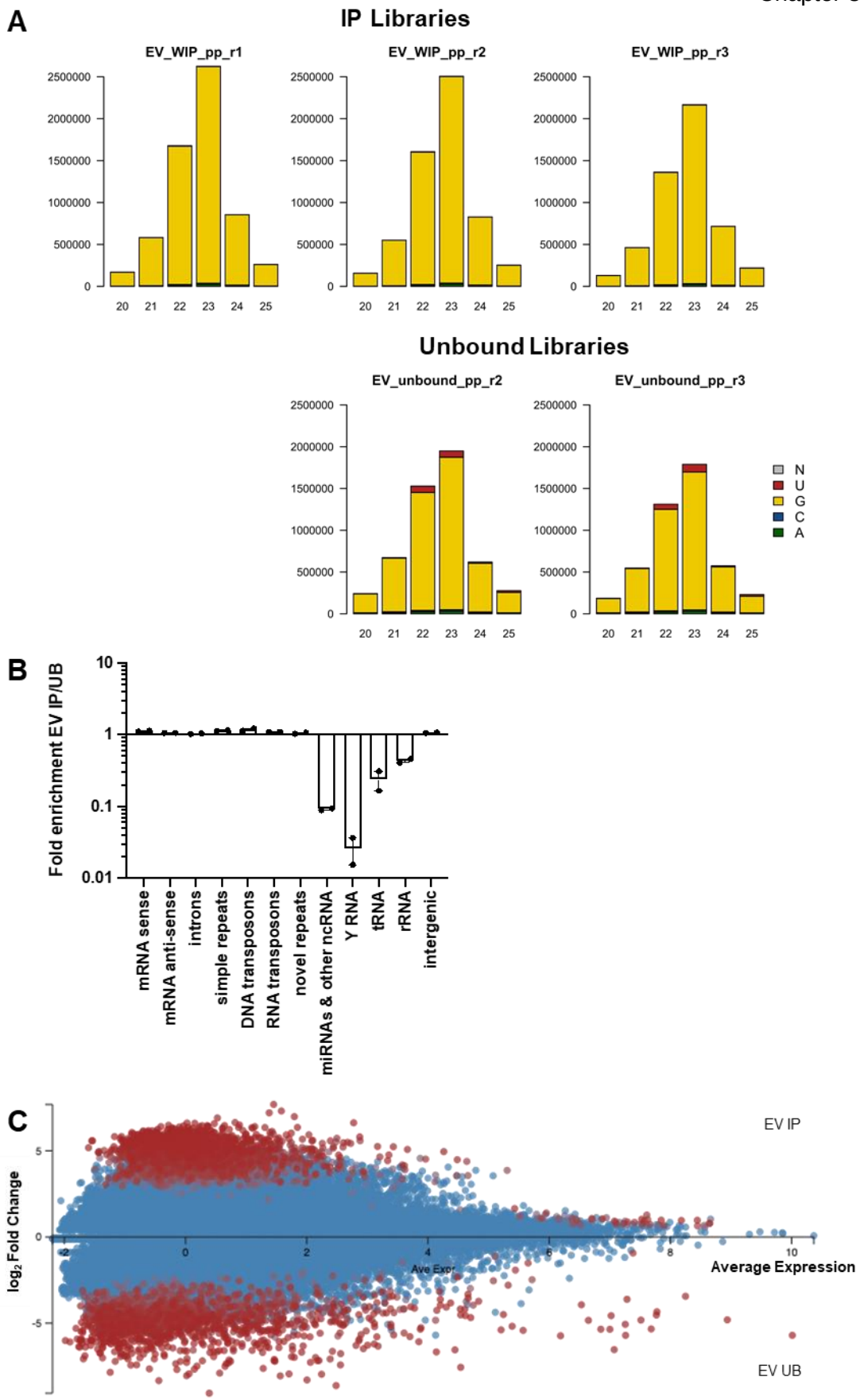


Figure legend on next page.

Figure 3.5 | The composition of sRNAs bound by exWAGO inside EVs (vesicular exWAGO) compared to the unbound.

A) Length distribution and first nucleotide plots of the sRNAs detected in EV exWAGO IPs and the unbound fraction. The x-axis represents the read length in nucleotides and the y-axis shows the number of reads. The figure was generated by Dr. Cei Abreu-Goodger. WIP = exWAGO IP, mp = non-polyP treated, pp = polyP-treated, r1-3 = biological replicates 1-3. B) Fold enrichment of 20-25 nt reads of IP/UB samples that map within annotated categories of the *H. bakeri* genome Data represent the average of 2-3 biological replicates ± S.E.M. C) Differential expression analysis of the reads between IP and UB samples. Data represent the average of 2-3 biological replicates. Average Expression is calculated as the mean of normalised reads in CPM across all samples and plotted on a log2 scale. sRNAs with FDR ≤ 0.01 are shown in red.

Table 3.3 | Summary table of information from the sRNA sequencing data regarding the sequences that were analysed in the RT-qPCR assay. Average Expression is calculated as the mean of normalised reads in CPM across all samples listed in Table 3.1 and is on a log2 scale.

		EV IP vs UB			SUP IP vs UB		
sRNA	Sequence	log2 FC EV IP/UB	FDR	Average Expression	log2 FC SUP IP/UB	FDR	Average Expression
nc16320	GATGACCAACCGGCTGTGGAAGC	0.073	0.969	10.366	2.120	0.000429	10.366
nc57384	GTAGTTGGGGTGGTTGTAGG	0.761	0.761	4.569	1.531	0.191	4.569
nc23553	GAACGACTGCTTCTATGCCACCCGA	-0.060	0.969	5.723	0.969	0.127	5.723
Y-RNA-3p	CGACAAAAGCTCGACCGGCGC	-4.166	0.0117	3.604	-11.683	7.650E-10	3.604
miR-100	AACCCGTAGATCCGAACTTGTGT	-4.519	4.05E-05	4.539	-8.486	2.090E-07	4.539

To place the above findings in context, I also sequenced the whole sRNA population that exWAGO binds inside the worm and the population of sRNAs bound by the exWAGO found outside of the worm in a vesicle-free environment (the non-vesicular form of exWAGO). We expect that a subset of the sRNAs in the adult worms will be exported to the extracellular environment of the worms inside or outside EVs.

3.1.3 sRNAs bound by exWAGO inside adult *H. bakeri* worms

To identify the sRNA guides bound to exWAGO inside the worm (intra-parasite exWAGO), I immunoprecipitated exWAGO from adult worms that had been 1 day in culture post-harvest from CBA x C57BL/6 F1 mice infected with *H. bakeri* for 14 days. For the IPs, 150 µg of adult worm lysate was used with 75 µl of Protein G beads and 25 µl of anti-exWAGO antiserum per replicate (n = 3 per condition). 85% of the IP was used for RNA extraction, after which the extracted RNA was divided in two equal parts. One part of the sample was treated with polyP enzyme, allowing us to capture 5'PPP molecules (Table 3.1) The other part of the sample remained untreated, meaning that only 5'P molecules were captured during sequencing (Table 3.1). Comparison of polyP-treated and untreated sRNAs libraries will allow us to define the exact sRNA population exWAGO binds inside adult worms. As previous research by Chow *et al*, (2019) showed that adult worms are enriched in 5'PPP sRNAs and exWAGO IPs from *H. bakeri* EVs showed the exWAGO is bound to 5'PPP sRNAs, we hypothesise that exWAGO will also be bound to secondary siRNAs inside the worms.

Indeed, our results indicate that exWAGO is bound mainly to 5'PPP sequences whilst still inside adult worms. This is shown by an increase in the amount of sRNAs present after the polyP treatment compared to the untreated libraries when analysed on a TBE gel (Fig. 3.2B). Length and category annotation analysis of 20-25 nt-long reads that map to annotated regions of the *H. bakeri* genome with 1 nt mismatch reveal that exWAGO is bound to sequences that start with G regardless of any enzymatic treatments (Fig 3.6A). It would be of interest in the future to understand if exWAGO has a preference to bind sequences that start with G over sequences that start with other nucleotides. Calculating the fold enrichment of reads mapping to annotated categories of the *H. bakeri* genome from polyP-treatment, suggests that the 5'PPP reads exWAGO binds come from regions of the *H. bakeri* genome annotated as DNA and RNA transposons and novel repeats (Fig. 3.6B). Differential expression analysis shows that 75,938 sequences are enriched during treatment with polyP, while only 5,389 are enriched in the untreated samples compared to polyP samples (FDR cut off = 0.01) (Fig 3.6C). The sequences enriched in the polyP-treated samples are hypothesised to possess a triphosphate modification at their 5' end. Interaction of exWAGO with 5'PPP guides which are considered to be secondary siRNAs due to their 5'PPP signature, suggests exWAGO is a Secondary AGO protein, similarly to its *C. elegans* orthologs SAGO-1, SAGO-2 and PPW-1.

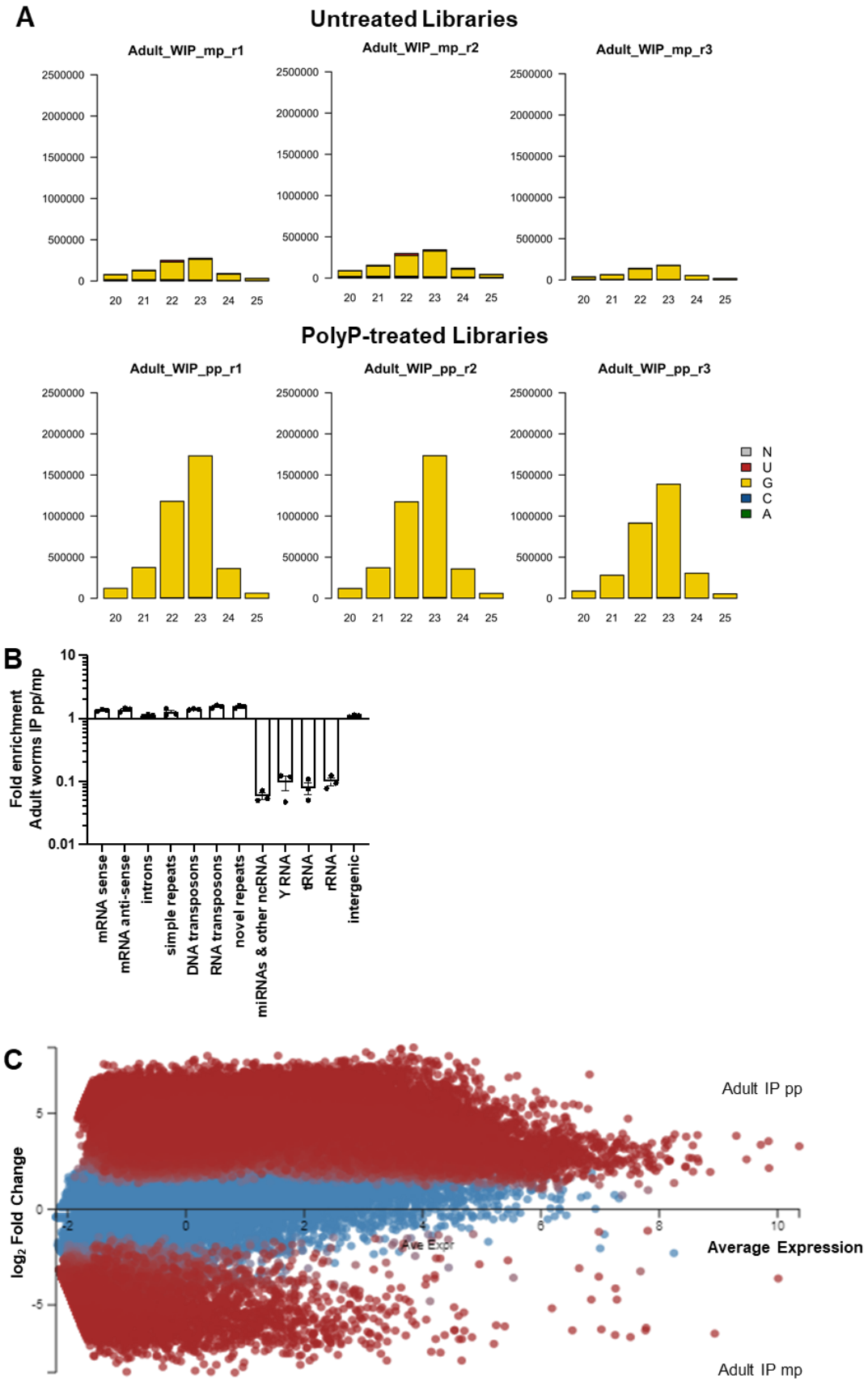


Figure legend on next page.

Figure 3.6 | The composition of sRNAs bound by exWAGO inside *H. bakeri* adult worms (intra-parasite exWAGO).

A) Length distribution and first nucleotide plots of the sRNAs bound by exWAGO inside adult worms with or without polyP treatment. The x-axis represents the read length in nucleotides and the y-axis shows the number of reads that map to the *H. bakeri* genome. The figure was generated by Dr. Cei Abreu-Goodger. Adult = worm lysate, WIP = exWAGO IP, mp = non-polyP treated, pp = polyP-treated, r1-3 = biological replicates 1-3. B) Fold enrichment of 20-25 nt reads of polyP/non-polyP samples that map within annotated categories of the *H. bakeri* genome. Data represent the average of 3 biological replicates ± S.E.M. C) Differential expression analysis of the reads between polyP- and non-polyP-treated samples. Data represent the average of 3 biological replicates. Average Expression is calculated as the mean of normalised reads in CPM across all samples and plotted on a log2 scale. sRNAs with FDR ≤ 0.01 are shown in red.

3.1.4 Non-vesicular exWAGO-sRNA cargo

exWAGO is not only secreted by the parasite inside EVs, but it also exists in a vesicle-free form. We hypothesise that the non-vesicular exWAGO is susceptible to proteases in contrast to the vesicular form of exWAGO which is protected by the lipid bilayer of EVs (Chow *et al*, 2019). However, the protein itself can act to protect the sRNA cargo it carries from RNases (Kim *et al*, 2020). It currently remains elusive whether the non-vesicular exWAGO is internalised by host cells, hence it is unknown whether it is functionally important in parasite-host communication. We know however, that factors of the EV-depleted HES mediate changes in the host immune responses. This was shown by suppression in gene expression of alternative activation markers when Bone Marrow-Derived Macrophages (BMDMs) were incubated with EV-depleted HES compared to incubation with media only (Coakley *et al*, 2017). Factors from the EV-depleted HES that could facilitate such changes in gene expression could be exWAGO and sRNAs.

How do the sRNAs bound to the non-vesicular exWAGO relate to the sRNAs bound by the vesicular exWAGO? To understand, therefore, if these two extracellular forms of exWAGO bind the same cargo, I generated sRNA libraries after exWAGO IP from EV-depleted HES (non-vesicular fraction of the total HES). exWAGO:sRNA complexes were immunopurified from 170 µg of EV-depleted HES. The EV-depleted

HES was obtained from the corresponding batches of HES that were used to purify the EVs used in Section 3.1.2, following ultracentrifugation. Per biological replicate (n = 3), 50 µl of Protein G beads with 10 µl of anti-exWAGO rat serum was used. The sRNAs purified from the exWAGO bound and unbound fractions were treated with polyP enzyme only (Table 3.1). Length distribution and first nucleotide plots show that the non-vesicular exWAGO binds sRNAs that start with G, similarly to those the vesicular exWAGO binds (Fig. 3.7A). In contrast, most reads are 22 nt in length whereas most of reads bound by the vesicular exWAGO are 23 nt long (Fig. 3.7A). Fold enrichment analysis of the reads found in the IP compared to the unbound that map within annotated categories of the *H. bakeri* genome indicates that the non-vesicular exWAGO is depleted of reads mapping to miRNAs, Y-RNAs, tRNAs and rRNAs, similarly to the vesicular exWAGO (Fig. 3.7B). On the other hand, the non-vesicular exWAGO binds sRNAs that are derived from genomic regions annotated as DNA and RNA transposons, simple and novel repeats, and mRNAs. These results demonstrate that exWAGO possesses binding selectivity for specific sRNA species. This suggests that exWAGO might have a structural feature that allows it to bind these sRNAs and/or it might interact with another protein to facilitate the loading of the sRNA in the exWAGO protein, and/or that there is cellular/subcellular compartmentalisation of exWAGO-sRNA loading resulting in selective loading of sRNAs.

To identify specifically which sequences are preferentially bound by exWAGO outside EVs, I performed differential expression analysis using the fold change and average expression of each sRNA in the bound and unbound fractions. 45,070 sRNAs are preferentially bound by the non-vesicular exWAGO while 4,182 sRNAs remain preferentially unbound (FDR cut off = 0.01) (Fig. 3.7C). Examination of individual sequences indicates that Y-RNA-3p and the miR-100 are enriched in the unbound fraction, in agreement with the sequencing data and qPCR data from the vesicular exWAGO (Table 3.3). From the secondary siRNAs tested by qPCR, only the sRNA from the locus nc16320 is statistically enriched in the bound fraction (FDR cut off = 0.01). The rest of the sRNAs, although not statistically enriched, tend to have a positive Fold Change value of the IP versus unbound (Table 3.3). The differences in the normalisation between the qPCR and sequencing data, as well as the possibility of some exWAGO remaining in the unbound fraction, could explain the differences seen between the qPCR and sequencing data, as discussed in Section 3.1.2.

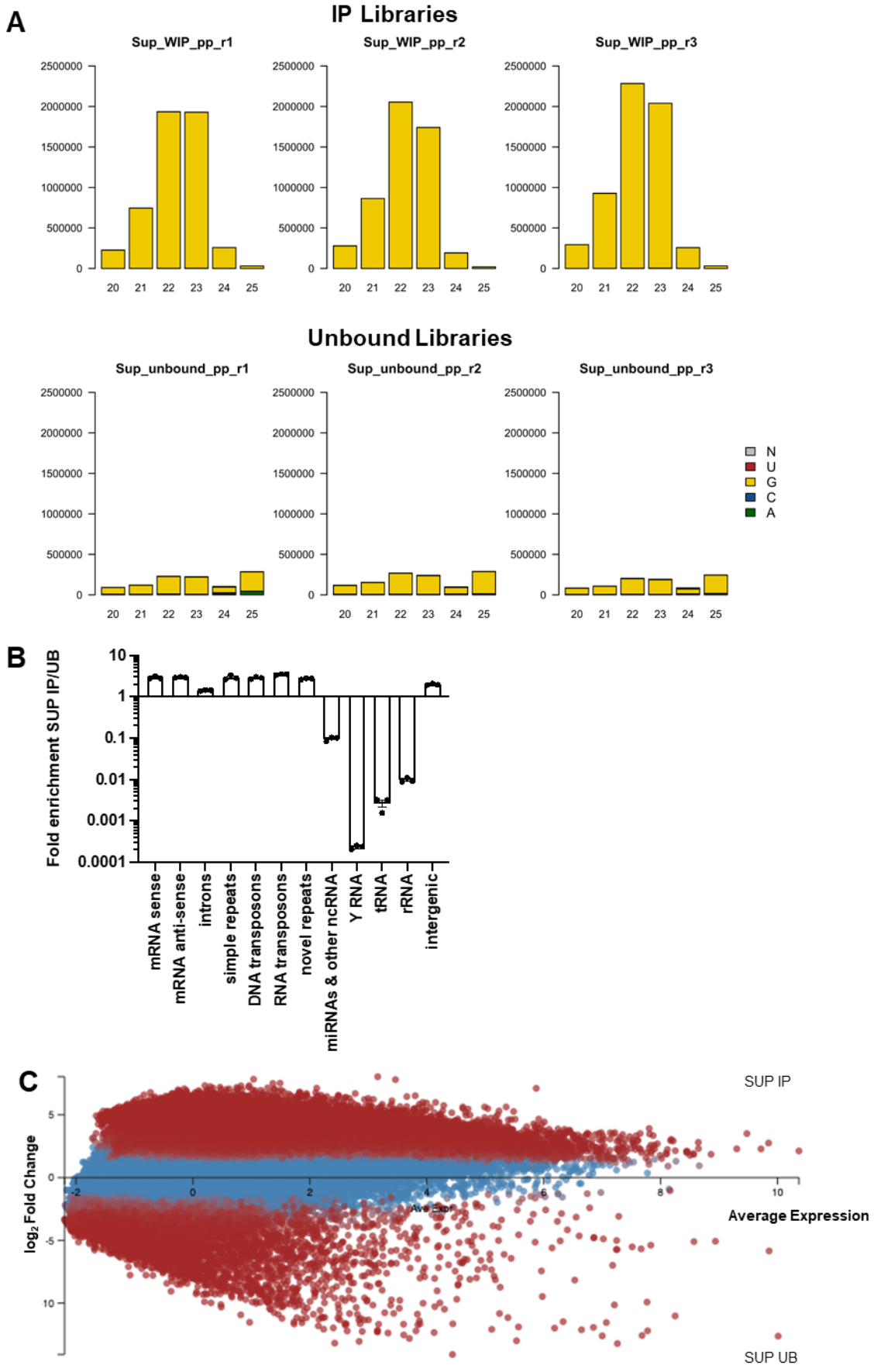


Figure legend on next page.

Figure 3.7 | The composition of sRNAs bound by exWAGO outside of EVs (non-vesicular exWAGO) compared to the unbound.

A) Length distribution and first nucleotide plots of the sRNAs bound by the non-vesicular exWAGO. The x-axis represents the read length in nucleotides and the y-axis shows the number of reads. The figure was generated by Dr. Cei Abreu-Goodger. SUP = EV-depleted HES; WIP = exWAGO IP, mp = non-polyP treated, pp = polyP-treated, r1-3 = biological replicates 1-3. B) Fold enrichment of 20-25 nt reads of IP/UB samples that map within annotated categories of the *H. bakeri* genome. Data represent the average of 3 biological replicates ± S.E.M. C) Differential expression analysis of the reads between IP and UB samples. Data represent the average of 3 biological replicates. Average Expression is calculated as the mean of normalised reads in CPM across all samples and plotted on a log2 scale. sRNAs with FDR ≤ 0.01 are shown in red.

3.1.5 Comparing the sRNAs bound by the vesicular and non-vesicular exWAGO

Analysis of the sRNA cargo bound by the vesicular and non-vesicular exWAGO indicates that they bind guides with similar properties (Sections 3.1.2 and 3.1.4). However, there are notable differences too. For example, most reads bound by the vesicular exWAGO are 23 nt long whereas most reads bound by the non-vesicular exWAGO are 22 nt long (Fig. 3.5A & 3.7A). Additionally, the unbound fraction of the vesicular exWAGO immunoprecipitation seems to be similar to what exWAGO binds too, whereas the unbound fraction of the non-vesicular exWAGO immunoprecipitation can be regarded as more diverse in terms of sRNA molecules. These differences, however, do not directly indicate whether the vesicular and non-vesicular forms of exWAGO bind the same guide sRNAs.

To understand whether the vesicular and non-vesicular forms of exWAGO bind the same sRNA cargo, I compared the sRNAs identified by exWAGO IP from EVs and EV-depleted HES directly. Identifying if the two extracellular forms of exWAGO bind the same guides will give us a hint as to whether they could also target the same genes. Therefore, I performed a differential expression analysis between all the reads identified to be bound by the vesicular and non-vesicular exWAGO protein. The results indicate that 59,132 sequences are enriched in the non-vesicular exWAGO IP (FDR cut off = 0.01), and 10,153 sequences are enriched in the vesicular exWAGO IP (FDR cut off = 0.01) (Fig. 3.8A). This finding indicates that the relative abundance of the guides identified to be bound to the vesicular and non-vesicular exWAGO varies. It also generates the hypotheses that (1) the two forms of exWAGO bind the same guides but not to the same extent in terms of relative abundance, and/or (2) they bind distinct sequences. Further analysis of the sequences indicates that both hypotheses stand. Inspection of the sequences bound by both the vesicular and non-vesicular exWAGO forms indicates that > 9,000 sequences are shared between the vesicular and non-vesicular exWAGO IPs but are deemed to be differentially expressed (EV and EV-depleted HES IPs samples > 0 CPM, $\log_2$FC ≤ -1 and $\log_2$FC ≥ 1, FDR cut off = 0.01). These results support the idea that the two forms of exWAGO bind the same guides but not to the same extent in terms of the relative abundance they are bound. Further analysis indicates that 1,888 sequences can be deemed exclusive to the non-vesicular exWAGO (EV IP counts = 0, EV-depleted HES counts > 10, $\log_2$FC ≥ 1, FDR cut off = 0.01) (Supp. Table 3.1), while only 43 sequences are

exclusive to the vesicular exWAGO (EV IP counts > 10, EV-depleted HES counts = 0, $\log_2$FC ≥ 1, FDR cut off = 0.01) (Supp. Table 3.2). Therefore, these data indicate that there are unique guides bound by each extracellular form of exWAGO, with the non-vesicular exWAGO binding more distinct guides that are not found in the vesicular exWAGO IP.

Multidimensional scaling analysis with exWAGO IPs from adult worms, EVs and EV-depleted HES was also performed and supports the differential expression analysis findings. The largest variance component between all samples (accounting for ~70%) separates the vesicular and non-vesicular exWAGO IPs, with the exWAGO IPs from adult worms clustering between them (Fig. 3.8B&C). The fact that the two forms of extracellular exWAGO bind guides that are unique to each form further suggests that the vesicular and non-vesicular forms of exWAGO may target different genes. Depending on how these guides would associate with their targets (e.g. via near-perfect complementarity, via seed pairing, etc) and considering that they also bind the same guide sequences but at a different relative abundance, we cannot disregard the hypothesis that these guides might also affect the same targets but to a different extent.

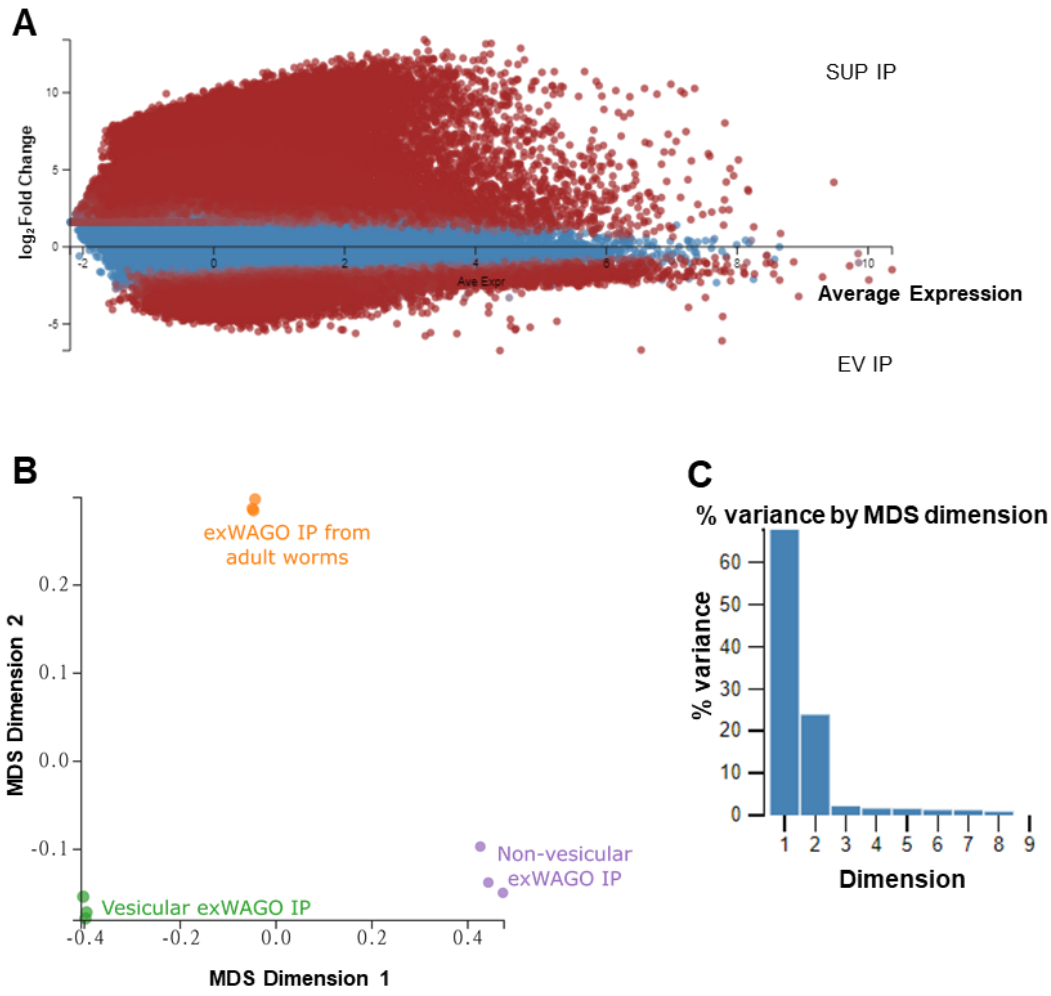


Figure 3.8 | The vesicular and non-vesicular exWAGO are not bound to identical sRNA cargo.

A) Differential expression analysis of the reads between SUP IP and EV IP samples. Data represent the average of 3 biological replicates. Average Expression is calculated as the mean of normalised reads in CPM across all samples and plotted on a log2 scale. sRNAs with FDR ≤ 0.01 are shown in red. B) MDS plot of the sRNA libraries from EV IPs, SUP IPs and worm lysate IPs. Data represent 2-3 biological replicates. C) Bar plot indicating the percentage variance by MDS dimension for the MDS plot shown in (B).

3.1.6 Extracellular sRNAs released by *H. bakeri*

The data presented so far in this chapter, show that exWAGO associates with 5'PPP secondary siRNAs and that the two extracellular forms of exWAGO can bind shared sequences but to a different relative abundance and can also bind unique sequences compared to each other. These findings suggest that these secondary siRNAs might guide exWAGO to targets inside the mouse host once the protein is internalised into host cells. To thus understand the population of extracellular sRNAs bound by exWAGO (vesicular and non-vesicular), we should also put this in context of the total sRNAs that are found in the *H. bakeri* excretory-secretory (HES) products.

RNA was extracted from 68 µg of HES from adult *H. bakeri* worms that were in culture for 1-8 days with a media change on day 4 (media after 24 h was discarded), per sample (n = 2, per condition). The experiment was designed to be able to determine the types of different sRNAs that are secreted by the parasite and whether they have a 5'P or a 5'PPP modification. The extracted RNA was divided in three equal parts where one part was treated with polyP enzyme, the other part was treated with terminator, and the last part remained untreated.

Multidimensional scaling analysis indicates that the polyP- and terminator-treated samples cluster together on the right-hand side of the graph (Fig. 3.9A&B). The untreated samples cluster on the opposite side of the graph. The clustering of these samples suggests that the sRNAs captured in the polyP-treated samples are similar to those captured in the terminator-treated samples. The only difference between these two samples is that the terminator treatment will degrade 5'P molecules, hence the 5'PPP will be sequenced. Length and first nucleotide plots indicate that most sRNAs secreted by *H. bakeri* start with a G (guanosine) despite the differences in the enzymatic treatments (Fig. 3.10). Closer investigation, however, shows that more 22 nt reads starting with U (uridine) are captured when no treatment is applied. This length and starting nucleotide bias is often associated with miRNAs (Seitz *et al*, 2011). Mapping the reads to annotated categories of the parasite genome shows that the annotation profile of the untreated samples differs to that of the polyP- and terminator-treated samples (Fig. 3.9C&D). The reads from the untreated samples are derived from miRNAs, Y-RNAs, tRNAs and rRNAs while the reads from the terminator and polyP treatments map to transposons and novel repeats.

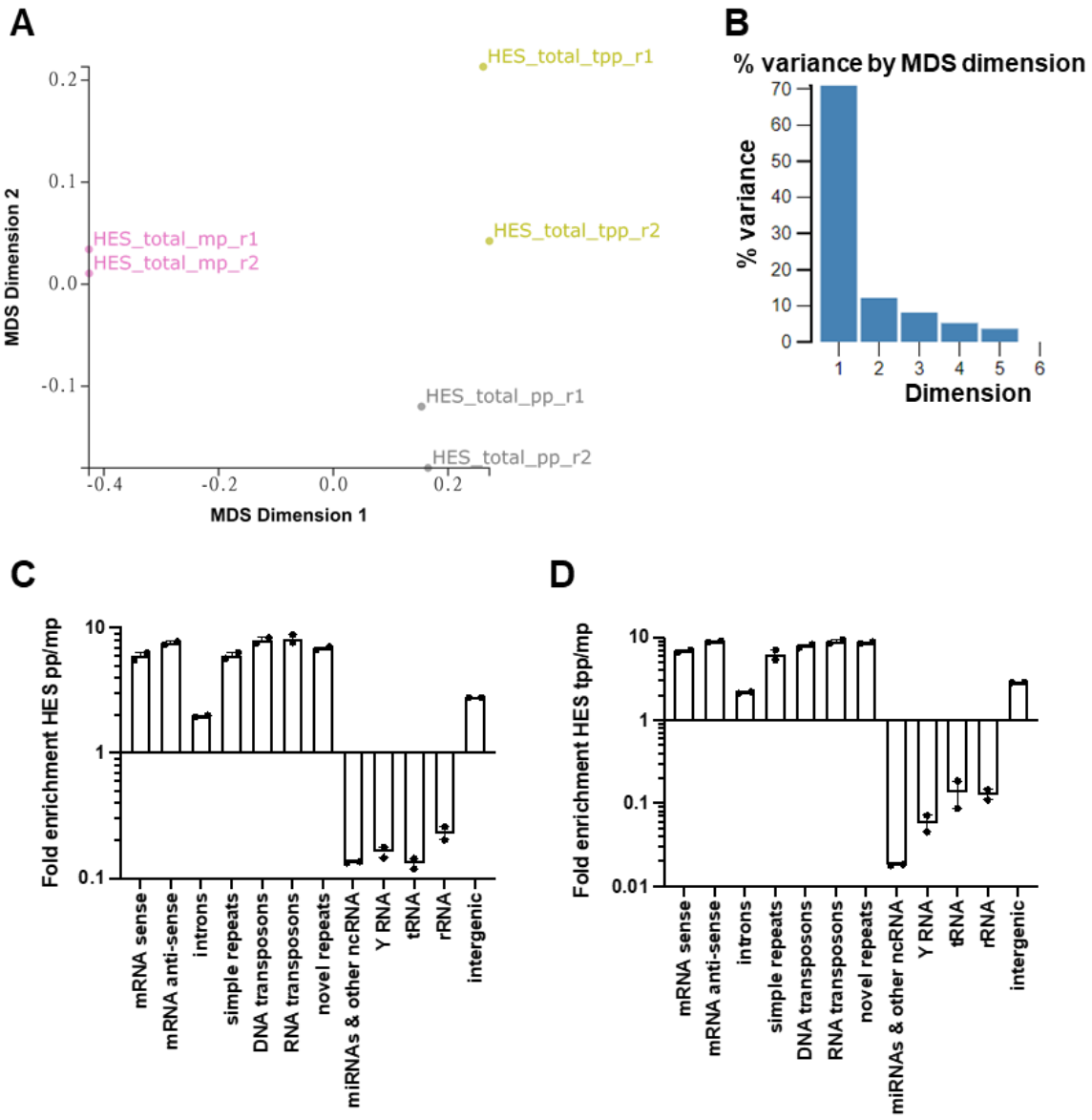


Figure 3.9 | Clustering analysis and the composition of sRNAs secreted by *H. bakeri* adult worms (HES).

A) MDS plot of all the sRNA libraries generated from HES. Data represent the average of 2 biological replicates. mp = non-polyP treated, pp = polyP-treated, tpp = terminator & polyP-treated, r1-3 = biological replicates 1-3. B) Bar plot indicating the percentage variance by MDS dimension for the MDS shown in (A). C & D) Fold enrichment of 20-25 nt reads of polyP- or terminator- samples over non-polyP-treated samples that map within annotated categories of the *H. bakeri* genome. Data represent the average of 2 biological replicates ± S.E.M.

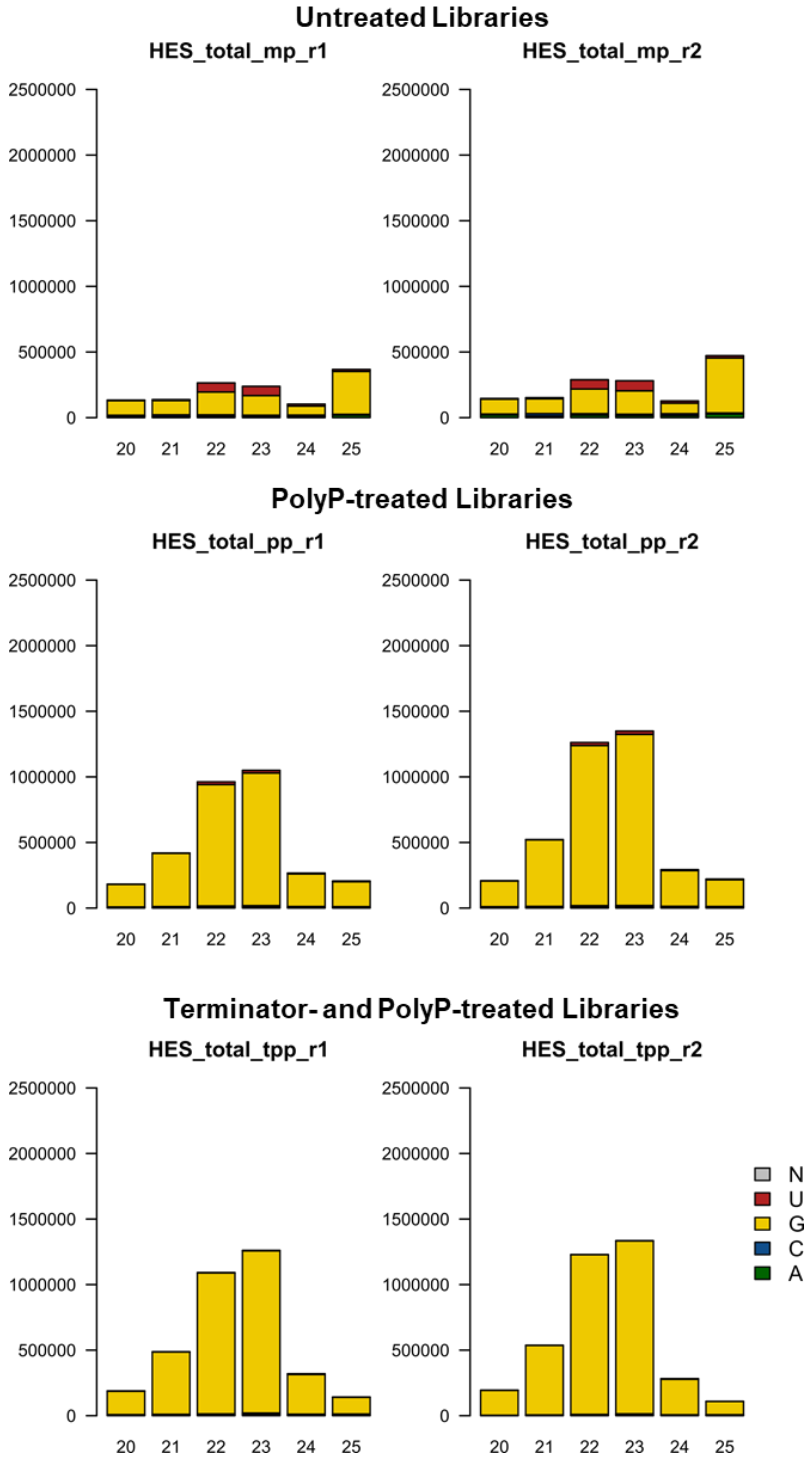


Figure 3.10 | Length distribution and first nucleotide plots of the reads sequenced from the HES samples that map to the *H. bakeri* genome. The x-axis represents the read length in nucleotides and the y-axis shows the number of reads. The figure was generated by Dr. Cei Abreu-Goodger. mp = non-polyP treated, pp = polyP-treated, tpp = terminator & polyP-treated, r1-3 = biological replicates 1-3.

To assess whether the terminator treatment could be used as a method to diminish 5'P molecules and thus enrich 5'PPP sequences compared to the polyP-only treatment, I performed a differential expression analysis. 1,075 sequences are enriched in the polyP-treatment compared to 61 enriched in the terminator-treated samples (FDR cut off = 0.01) (Fig. 3.11). The sequences depleted from the terminator-treated samples but enriched in the polyP-treated samples could reveal true 5'P sequences (such as miRNAs). To validate this hypothesis, we first mapped the sequences that are significantly enriched or depleted in the terminator treatment to annotated categories of the *H. bakeri* genome (FDR cut off = 0.01). Then I calculated the fold change of each category for these sequences to identify if specific categories are identified as enriched in the polyP- or terminator-treated samples. Interestingly, none of the reads that were enriched in the terminator-treated samples map to known miRNAs or Y-RNAs which have a 5'P terminus, consistent with the above hypothesis (Table. 3.4). Unexpectedly, the percentage of sequences derived from tRNAs is 38 times greater in the terminator-treated sequences that are significantly enriched over the polyP-treated sequences. tRNA fragments are not expected to carry 5'PPP unless these fragments contain the 5' leader sequence of tRNA precursors (Raina & Ibba, 2014). BLAST analysis of the tRNA sequences enriched in the terminator-treated samples suggests that these do not come from the 5' leader sequence of tRNA precursors but are rather derived from structured parts of the tRNA. These results could indicate that the terminator enzyme cannot access these highly structured parts of the tRNA hence the terminator treatment was not able to degrade them.

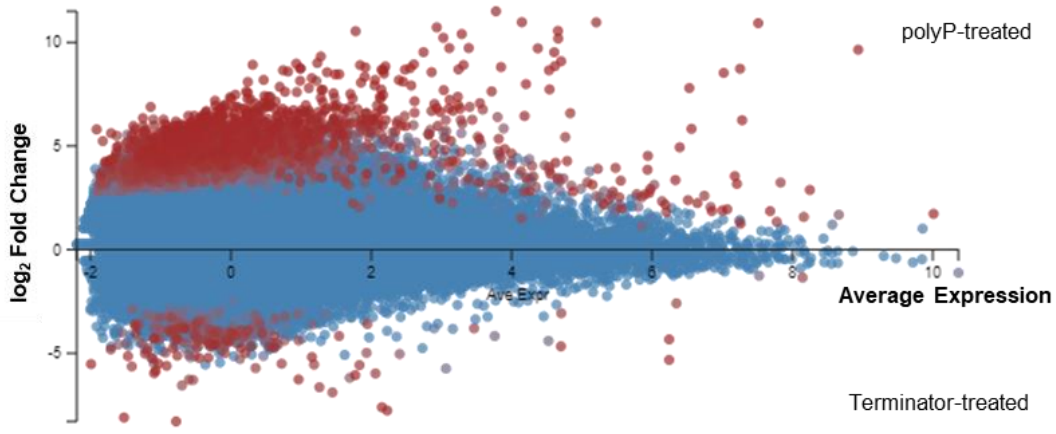


Figure 3.11 | Differential expression analysis of the reads between polyP- and terminator-treated HES samples. Data represent the average of 2 biological replicates. Average Expression is calculated as the mean of normalised reads in CPM across all samples and plotted on a log2 scale. sRNAs with FDR ≤ 0.01 are shown in red.

Table 3.4 | Summary table of enriched sequences from polyP- and terminator-treated HES samples that map to annotated categories of the *H. bakeri* genome. Sequences are defined as enriched by an FDR ≤ 0.01. AS = anti-sense; S = sense; rep = repeats; unk = unknown; snRNA = small nucleolar RNA.

Annotated Category	No. of polyP enriched sequences	No. of terminator enriched sequences	% of polyP enriched sequences	% of terminator enriched sequences	Fold Change terminator vs polyP
exons_AS	21	2	2	3.3	1.68
exons_S	6	1	0.6	1.6	2.94
intergenic	264	10	24.6	16.4	0.67
introns_AS	20	1	1.9	1.6	0.88
introns_S	15	0	1.4	0	0
miRNA_S	102	0	9.5	0	0
other_ncRNA_S	1	0	0.1	0	0
piRNA_AS	1	0	0.1	0	0
rep_dna_AS	46	3	4.3	4.9	1.15
rep_dna_S	8	1	0.7	1.6	2.2
rep_rna_AS	112	6	10.4	9.8	0.94
rep_rna_S	27	2	2.5	3.3	1.31
rep_simple	4	0	0.4	0	0
rep_unk_AS	23	2	2.1	3.3	1.53
rep_unk_S	41	3	3.8	4.9	1.29
rRNA_AS	155	11	14.4	18	1.25
rRNA_S	138	6	12.8	9.8	0.77
snRNA_S	29	0	2.7	0	0
tRNA_S	6	13	0.6	21.3	38.18
yRNA_AS	2	0	0.2	0	0
yRNA_S	54	0	5	0	0
Total	1075	61	100	100	

From the above data, we denote that the total extracellular RNA found in HES is not only made of secondary siRNAs but also contains other types of non-coding RNAs, such as miRNAs, Y-RNAs, tRNAs and rRNAs. These results are consistent with previous data presented in Sections 3.1.2 and 3.1.4, where exWAGO is found to be preferentially bound to 22G sequences despite the presence of other non-coding RNAs.

3.2 Exploring the binding properties of exWAGO using gel shift assays

The sequencing data clearly indicate that exWAGO binds to specific types of sRNAs. Regardless of whether exWAGO is inside or outside vesicles, it binds 22-23 nt, 5'PPP sRNA molecules starting with a G (secondary siRNAs), while it does not appear to bind miRNAs. To understand if the binding selectivity demonstrated by exWAGO is an intrinsic feature of the protein, I sought to study the binding selectivity of exWAGO *in vitro* using biochemical techniques.

3.2.1 exWAGO binds 5'PPP 23G RNAs with high affinity in contrast to mAGO2

The RNA-binding affinity of exWAGO was examined using concentrations of recombinant FLAG-tagged exWAGO ranging from 0 to 500 nM. Recombinant exWAGO was incubated with 5 nM of 5'PPP nc16320 secondary siRNA guide which was generated using a T7 *in vitro* transcription kit and radiolabelled with [32P]pCp at the 3' end. exWAGO:5'PPP nc16320 complexes were resolved on a native gel using an electrophoretic mobility shift assay (EMSA). Quantification of the dissociation equilibrium constant (Kd) indicates that exWAGO binds 5'PPP nc16320 with high affinity (Kd = 9.20 nM ± 1.03 S.E.M) (Fig. 3.12A&C, Table 3.5). This is comparable to the binding affinity of GST-tagged mouse AGO2 to 5'P let7a (Kd = 9.8 nM) as measured by Tan *et al* (2009). To determine if binding of 5'PPP molecules is a feature exhibited only by exWAGO in comparison to the mouse host AGO protein, I tested whether recombinant mouse AGO2 (mAGO2) can bind 5'PPP 23G sRNAs. The results show that, in contrast to exWAGO, mAGO2 binds the 5'PPP nc16320 guide with lower affinity (Kd = 56.69 nM ± 8.53 S.E.M.) (Fig. 3.12B&C, Table 3.5). Additionally, the maximum fraction of bound 5'PPP nc16320 guide at saturating AGO protein concentrations is 2.3-fold lower for the mAGO2 (Bmax = 0.43 ± 0.029 S.E.M.) compared to the exWAGO (Bmax = 0.99 nM ± 0.031 S.E.M.) (Fig. 3.12C, Table 3.5). These results highlight that exWAGO binds 5'PPP sRNAs with high affinity for an AGO protein compared to the mAGO2 which binds 5'PPP molecules less strongly.

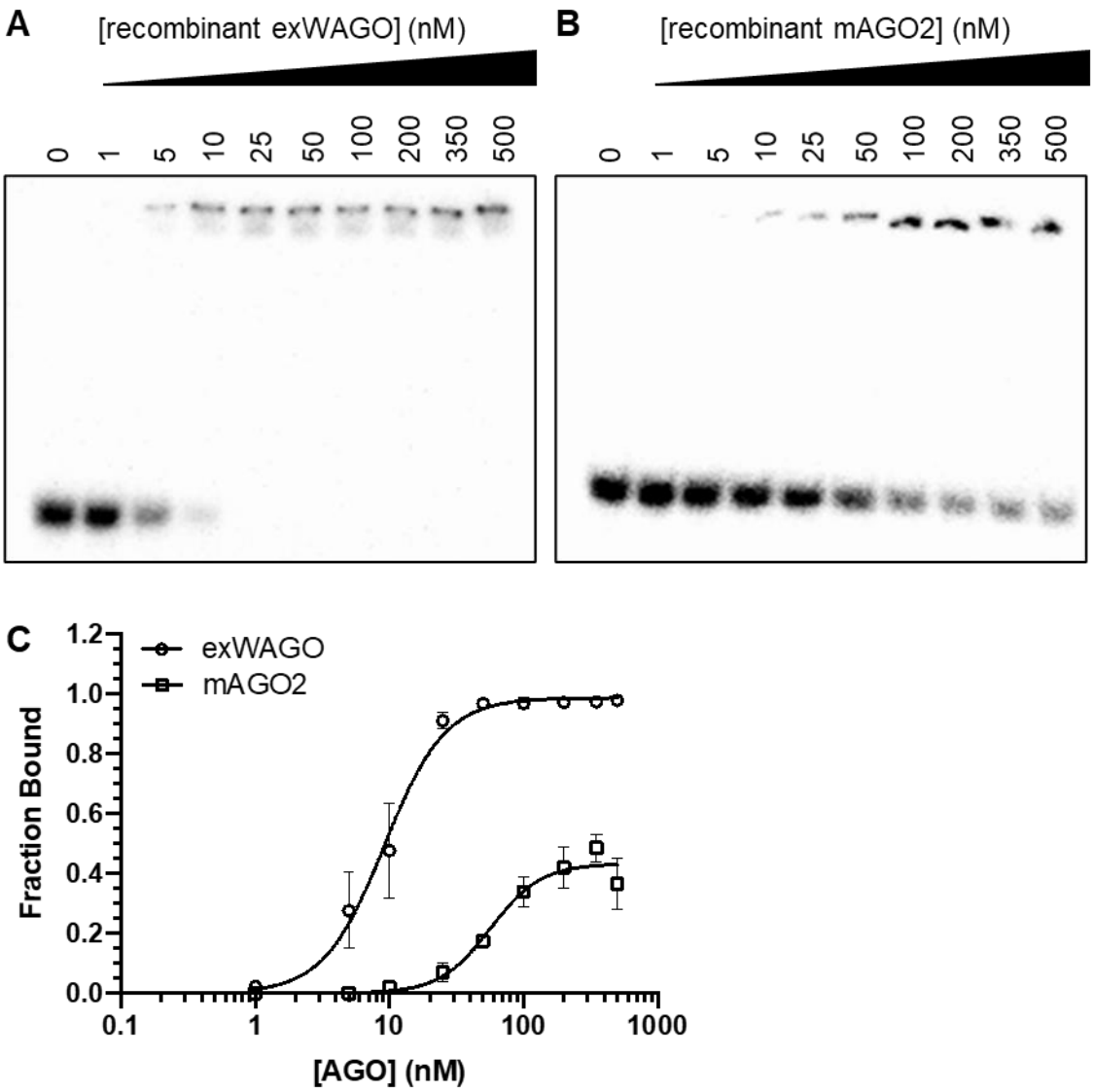


Figure 3.12 | exWAGO binds 5'PPP 23G sRNAs with high affinity whereas mAGO2 does not.

pCp-radiolabelled 5'PPP nc16320 (5 nM) was incubated with increasing concentrations (0-500 nM) of recombinant A) exWAGO or B) mAGO2. AGO:guide complexes were resolved on a native gel. C) Quantitative binding analysis of the fraction of guide bound by exWAGO and mAGO2 plotted against increasing concentration of the AGO proteins. Data were quantified using ImageJ and show the mean±S.E.M from two to six independent experiments.

Table 3.5 | Summary of Kd and Bmax values from the quantitative binding analysis of the fraction of guide bound by exWAGO or mAGO2. N is the number of replicates used; S.E.M. = Standard Error of the Mean.

Recombinant AGO protein	Radiolabelled sRNA guide	n	Kd (nM)	Kd S.E.M (nM)	Bmax	Bmax S.E.M
exWAGO	5'PPP nc16320	6	9.20	1.03	0.99	0.031
mAGO2	5'PPP nc16320	2	56.69	8.53	0.43	0.029
exWAGO	5'P nc16320 (polyP-treated)	6	18.63	2.32	0.89	0.034
exWAGO	5'P nc16320 (synthetic)	2	8.43	2.93	0.97	0.086
exWAGO	5'P miR100	2	3.26	0.42	0.99	0.021
mAGO2	5'P miR100	2	11.30	2.13	0.81	0.040

3.2.2 exWAGO binds 5’P sRNAs with high affinity *in vitro*

To understand if exWAGO has a binding preference for 5’PPP sRNAs compared to 5’P molecules, I generated a 5’P nc16320 molecule by treating the 5’PPP nc16320 with polyP enzyme and also used a synthetic 5’P nc16320 to test the ability of exWAGO to bind 5’P molecules. The 5’P nc16320 molecules were radiolabelled with [32P]γ-ATP at the 5’ end and tested using EMSAs as described above. exWAGO binds both polyP-treated and the synthetic 5’P nc16320 with high affinity (polyP-treated: Kd = 18.63 nM ± 2.32 S.E.M.; synthetic: Kd = 8.43 nM ± 2.93 S.E.M.) (Fig. 3.13, Table 3.5).

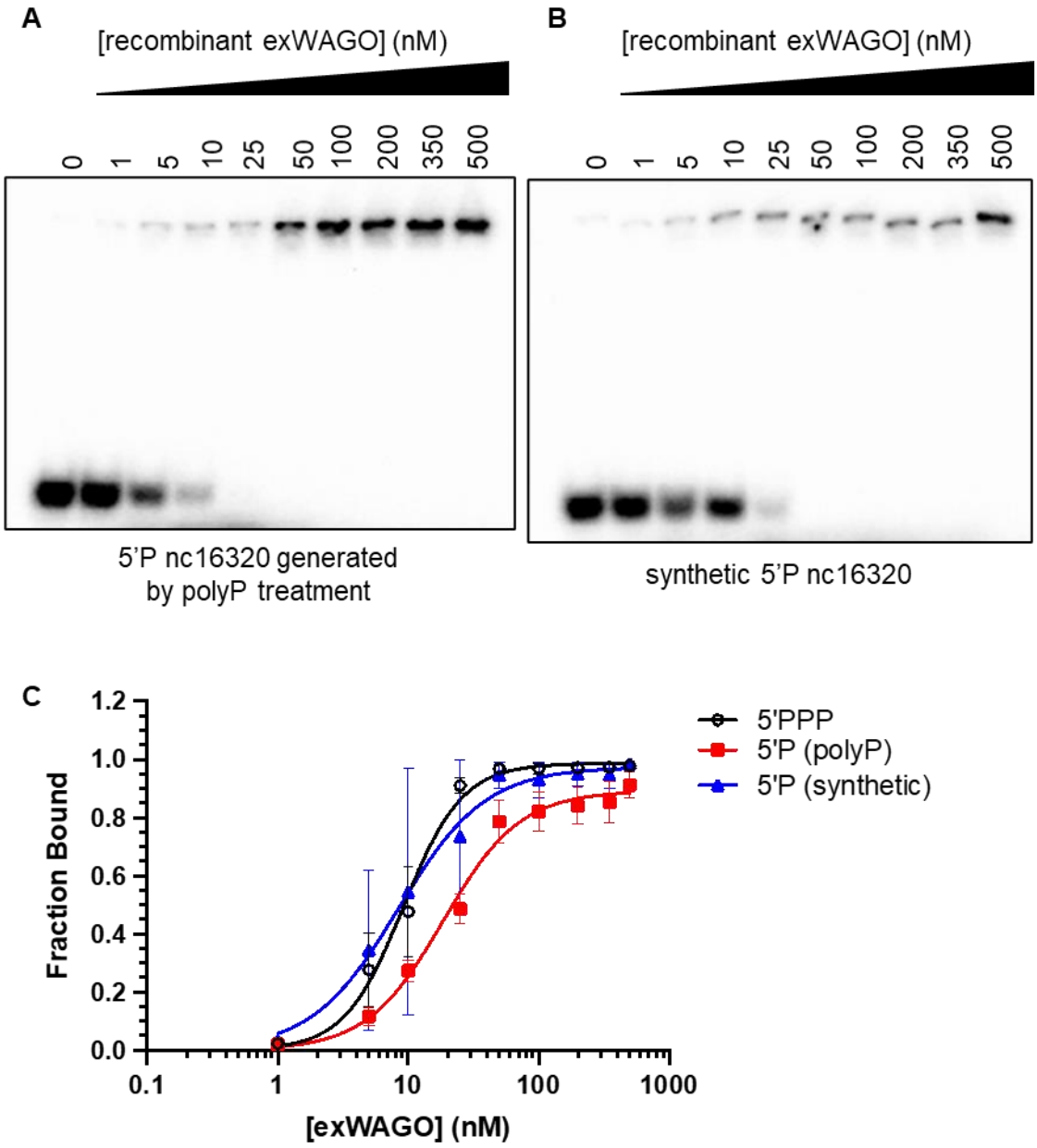


Figure 3.13 | exWAGO binds 5'P 23G sRNAs with high affinity.

A) 5'P nc16320 generated by polyP treatment or B) synthetic 5'P nc16320 were [32P]γATP-radiolabelled and 5 nM were incubated with increasing concentrations of recombinant exWAGO (0-500 nM). AGO:guide complexes were resolved on a native gel. C) Quantitative binding analysis of the fraction of guide bound by exWAGO. Data were quantified using ImageJ and show the mean±S.E.M from two to six independent experiments.

Consistent with the findings that exWAGO binds 5'P sRNAs with high affinity, I show that exWAGO can also bind miR100 (which has a 5'P) *in vitro* (Kd = 3.26 nM ± 0.42 S.E.M) (Fig. 3.14A&C, Table 3.5). As expected, mAGO2 binds strongly to 5'P miR100 as well (Kd = 11.30 nM ± 2.13 S.E.M) (Fig. 3.14B&C, Table 3.5). The *in vitro* data contrast with the data from the native extracellular exWAGO complexes, where exWAGO is bound to 5'PPP secondary siRNAs but not miRNAs.

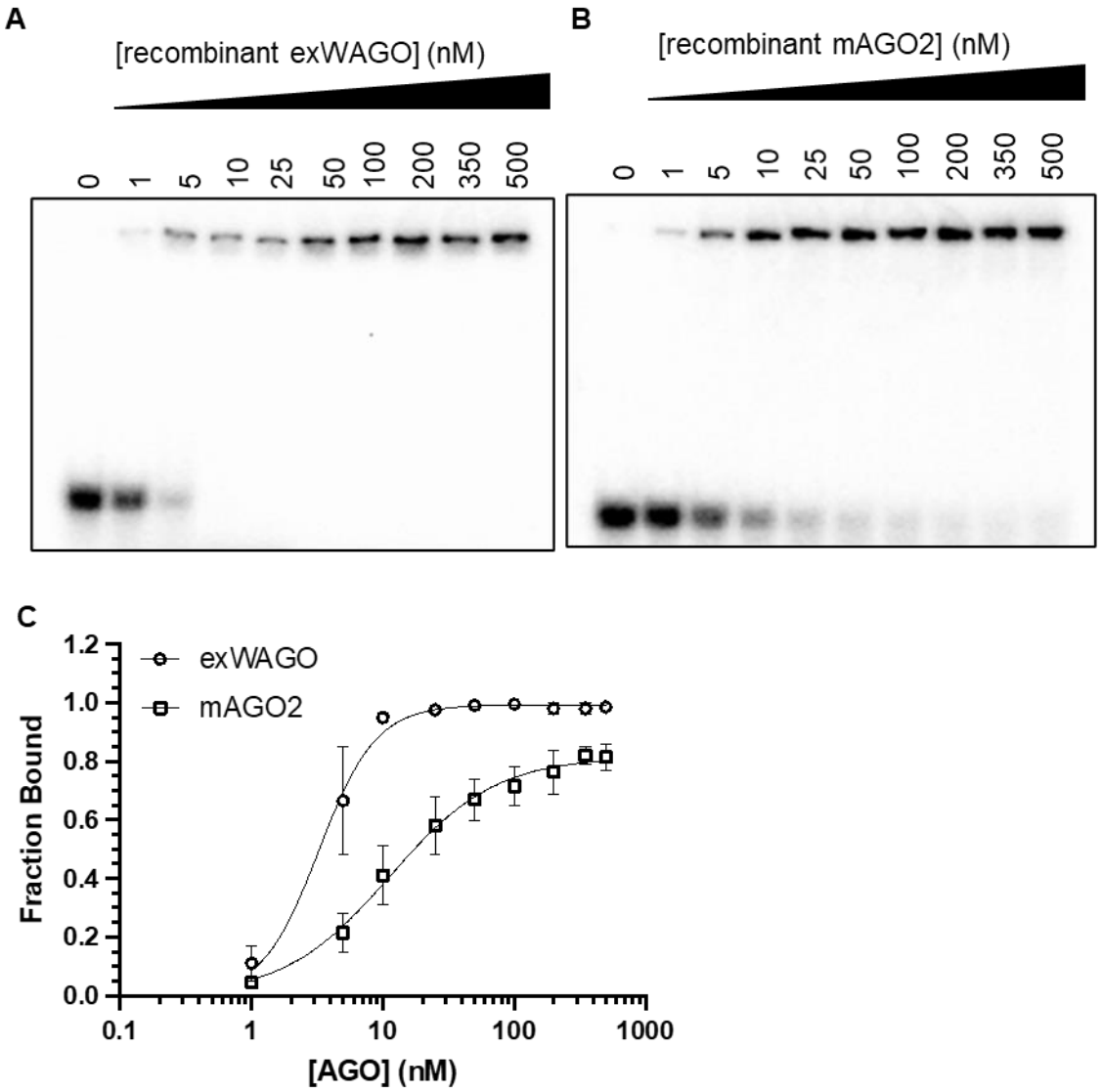


Figure 3.14 | exWAGO binds 5'P miR100 *in vitro*.

[32P]γATP-radiolabelled 5'P miR100 (5 nM) was incubated with increasing concentrations (0-500 nM) of recombinant A) exWAGO or B) mAGO2. AGO:guide complexes were resolved on a native gel. C) Quantitative binding analysis of the fraction of guide bound by exWAGO or mAGO2 plotted against increasing concentration of the AGO proteins. Data were quantified using ImageJ and show the mean±S.E.M from two independent experiments.

Discussion

3.3 Summary

In this chapter, I employed sRNA sequencing to detect which sRNA guides are loaded into the various forms of the exWAGO protein. Consistent with the fact that exWAGO is an orthologue of *C. elegans* Secondary AGOs, I show that exWAGO binds secondary siRNAs. These secondary siRNAs have a 5'PPP signature, are 22-23 nt in length and show a strong propensity for G as their starting nucleotide (Fig. 3.5A, 3.6A, 3.9A, 3.7A). The analysis of sRNAs found in libraries with different enzymatic treatments further supports the idea that the siRNAs bound by exWAGO have a 5'PPP. Consistent with this, I show that the extracellular forms of exWAGO preferentially bind sequences derived from DNA and RNA transposons, and simple and novel repeats, but are depleted from sequences that map to miRNAs, Y-RNAs, tRNAs and rRNAs (Fig. 3.7B, 3.8B). Intriguingly, binding assays indicate that exWAGO can bind both 5'P and 5'PPP molecules *in vitro* individually with high affinity (Fig. 3.12, 3.13). Comparison of the vesicular and non-vesicular exWAGO illustrates some overlap in the sequences bound but the relative abundance of different sequences varies in the two forms of exWAGO (Fig. 3.8). Additionally, there are some sequences present in only one or the other (Fig. 3.8, Supp. Table 3.1, 3.2). This suggests that the two forms of extracellular exWAGO could have different targets but also affect the same targets to a different extent.

3.4 exWAGO and its secondary siRNA guides

The work presented in this chapter clearly demonstrates that exWAGO binds 22-23G secondary siRNAs. Despite the presence of other types of sRNAs including miRNAs in the environment(s) where exWAGO exists, our data indicate that exWAGO is preferentially bound to secondary siRNAs. These results suggest that exWAGO might be implicated in selective export of some of these secondary siRNAs in the extracellular environment (inside and outside EVs). The idea of an RNA binding protein involved in RNA export is not novel. In fact, other RNA-binding proteins including the *Arabidopsis* AGO protein AGO1 have been described to be implicated in the loading of sRNAs into EVs secreted by *Arabidopsis* (He *et al*, 2021).

Moreover, exWAGO binds secondary siRNAs similarly to other AGO proteins of the WAGO subfamily in the free-living nematode *C. elegans*. Considering that *C. elegans* WAGOs are involved in gene silencing through various mechanisms, it is of great interest to understand what role exWAGO particularly plays inside the parasite, as

well as where it is expressed and where it is localised, both at the cellular and organismal level. The AGO gene family of *C. elegans*, which belongs to Clade V of nematodes like *H. bakeri*, has undergone a major expansion. *C. elegans* possess 27 AGO-like genes (Yigit *et al*, 2006), although only 19 of these encode functional proteins (Charlesworth *et al*, 2021). *H. bakeri* is thought to express 10 AGOs. Twelve of these 19 functional AGOs in *C. elegans*, belong to the WAGO subfamily and associate with 22G RNAs (Gu *et al*, 2009). SAGO-1, SAGO-2, PPW-1 (collectively termed intestinal secondary AGOs, iSAGOs (Claycomb lab, unpublished) belong to the WAGO subfamily. The functions of the iSAGOs are largely unknown but recent findings indicate that SAGO-2 and PPW-1 are loaded with secondary siRNA guides that are fully complementary to genes involved in immune responses (Seroussi, unpublished) and thus they seem to play a role in immune modulation. Another WAGO, NRDE-3 which is expressed in the soma, is involved in the transport of secondary siRNAs from the cytoplasm to the nucleus and nuclear silencing (Guang *et al*, 2008). Interestingly, WAGO-4 which also binds 22G RNAs, is expressed in the germline and is involved in RNAi inheritance (Xu *et al*, 2018). These examples illustrate a diverse repertoire of functions WAGO proteins exhibit using secondary siRNAs, despite that they are thought to share redundant roles in *C. elegans* (Gu *et al*, 2009). Work by Richard Davis lab shows that the Clade III parasitic nematode *Ascaris* expresses 10 AGO proteins, 5 of which are WAGOs (Zagoskin *et al*, 2021). Using similar methods to the ones I used in this chapter (i.e. generation of sRNA libraries with polyP treatment following AGO IPs from parasite intracellular material), the authors show that the *Ascaris* WAGOs bind 22-24G secondary siRNAs. Interestingly, *Ascaris* WAGO-3 and NRDE-3 show target plasticity as they can switch from targeting repetitive elements in the early embryo to targeting mRNAs during spermatogenesis (Zagoskin *et al*, 2021).

The idea that the worm secretes 5’PPP sRNAs in the mouse gut raises the question whether these would trigger an anti-viral-like response. The data here show that exWAGO is preferentially bound to the secondary siRNAs. Thus, we can predict that the 5’PPP moiety of those secondary siRNAs will not be exposed to any pattern-recognition receptors, as it will be ‘hidden’ in the binding pocket of the MID domain. However, our data also suggest that some 22G RNAs might remain in the unbound fraction, particularly inside the EVs. Nevertheless, further experiments are required to confirm that what we detect in the unbound is not derived from exWAGO remaining in the unbound fraction of the IP. In theory, unbound 22G RNAs could be

recognised by the Toll-like receptors (TLRs) TLR7 and TLR8. These are intracellular receptors which recognise single-stranded RNA (ssRNA) and act within endosomal compartments (Heil *et al*, 2004; Blasius & Beutler, 2010). Detection of 5'PPP ssRNAs by these TLRs can induce production of proinflammatory cytokines (like the Tumor necrosis factor and IL-12) and type I interferon cytokines (IFN-α/β) (Blasius & Beutler, 2010).

Gel shift assays testing the ability of exWAGO to bind sRNAs *in vitro* show that the recombinant exWAGO can bind the 5'P nc16320 and the 5'P miRNA miR100 with high affinity, comparable to the affinity it exhibits for the 5'PPP nc16320 (Fig. 3.13, 3.14A&C, Table 3.5). These results contrast the sequencing and qPCR data from the native complexes, which show that miR100 is not bound by exWAGO both inside and outside EVs (Fig. 3.4, Table 3.3). Thus, it is important to consider the native environment where exWAGO would be associating with its sRNA cargo and try recapitulating that *in vitro.* This would enable us to obtain results that are more likely to reflect what is happening *in vivo*, since we do not expect the recombinant exWAGO to be identical to the native exWAGO. For example, inside adult worms there is clear enrichment of 5'PPP sRNAs compared to any other 5'P molecules (Chow *et al*, 2019). Hence, since the frequency of secondary siRNAs is much higher in the environment where exWAGO exists, it is unreasonable to test the ability of exWAGO to bind 5'P sRNAs in the absence of 5'PPP sRNAs. Additionally, exWAGO might associate with certain protein factors that dictate the specificity for 5'PPP sequences observed *in vivo* or it might possess different association and dissociation rates for 5'P and 5'PPP, something which should be examined in the future.

How do proteins of the WAGO clade bind secondary siRNAs? To this day it remains unclear how the ability of WAGO proteins to associate with sRNA guides possessing a 5'PPP modification is facilitated (Faehnle & Joshua-Tor, 2007; Swarts *et al*, 2014). Studying the crystal structure of another AGO protein from *Archaeoglobus fulgidus* which belongs to the PIWI clade, AfPIWI, revealed a highly conserved region involved in the recognition of the 5'P terminus of the guide RNA near the MID and PIWI domains of the protein (Fig. 1.5, Fig. 3.16) (Parker *et al*, 2005; Parker & Barford, 2006). It has been noted that this highly conserved region is divergent in the WAGO clade of *C. elegans* by one amino acid where the equivalent of AfPIWI Tyr123 is replaced by a His in the WAGO proteins (Fig. 3.16) (Parker & Barford, 2006). The Tyr>His change is also observed in exWAGO (Fig. 3.16). This amino acid change

might be responsible for accommodating the 5'PPP feature of guides in the binding pocket of WAGO proteins (Faehnle & Joshua-Tor, 2007; Swarts *et al*, 2014). The ability of exWAGO and other WAGO proteins to bind 5'PPP sRNAs could be explained by their protein structure which have not yet been solved.

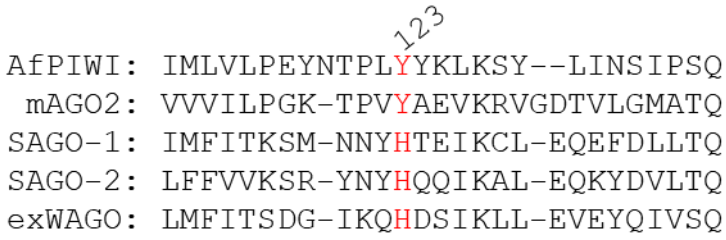


Figure 3.16 | exWAGO and other *C. elegans* WAGO proteins do not possess a Tyr residue part of a highly conserved motif involved in the binding of the 5' end of guide RNAs. Protein sequence alignment of the *Archaeoglobus fulgidus* PIWI protein (AfPIWI), the mouse AGO2 (mAGO2), the *C. elegans* WAGO proteins (SAGO-1 and SAGO-2) and of the exWAGO. The Tyr residue at the AfPIWI amino acid position 123 is indicated in red. The alignment was performed using Clustal Omega (1.2.4) (Sievers *et al*, 2011).

3.5 Research considerations & limitations

3.5.1 sRNA data analysis & normalisation

One of the main caveats of the sRNA sequencing data analysis presented in Section 3.1 is the normalisation used here. The data is by default normalised by distribution using the Trimmed Mean of the M-values (TMM) and then Counts Per Million (CPM) are computed using the estimated libraries by the TMM method. This method of normalisation assumes that differential expression across the samples compared is symmetrical and that technical effects do not differ between the differentially expressed and non-differentially expressed sRNAs (Evans *et al*, 2018). Our data violates these assumptions, especially when it comes to analysing bound vs unbound samples. Generally, this type of data violates all typical normalization methods thus it is arguable that there is no 'perfect' way of analysing. We do, however, need to be mindful of the caveats such analysis entails. Due to these normalisation issues, the number of sRNAs identified as differentially expressed could change drastically with different types of normalisation.

3.5.2 Vesicular & non-vesicular exWAGO guides: functional implications

Analysis using Degust identified tens of thousands of sRNA sequences differentially expressed when comparing the exWAGO IPs from EVs and EV-depleted HES (Fig 3.8A). However, it is now important to consider which of these sequences, that are bound by the vesicular and non-vesicular exWAGO proteins, can induce an effect in the mouse host. Additionally, we need to consider what the rules for a parasite secondary siRNA to base pair with a host mammalian target and successfully induce target silencing are. Evidence suggests that several factors can be implicated in efficient target repression that concern both the guide sRNA and the target gene. These include the target's (1) dosage sensitivity (i.e. the extent to which the guide needs to supress the target so that the protein levels of the target fall below the functional threshold of the protein), (2) position in relation to its gene regulatory network, and (3) spatiotemporal properties (Ketting & Cochella, 2020). The abundance of the guide is also important. Different guide sRNAs might co-target the same transcript to achieve an additive repressive effect that would otherwise not be reached by the individual contribution of each guide (Ebert & Sharp, 2012). Furthermore, association with specific protein factors such as GW182, which forms part of the High Molecular Weight RNA-induced Silencing Complex (HMW-RISC), has been reported to be important in gene regulation (Lian *et al*, 2009; Rocca *et al*, 2015), although not always essential (Liu *et al*, 2019b). Based on the aforementioned observations, one might expect that the extent of silencing and any other effects caused by the sRNA guides that the vesicular or non-vesicular exWAGO is loaded with, will depend on: (1) how the exWAGO:sRNA complex associates to its RNA target (eg. co-targeting or individual guide targeting), and (2) the availability of complexes to interact with any host factors required to induce gene suppression. In turn, these reasonings depend on how many of the exWAGO:sRNA complexes are internalised by the target cell(s), which in turn relies on how much material is being secreted by the parasites, which depends on the level of infection (i.e. the number of adult worms present in the host gut environment).

Lastly, these findings provoke us to ask which mouse cells are targeted by each extracellular form of exWAGO? The intestinal epithelium is arguably composed of at least seven cell types (Gerbe *et al*, 2012), while the immune cells residing in the lamina propria beneath the intestinal epithelium such as macrophages and dendritic cells can also be involved in the uptake of helminth ES products (Maizels *et al*, 2018). In theory, any of these mouse cells could be targeted by either the vesicle-free

exWAGO or by the EVs carrying the vesicular exWAGO, while the gut microbiome and/or other co-infecting parasites could also be the recipient organisms of exWAGO. Identifying which host cells (or other organisms) are the targets of the two extracellular forms of exWAGO will be important in understanding the impact of the sRNA cargo they carry.

3.5.3 Gel shift assays

The EMSA experiments were performed with a recombinant version of the exWAGO and mAGO2 proteins. The purification steps these proteins underwent do not select for 'empty' AGO protein. This needs to be taken into consideration when interpreting the Kd values reported here (Table 3.5), as the radiolabelled guide would need to compete with any pre-loaded guides for binding by the recombinant AGO. Additionally, it should be noted that [32P]pCp-labelling results in the addition of an extra C (cytidine) at the 3' end of the sRNA labelled thus the 23 nt 5'PPP-nc16320 becomes 24 nt in length.

3.6 Future Work

sRNA datasets generated in this project can complement existing datasets to further investigate the biology of exWAGO and adult *H. bakeri* worms. For example, the sequences identified in the exWAGO IPs from adult worms after polyP- and non-polyP treatment can be compared to the sequences identified in the sRNA libraries generated from total adult worm lysate with and without polyP treatment by Chow *et al* (2019). This comparison would identify the sRNA population enriched in the exWAGO IPs compared to the total sRNA population that is present inside the parasite. This can also provide information as to the function of exWAGO inside worms. Comparing the sRNAs identified in the total HES libraries (polyP- and non-polyP treated) to those detected in total adult worm lysate by Chow *et al* (2019), can identify which sRNAs are not secreted by *H. bakeri*, regardless of whether they are bound by exWAGO. Sequences that are not exported by the parasite can be considered as sRNAs that are not involved in intra- or inter-species communication. Moreover, one could also analyse the sRNA cargo that is bound by exWAGO inside the EVs and the *H. bakeri* sequences identified to be internalised by MODE-K cells (mouse intestinal epithelial cells) after incubation with EVs (Bermúdez-Barrientos *et al*, 2019). This analysis would provide suggestive evidence that exWAGO can successfully transport sRNAs to mouse host cells *in vitro* via vesicles. The same experimental set up can be executed for the non-vesicular exWAGO, however a

dataset of which non-vesicular sRNAs are internalised by MODE-K cells following incubation with EV-depleted cells would also be required. This experiment would provide evidence that non-vesicular sRNAs are successfully delivered to host cells *in vitro* and this might be facilitated by direct internalisation of exWAGO.

Data analysis of samples treated with the terminator exonuclease enzyme (Section 3.1.6) indicates that this treatment successfully reduces the signal from 5'P sequences, resulting in the enrichment of signal from 5'PPP reads. This enzymatic treatment can be applied to find parasite sequences inside the host, as parasite reads are expected to be of much lower abundance to those of the host. To exemplify, using this method we can ask if 5'PPP parasite sRNAs can be detected in host organs that are distal to the site of infection or in the blood circulation of the host, to understand whether the parasite cargo can have an effect beyond the site where the worms reside.

To understand the function of the secondary siRNAs loaded in the exWAGO protein secreted by *H. bakeri* identified here, we then aimed to identify the genes targeted by these guides in the next chapter.

Chapter 4: Identification of exWAGO host targets

Introduction

The main goal of this project is understanding the function of exWAGO, and we hypothesise that exWAGO could be involved in manipulating host gene expression through the sRNA guides it associates with. Identifying which host genes are being targeted by exWAGO:guide complexes is crucial in elucidating the function of exWAGO.

A popular method to detect sRNA targets is computational target prediction, for which there are at least 60 tools available (Lukasik *et al*, 2016). The computational tools available generally focus on predicting potential miRNA targets specifically centred on: (1) seed-based guide:target complementarity (nucleotides at position 2-8 of the miRNA from the 5' end), (2) sequence conservation across species of the miRNA and/or the 3'UTR or 5'UTR of the transcript, (3) Gibbs free energy, and (4) site accessibility (Peterson *et al*, 2014). Such computational tools have been used to predict worm guide:mouse target interactions. For example, Buck *et al*, (2014) reports that over a hundred genes are differentially expressed when mouse intestinal epithelial cells are incubated with *H. bakeri* EVs. One such target is the gene expressing Dusp1, a protein implicated in the control of anti-inflammatory genes. Following miRNA target prediction, Buck *et al*, (2014) identified miRNA target sites on the 3'UTR of *Dusp1* and using a reporter assay the authors show that *H. bakeri* miRNAs miR-200, miR-425 and let-7 found in EVs reduce the expression of luciferase fused to the 3'UTR of *Dusp1*. These findings indicate that *H. bakeri* miRNAs can regulate host gene expression *in vitro* and the sRNAs secreted by *H. bakeri* can also be implicated in modulating host immune responses (Buck *et al*, 2014). Furthermore, several genes expressed by human innate immune cells are predicted to be targeted by miRNAs of the parasitic worm *Fasciola hepatica* (Ricafrente *et al*, 2021). More specifically, Tran *et al* (2021) identified that the miR-125b released in EVs by *F. hepatica* has hundreds of immune-related genes as putative targets. As the *F. hepatica* miR-125b is homologous to the human miR-125b which is known for regulating the activation of the pro-inflammatory M1 macrophages (Chaudhuri *et al*, 2011), the authors restricted their target prediction analysis to identify targets in peritoneal macrophages (Tran *et al*, 2021). This led to identification of Traf6 as a putative target of miR-125b (Tran *et al*, 2021). Traf6 is required for activation of NF-κB pathways and it interacts with other proteins important for the regulation of

inflammatory responses (Lalani *et al*, 2017). Tran *et al* (2021) show that *in vitro* transfection of *F. hepatica* miR-125b significantly repressed Traf6 expression in macrophages. These results show that *F. hepatica* miRNAs can manipulate host gene expression *in vitro* and reveals that parasite sRNAs are implicated in immunomodulation of the host.

The computational target prediction approach has several limitations, including: (1) the physiological relevance of any predictions which will need to be validated, (2) identification of false positive sRNA:target associations, and (3) the unknown rules for target prediction of sRNAs other than miRNAs. Nevertheless, the aforementioned studies provide a proof of concept that parasite sRNAs can regulate host gene expression *in vitro* and highlight how computational tools can successfully predict target genes. However, validation experiments are required to confirm that the guide and the target RNAs truly associate under physiological conditions (rather than just *in vitro*) and to understand the outcome of such interactions with regards to the expression of the target.

As detailed in Chapter 3 (Section 3.1.6), the majority of sRNAs secreted by *H. bakeri* are secondary siRNAs rather than miRNAs. Therefore, it might be expected that secondary siRNAs potentially are more important effector molecules in cross-species RNAi since they are present at much greater frequency than miRNAs. Additionally, we cannot confidently predict how the parasite secondary siRNAs would be targeting host genes using computational tools. This is because secondary siRNAs associate with targets using full or near-perfect complementary base pairing, but miRNAs mainly induce gene silencing using a seed match (Hoogstrate *et al*, 2014). So how would worm secondary siRNAs associate with mouse targets? Other members of the Buck lab are testing various computational prediction methods to predict targets. The goal in this chapter is to obtain biochemical data on what host genes might associate with exWAGO.

To achieve this goal, I tested a biochemical approach to directly detect the exWAGO guide sRNAs *in vivo* and their specific mRNA targets. This involves UV crosslinking of the RNA to the AGO protein to generate covalent bonds between the protein and associated RNAs, followed by immunoprecipitation of the Argonaute:RNA complex under stringent conditions and intermolecular ligation of the guide sRNA to the target mRNA to generate chimeric reads (also called RNA hybrids) which can then be sequenced (Helwak & Tollervey, 2014; Moore *et al*, 2015). Several variations of the

methods have been developed and adapted by different research groups to do this. These include the CLASH (Cross-linking, Ligation and Sequencing of Hybrids) method which was developed using cells from culture as starting material (Helwak *et al*, 2013; Helwak & Tollervey, 2014), and the CLEAR-CLIP (Covalent Ligation of Endogenous Argonaute-bound RNAs Cross-linking and Immunoprecipitation) technique applied on cultured cells and brain tissue (Moore *et al*, 2015).

To identify exWAGO-specific guide:host target chimeras I first developed a protocol to immunopurify exWAGO *in vivo* from *H. bakeri*-infected mouse gut tissue and purify its associated guides. Then I used a modified version of the CLASH protocol to generate parasite guide:host target reads, however I omitted cross-linking, radioactive labelling of RNA and gel purification of AGO:RNA complexes. This approach was based on data generated by others in the lab showing that highly stable miRNA-target interactions in human AGO2 could be specifically captured without crosslinking (Sarah Ressel, unpublished data). As a positive control, in parallel with the immunoprecipitation of exWAGO I also immunoprecipitated host mAGO2 for comparison.

Results

4.1 Development of *in vivo* gut IPs

To examine host targets of exWAGO under physiologically relevant conditions, I developed and optimised the purification of exWAGO from *H. bakeri*-infected mouse gut tissue using rat polyclonal antibodies (Fig. 4.1). *H. bakeri*-infected gut tissue was obtained from CBA x C57BL/6 F1 mice after 14 days of infection. Briefly, the duodenum was flash frozen in liquid nitrogen immediately after the mice were culled and stored at -80^{o}C until required. The tissue was ground to a fine powder under liquid nitrogen and lysed in the presence of protease and RNase inhibitors. The lysate was spun, and the non-lysed (pelleted) material was discarded. The supernatant was incubated with the beads conjugated to rat anti-exWAGO polyclonal antibodies for two different periods of time (45 min or 4 h). Western blot analysis shows that exWAGO can successfully be immunoprecipitated from infected gut tissue after incubating the lysate with the antibody-conjugated beads in as little as 45 minutes (Fig. 4.2A). For comparison, I also tested immunopurification of the host mAGO2 under the same conditions. mAGO2 was also successfully pulled down from gut tissue in 45 mins using the same protocol but with the mouse anti-mAGO2 monoclonal antibody (Fig. 4.2B). The anti-exWAGO and anti-mAGO2 antibodies used do not cross-react, as seen in Figure 4.4. To minimise formation of non-specific interactions and minimise the chance of RNA getting degraded, and since the AGO proteins were pulled down successfully in a short period of time, we decided to use the 45 min incubation time rather than longer incubation periods for all the *in vivo* gut IPs performed in this project, unless otherwise stated.

Furthermore, I tested how to best extract the RNA associated with exWAGO pulled down from gut tissue. This would allow us to minimise loss of RNA during the extraction process and increase RNA recovery. The exWAGO:RNA complexes were eluted in Qiazol, and the samples were spiked with a known concentration of RT4 synthetic RNA (a 22 nt sequence that is not found in worms or mouse) prior to RNA extraction, to assess if the method for RNA extraction was optimal. The RNA was extracted using either the Direct-zol RNA MiniPrep kit (Zymo Research, catalogue no. R2050) or the miRNeasy serum/plasma kit (Qiagen, catalogue no. 217184). After RNA extraction, the samples were also spiked with synthetic cel-miR-39-3p spike at the stage of cDNA synthesis to control for inhibitors in the PCR reactions. Then I performed RT-qPCR for RT4 and cel-miR-39-3p spikes. The data clearly show that

the sRNA recovery was far greater when using the miRNeasy serum/plasma kit (Fig. 4.2C), while cDNA synthesis was comparable between the different samples (Fig. 4.2D). As more RNA is recovered using the Qiagen miRNeasy serum/plasma kit, this kit was used to extract the RNA associated with exWAGO or mAGO2 immunoprecipitated from gut tissue.

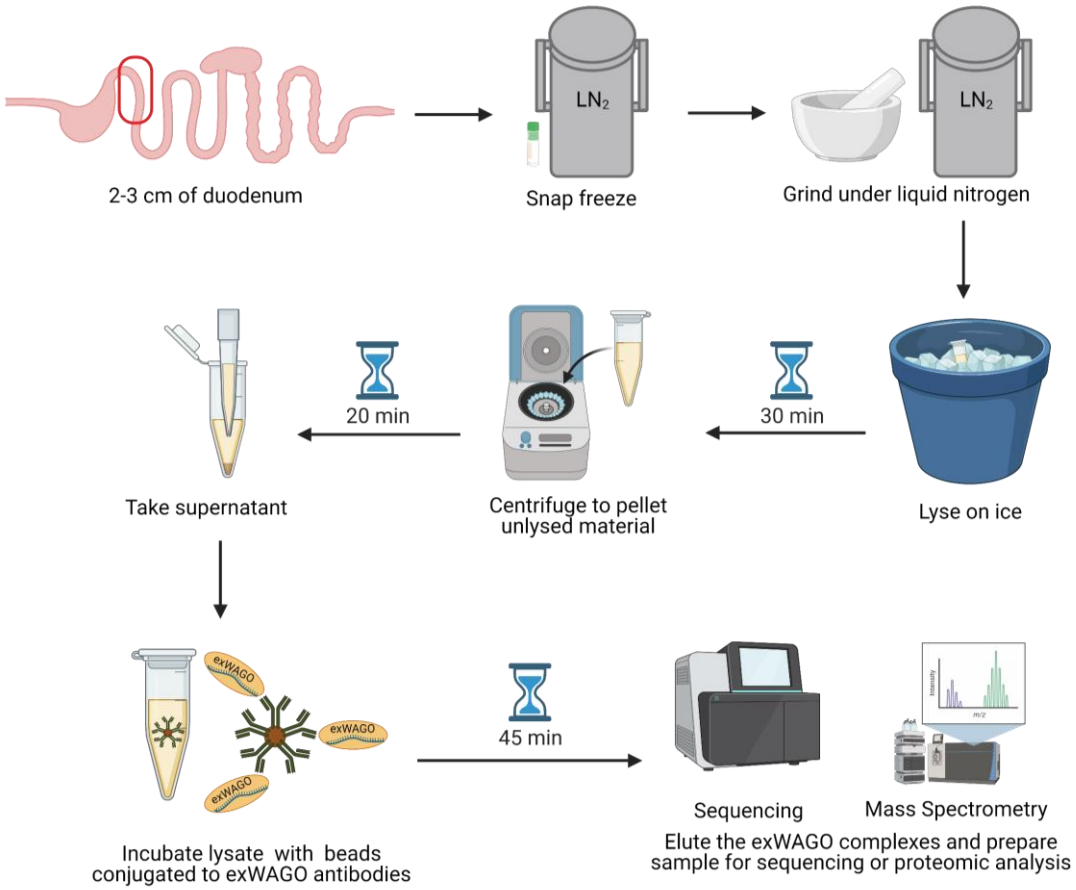


Figure 4.1 | Schematic representation of exWAGO immunoprecipitation from mouse gut tissue. The same protocol was also applied to mAGO2 with the appropriate antibody. Figure created with BioRender.com

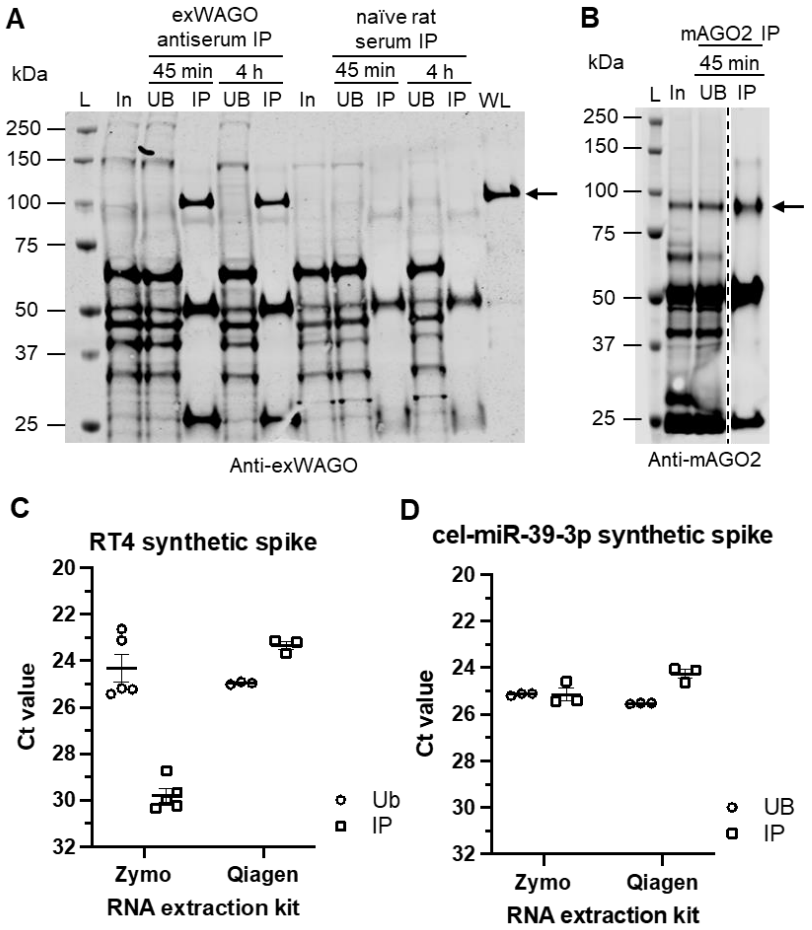


Figure 4.2 | Optimisation of *in vivo* gut IPs. A) Western blot analysis of exWAGO IPs from *H. bakeri*-infected gut tissue with rat polyclonal antiserum or naïve rat serum for negative control with 45 min or 4 h incubation of the lysate with beads conjugated with appropriate antibody. The western blot was probed with 1:2,000 rabbit anti-exWAGO polyclonal antibody, followed by 1:10,000 goat anti-rabbit IgG Dylight 800. The arrow indicates the expected size of exWAGO (102 kDa). B) mAGO2 IP from *H. bakeri*-infected gut tissue with anti-mAGO2 mouse monoclonal antibody, with 45 min incubation of the lysate with anti-mAGO2 antibody-conjugated beads. The western blot was probed with 1:4,000 mouse anti-mAGO2 monoclonal antibody, followed by 1:10,000 goat anti-mouse AlexaFluor 680. The arrow indicates the expected size of mAGO2 (97 kDa). Percentage/Amount of fraction analysed: 1% Input (In), 1% Unbound (UB), 6.67%. for panel A and 3.75% for panel B of eluate (IP), 5.5 µg of worm lysate (WL). L = ladder. C) RT-qPCR data of RT4 synthetic spike used prior to RNA extraction. For the samples processed using the Zymo kit, data show the mean±S.E.M from the unbound or eluate fractions of the following 5 samples (5 data points): 2 biological replicates of anti-exWAGO IPs, 2 biological replicates of naive rat serum IPs and 1 replicate of anti-mAGO2 IP. Each data point was calculated as the mean Ct value from 2 qPCR technical replicates. For the samples processed using the Qiagen kit, data show the mean±S.E.M from unbound or eluate fractions of the following 3 samples (3 data points): 1 replicate of anti-exWAGO IP, 1 replicate of naive rat serum IP and 1 replicate of anti-mAGO2 IP. Each data point was calculated as the mean Ct value from 2 qPCR technical replicates. D) RT-qPCR data of cel-miR39-3p synthetic spike used after RNA extraction but prior to cDNA synthesis. For the samples processed using the Zymo or the Qiagen kit, data show the mean±S.E.M from unbound or eluate fractions of the following 3 samples (3 data points): 1 replicate of anti-exWAGO IP, 1 replicate of naive rat serum IP and 1 replicate of anti-mAGO2 IP. Each data point was calculated as the mean Ct value from 2 qPCR technical replicates. For the IPs shown in in this figure, 100 µl of Protein G beads with 20-25 µl of appropriate antibody were used per sample. The IPs in C) and D) were performed with 45 min lysate incubation.

4.2 Generating RNA libraries from *in vivo* gut samples

Following optimisation of the gut IPs, I then applied a modified version of the CLASH protocol to identify host targets of exWAGO (Chapter 2, Section 2.13) (Fig. 4.3). In this modified version of the CLASH protocol, I omitted the cross-linking step, radiolabelling of RNA and gel purification of AGO:RNA complexes. I immunopurified exWAGO from gut tissue of CBA x C57BL/6 F1 mice infected with *H. bakeri* for 14 days. I used the rat anti-exWAGO polyclonal antiserum and three biological replicates (each representing a different mouse). The material captured by these IPs was divided in two. One part of the sample was treated with RNase as per protocol, but the other part of the sample was not subjected to RNase treatment. As gut tissue contains a lot of digestive enzymes including RNases, we considered that the RNA that was not protected by the protein may already be quite degraded and wanted to test the extent to which RNase treatment influenced the results. I designed the experiment to control for several factors. First, to assess the host background, I performed exWAGO IPs from uninfected mouse gut tissue (three biological replicates). Additionally, to assess the parasite background and the host background in the context of infection, where gene expression will be different compared to uninfected tissue, I performed an IP using rat naïve antiserum (one replicate) from infected gut tissue. To compare the sequences identified in the nematode exWAGO to the well-characterised mammalian AGO2, I pulled down mAGO2 from the corresponding *H. bakeri*-infected and uninfected tissues (using the same three biological replicates). The samples generated in this experiment are summarised in Table 4.1.

Following the various enzymatic treatments to achieve ligation of the guide and target on the beads (Fig. 4.3), I retained 90% of the eluted sample for RNA extraction in Qiazol and the remaining 10% was eluted in SDS-PAGE protein buffer for western blots. Western blots were carried out prior to generating RNA libraries to check the degree to which each AGO was captured. As expected, exWAGO was pulled down after 45 min anti-exWAGO IP from all the worm-infected *in vivo* samples (Fig. 4.4A, lanes 2,3,6,7,10 and 11) but not from the uninfected gut samples (Fig. 4.4B, lanes 2, 5, and 8). mAGO2 was immunoprecipitated from all the samples as expected (Fig. 4.4A, lanes 4, 8 and 12; and Fig. 4.4A, lanes 3, 6 and 9). The western blot also shows that there is no cross-reaction between the exWAGO and mAGO2 antibodies, as the same blot probed with anti-exWAGO primary antibody was re-probed with anti-mAGO2 primary antibody (Fig. 4.4).

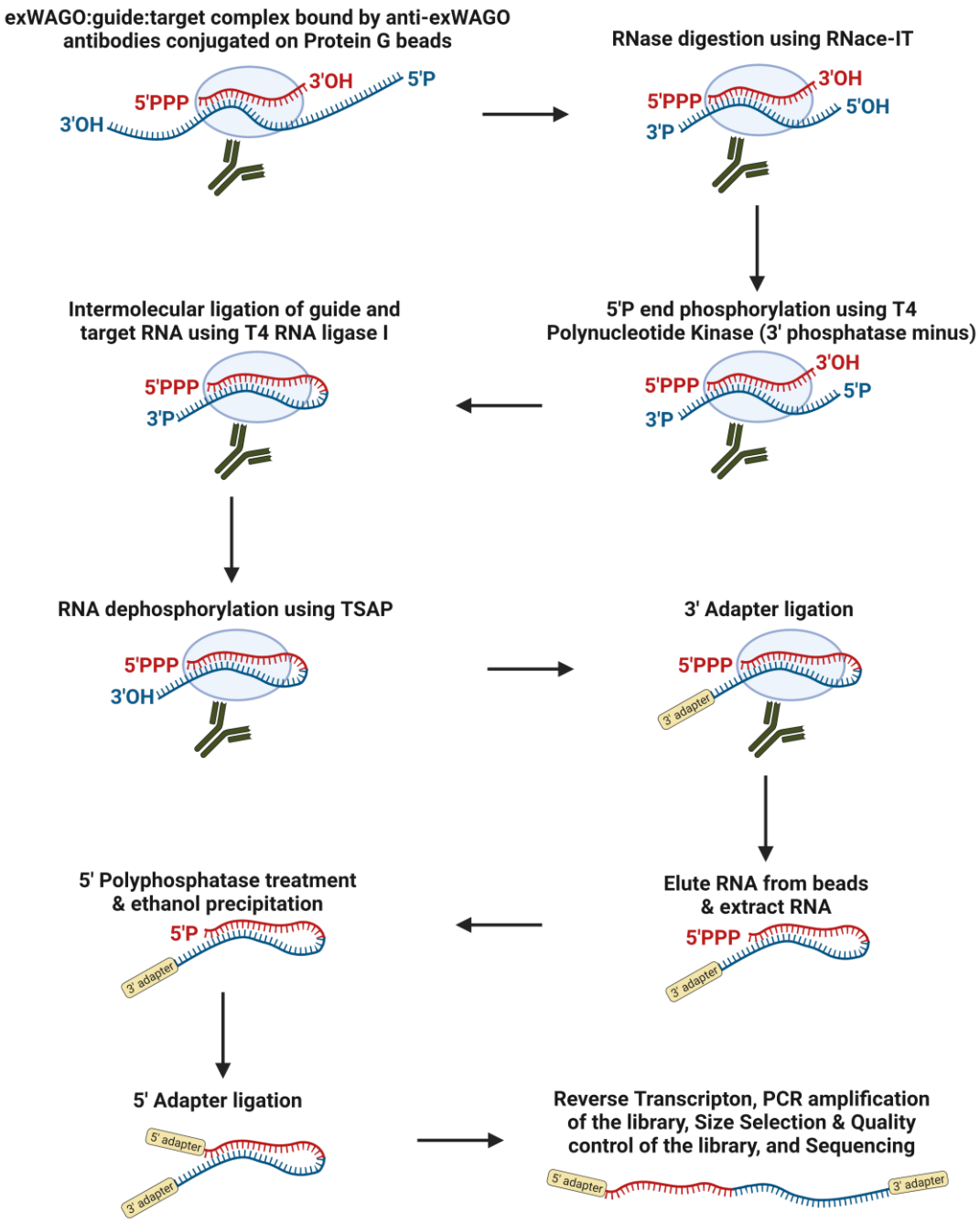


Figure 4.3 | Summary of the modified CLASH protocol employed for the direct identification of exWAGO host targets by ligation of the guide to the mRNA transcript. The same protocol was also applied to mAGO2 with the appropriate antibody. We do not expect mAGO2 to bind 5'PPP guides based on data obtained in Chapter 3 (Section 3.2.1). Figure created with BioRender.com

Table 4.1 | Summary of the experimental samples and controls used for CLASH. Each infected gut tissue from three mice was divided in three for exWAGO IP with and without RNase treatment and mAGO2 IP. Similarly, each uninfected gut tissue from three mice was divided in two for exWAGO and mAGO2 IPs. Another infected mouse gut tissue was exclusively used for IP with naïve rat serum. Per IP, 100 µl of Protein G beads conjugated to 20 µl of antibody was used and the sample was incubated with the antibody-conjugated beads for 45 min.

Material	IP antibody	RNase treatment
Infected mouse gut (n = 3)	exWAGO	No
	exWAGO	Yes
	mAGO2	Yes
Uninfected mouse gut (n = 3)	exWAGO	Yes
	mAGO2	Yes
Infected mouse gut (n = 1)	Naive rat serum	Yes

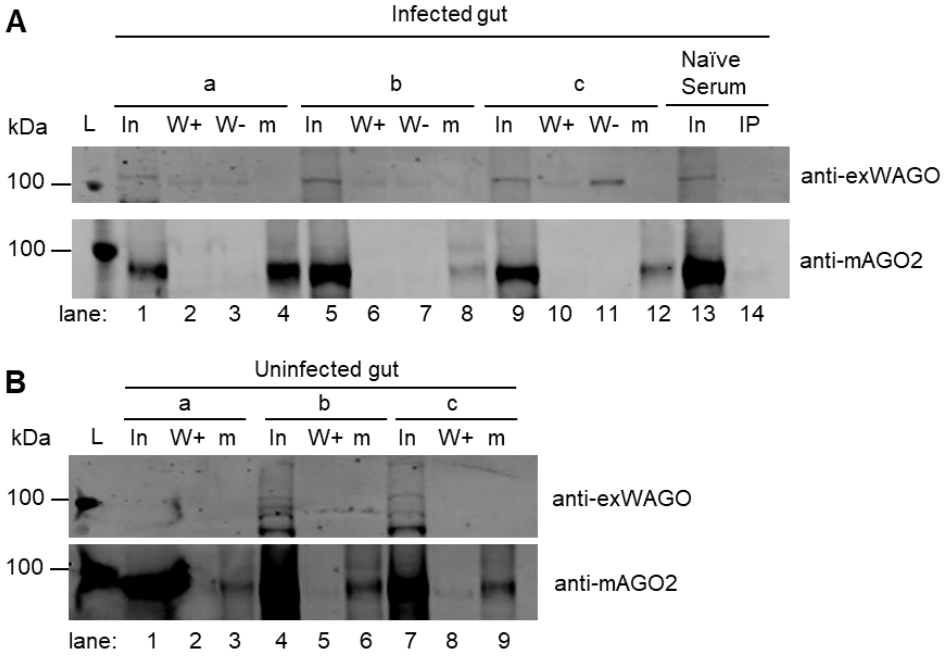
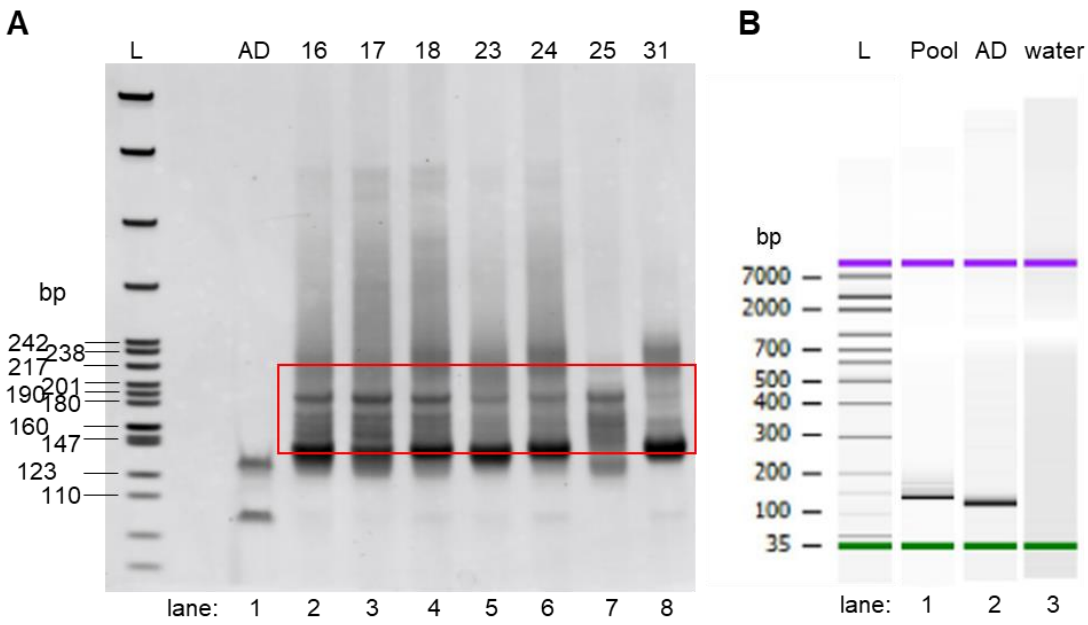


Figure 4. 4 | Quality control of the *in vivo* gut IPs following the CLASH modified protocol. A & B) Western blot analysis of exWAGO and mAGO2 IPs as detailed in Table 4.1 from (A) *H. bakeri*-infected gut tissue or (B) uninfected gut tissue. Percentage of fraction analysed: 2% input (In), 2.5% exWAGO IP using antiserum (W (+/-)) with (+) or without (-) RNase treatment, 2.5% mAGO2 IP (m). L = Ladder, a, b, c = biological replicates. For these IPs, 100 µl of Protein G beads with 20 µl of appropriate antibody were used per sample. Western blots were first probed with 1:2,000 rabbit anti-exWAGO primary antibody followed by 1:10,000 goat anti-rabbit IgG Dylight 800 and then re-probed with 1:4,000 mouse anti-mAGO2 primary antibody followed by 1:10,000 goat anti-mouse AlexaFluor 680.

The eluted RNA was extracted using the miRNeasy Serum/Plasma Kit (Qiagen) and was used to make RNA libraries. Unlike the generation of sRNA libraries performed in Chapter 3, the 3' adapter was already ligated to the RNA on the beads for the modified CLASH protocol here (Fig. 4.3). However, the 5' adapter is added after the RNAs have been eluted from beads. In Chapter 3, we showed that exWAGO is bound to triphosphorylated guides, thus I treated the RNA with polyP to convert any 5'PPP molecules to 5'P to permit ligation of the 5' adapter required for sequencing. The RNA was then reverse transcribed, PCR amplified, and the products examined using a TBE gel. The exWAGO and mAGO2 IPs are enriched with sRNAs, which run at approximately 140-144 bp (Fig. 4.5A, lanes 2-6 and 8). This band is not enriched to the same extent in the negative control (Fig. 4.5A, lane 7). Adapter dimers are seen just below the sRNAs at approximately 120-128 bp (Fig. 4.5A). Following this quality control analysis, the samples were pooled together at equal volumes and re-analysed on a TBE gel. A gel band was excised to include RNAs of approximately 142-217 bp and to exclude adapter dimers. The libraries were gel purified, examined using the Bioanalyzer High Sensitivity DNA chip (Fig. 4.5B, lane 2; Supp. Fig. 4.1), and then sequenced using the Illumina NovaSeq platform by Edinburgh Genomics (100 bp single end).



Sample number	Gut tissue	IP	RNace-IT treatment
16	Infected - a	exWAGO	Yes
17	Infected - a	exWAGO	No
18	Infected - a	mAGO2	Yes
23	Infected - c	exWAGO	No
24	Infected - c	mAGO2	Yes
25	Infected - d	naïve serum	Yes
31	Uninfected - c	mAGO2	Yes

Figure 4.5 | Quality control analyses of the libraries generated by the modified CLASH protocol from gut tissue.

A) A 6% TBE gel was run to assess the quality of the libraries for a subset of the samples as designated in the table. The red box indicates the size of bands I proceeded to cut after all the samples were pooled and run on another TBE gel. L = Ladder; AD = Adapter Dimer control; bp = base pairs. B) High Sensitivity Bioanalyzer chip electropherogram after size purification of the final pool. L = Ladder; AD = Adapter Dimer control; bp = base pairs.

Subsequently, sequenced reads were processed by Dr. Sujai Kumar to remove the adapter sequences and retain only reads ≥ 18 nt in length. These reads were then mapped to *H. bakeri* and mouse genomes, allowing up to 1 nt mismatch. As expected, only the exWAGO IPs from infected tissue are enriched with reads mapping to the nematode while the mAGO2 IPs (from infected and uninfected tissue) are enriched in reads mapping to the mouse genome (Fig. 4.6). The exWAGO IPs from uninfected tissue have a much smaller number of reads (< 420,000 reads) compared to IPs from infected samples (> 16,000,000 reads) which could represent background sequences sticking to beads and the anti-exWAGO antibody. The naïve serum IP has a comparable count of reads to that of the exWAGO IPs from infected tissue (Fig. 4.6A). Nevertheless, the naïve serum IP from infected tissue pulls down mostly reads that map to the mouse genome compared to the anti-exWAGO IP from infected tissue which shows nematode reads, demonstrating specificity in the assay (Fig. 4.6B).

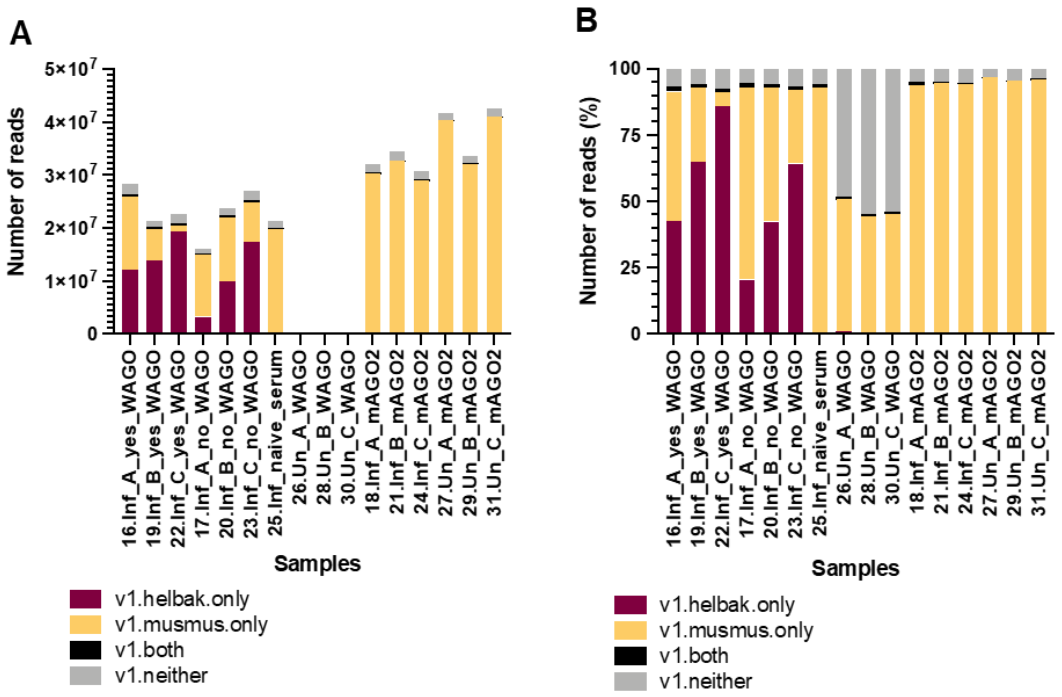


Figure 4.6 | The reads detected in the CLASH libraries from gut tissue IPs as (A) counts or as (B) percentage, prior to ShortStack analysis. v1 = the reads were mapped with up to 1 nt mismatch, helbak.only= the reads map to the nematode genome only, musmus.only = the reads map to the mouse genome only, both = the reads map to both the nematode and the mouse genome, neither = the reads do not map to either the nematode or mouse genomes, Inf = infected gut tissue, Un = Uninfected gut tissue, A/B/C = replicate number, WAGO = exWAGO, yes/no = refers to whether RNase treatment was performed. The number in the sample name indicates the index used during RNA library preparation. Data presented here were quantified by Dr. Sujai Kumar.

The reads mapping to the mouse genome only (with up to 1 nt mismatch) were further categorised into RNA biotypes by Dr. Sujai Kumar, as shown in Figure 4.7A. This was a quick quality control analysis to check that the mAGO2 pulled down mainly miRNAs as expected. Indeed, the sequences pulled down by mAGO2 are enriched with miRNAs, regardless of whether the tissue was infected or not. However, the mAGO2 IPs from infected tissue also pull down rRNA. Examination of the sequences mapping to mouse after IP with exWAGO from both infected and non-infected tissue, as well as the naïve serum IP pull down, indicates enrichment in rRNA (Figure 4.7). rRNA is considered to be the most common background in these types of libraries (Fowler *et al*, 2018). Interestingly, 0.38-2.48% of the reads (mapping to the mouse genome only) pulled down from the exWAGO IPs from infected tissue map to mouse miRNAs (Figure 4.7). This might imply that exWAGO binds mouse miRNAs and utilises them as guides or it binds them to inhibit their function. Nevertheless, as reads mapping to mouse miRNAs are also identified in the control samples (exWAGO IP from uninfected tissue) (Figure 4.7), close examination of these reads is required to specify if there are enriched mouse miRNAs bound by exWAGO only. These reads could also be a result of barcode 'hopping' between libraries that have some highly abundant reads which are not found in other libraries (Griffiths *et al*, 2018).

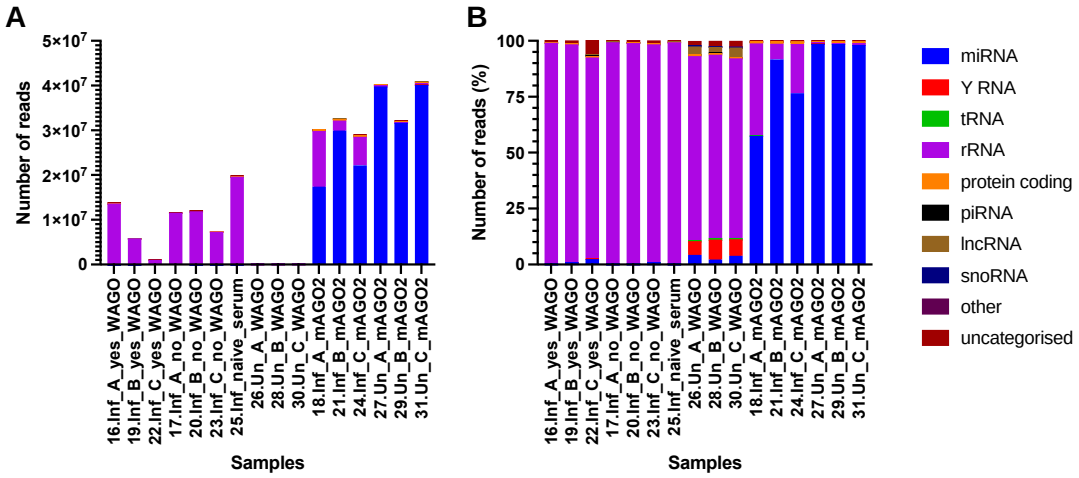


Figure 4.7 | RNA biotypes identified from the reads detected in the CLASH libraries from gut tissue IPs that map only to the mouse genome with up to 1 mismatch. The data are shown as counts (A) or as percentage (B). Inf = infected gut tissue, Un = Uninfected gut tissue, A/B/C = replicate number, WAGO = exWAGO, yes/no = refers to whether RNase treatment was performed. The number in the sample name indicates the index used during RNA library preparation. Data presented here were quantified by Dr. Sujai Kumar.

4.3 Identification and characterisation of exWAGO guides *in vivo*

The modified CLASH protocol applied on the *in vivo* gut samples detailed in Figure 4.3 and Table 4.1 can potentially give us two main pieces of information: (1) the guide sRNAs associated with exWAGO or mAGO2 and (2) the host genes targeted by the guides. Identification of host targets will be discussed in Section 4.4.

Since only a small proportion of the reads will be guides ligated to the target, here we investigate guide reads that are not in a chimeric read. I focused on the exWAGO-bound sRNAs mainly rather than the mAGO2-bound sRNAs as exWAGO is considered to be less abundant compared to the host mAGO2 in the starting mouse gut material used for the IPs. Hence, exWAGO-bound guides would be considered more difficult to detect than mAGO2-bound guides due to the abundance of the AGO proteins in the gut tissue. Analysis of the sRNA guides in this Chapter was carried out by Dr. Jose Roberto Bermudez-Barrientos and Dr. Cei Abreu-Goodger. All reads were mapped simultaneously to the nematode and mouse genome using ShortStack to define clusters of locations on both genomes. ShortStack allows for reads that map with up to 1 nt mismatch better to the *H. bakeri* genome to be assigned there while the same applies for the other genome. If a read maps equally well to more than one location, including one location in the worm and one location in the mouse genome, then the read will be assigned to the location based on how many uniquely-placed reads are nearby. In cases where multiple locations cannot be distinguished by uniquely-placed reads, the read will get assigned randomly.

Clusters of 18-25 nt identified in exWAGO IPs from infected gut tissue mapping to the *H. bakeri* genome were analysed for length distribution and first nucleotide. The results show that most of these clusters are 22 nt in length and start with a G (Fig. 4.8). The characteristics of the clusters are reminiscent of the exWAGO guides identified in Chapter 3. The profile of the plots is almost identical between the samples that were not treated with RNase compared to those that were. Interestingly, the parasite reads found in the control samples (where exWAGO is immunoprecipitated from uninfected guts) also are 22G RNAs. However, the number of these clusters identified is very small (Fig. 4.9A). The reads making up these clusters can be considered as background molecules which stick to the beads and/or the antibody during the IP as exWAGO is absent from the uninfected gut tissue. These reads could also represent contamination from other samples due to index ‘hopping’/switching (Griffiths *et al*, 2018).

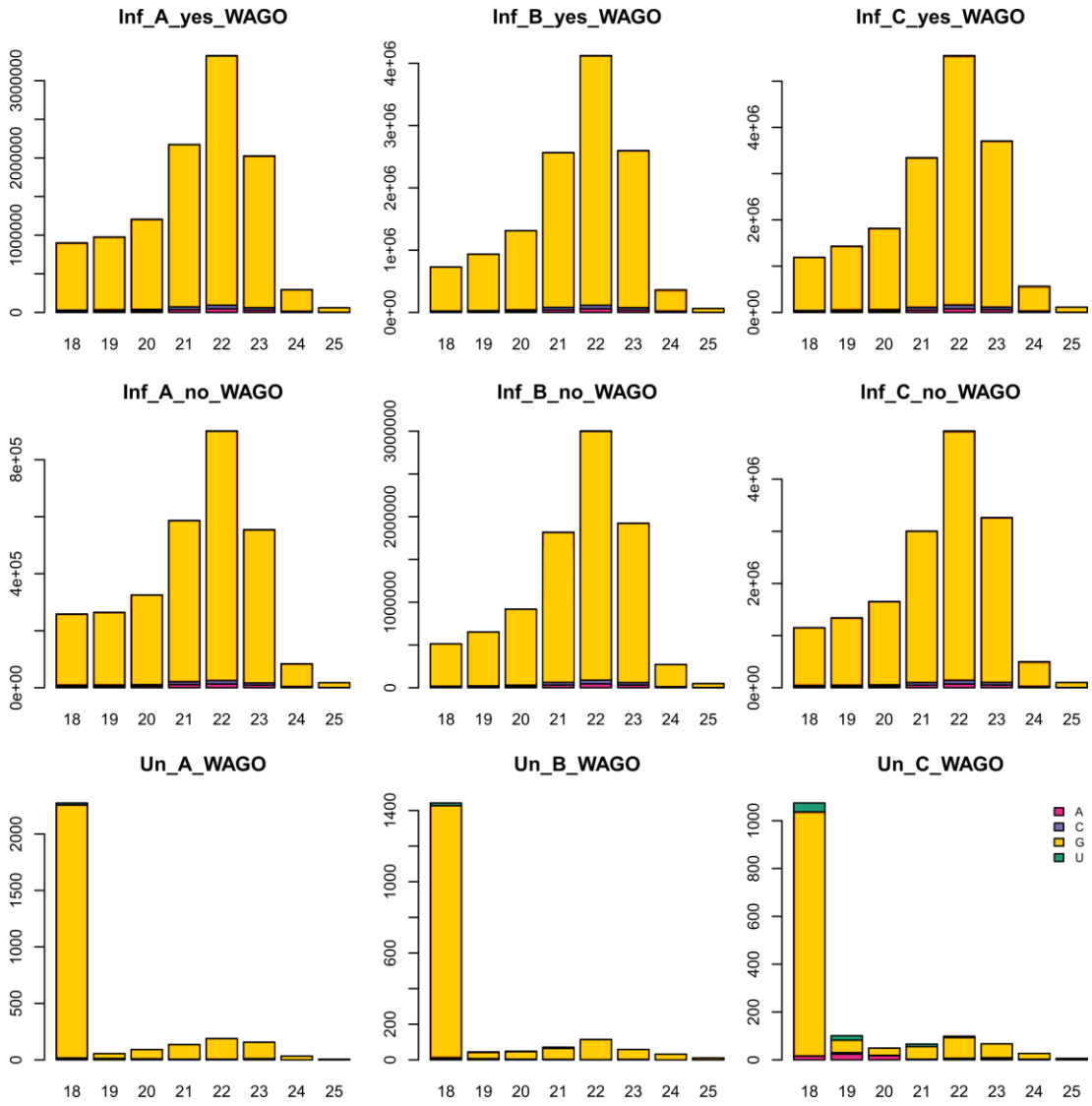


Figure 4.8 | Length distribution and first nucleotide plot analyses of 18-25 nt clusters mapping to the *H. bakeri* genome after ShortStack analysis of the sRNAs detected in the CLASH libraries from gut tissue exWAGO IPs. The x-axis represents the length of the reads in nucleotides and the y-axis shows the number of clusters. The figure was generated by Dr. Cei Abreu-Goodger. Inf = infected gut tissue, Un = Uninfected gut tissue, A/B/C = replicate number, WAGO = exWAGO, yes/no = refers to whether RNace-IT treatment was performed.

The reads mapping to the *H. bakeri* genome were further categorised to annotated regions of the *H. bakeri* genome (Fig. 4.9). From the results in Chapter 3, we expect that the reads would map to transposons and novel repeats and be depleted of miRNAs, tRNAs, rRNA and Y-RNAs. Indeed, this is what we observe (Fig. 4.9), suggesting that exWAGO IPs from infected gut tissue can be implemented to identify the sRNA guides that associate with it. Intriguingly, some reads map to the worm genome after immunoprecipitation with mAGO2. These mainly map to miRNAs (Fig. 4.9B). In the future, it will be of interest to investigate if these miRNAs are true worm sequences as this result would suggest that worm miRNAs have the ability to hijack the host mAGO2 machinery.

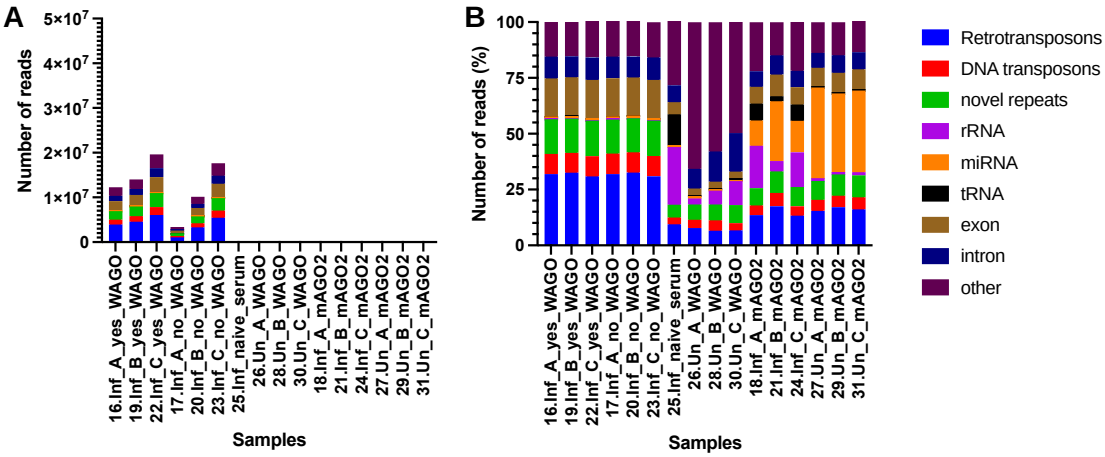


Figure 4.9 | RNA biotypes identified from the reads mapping only to the worm genome with up to 1 mismatch. The data is shown as counts (A) or as percentage (B). Inf = infected gut tissue, Un = Uninfected gut tissue, A/B/C = replicate number, WAGO = exWAGO, yes/no = refers to whether RNase treatment was performed. The number in the sample name indicates the index used during RNA library preparation. The numbers for the read categories were generated by Dr. Cei Abreu-Goodger.

ExWAGO exists in several environments and in Chapter 3 we found that even the vesicular and non-vesicular exWAGO forms bind guides with different abundance and guides that are unique to each. The infected gut samples are thought to represent all forms of exWAGO: (1) the exWAGO that is inside the *H. bakeri* adult worms present in the host tissue, (2) the vesicular form outside the parasite that has not been internalised by host cells, (3) the non-vesicular extracellular form outside the parasite

that has not been internalised by host cells, and (4) the exWAGO protein that is internalised in the mouse host cells. It is thus important to understand if the exWAGO immunoprecipitated from *in vivo* gut samples represents a particular form of the protein. To achieve this, reads of 20-25 nt length mapping to the *H. bakeri* genome only from exWAGO IP samples generated for sRNA sequencing discussed in Chapter 3 (Table 3.1), and from exWAGO IPs from *H. bakeri*-infected gut executed by myself using the modified CLASH protocol and by Mr. Xiaochen Du who generated sRNA sequencing libraries, were collated and analysed by Dr. Cei Abreu-Goodger. The results indicate that the first dimension of the MDS plot separates the sRNAs by protocol (CLASH vs sRNA libraries) (Fig 4.10). The second dimension of the MDS plot separates the sRNAs according to the environment they were detected in (i.e. worms, EVs, EV-depleted HES). Figure 4.10 shows that the sRNAs derived from the *in vivo* exWAGO IPs (Mmus_WIP_d14 & Inf_WIP_pp) cluster more closely to the exWAGO IPs from adult worms (Adult_WIP_pp), based on the fact that the sRNAs from these samples are in the same position in the second dimension of the MDS plot. This can be explained by the fact that there are still worms present in the gut tissues used in the *in vivo* exWAGO IPs which likely contributes the greatest amount of exWAGO compared to the secreted/excreted or host-internalised exWAGO complexes.

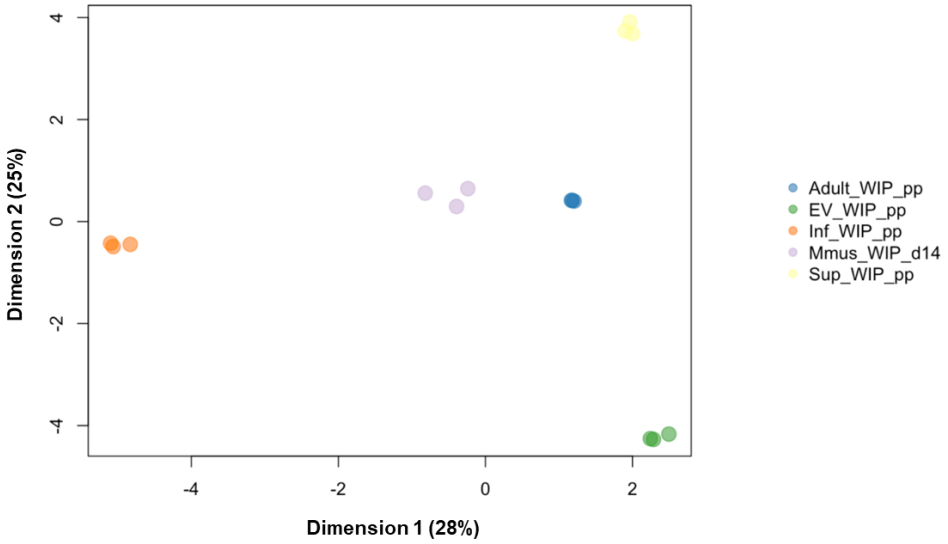


Figure 4.10 | MDS plot of small RNAs of 20-25 nt in length mapping to the worm genome only from *in vitro* and *in vivo* samples. All the samples used for the MDS plot are immunoprecipitations with anti-exWAGO antiserum and the RNAs were polyP treated. Inf_WIP_pp samples were prepared using the CLASH protocol while the rest of the samples were generated as sRNA libraries. The figure was generated by Dr. Cei Abreu-Goodger. Adult = *H. bakeri* worms, Mmus = gut tissue from infected mouse, Inf = *H. bakeri*-infected mouse gut tissue, Sup = EV-depleted HES, WIP = exWAGO IP, pp = treated with polyP, d14 = 14 days post-infection.

4.4 Identification and characterisation of exWAGO host targets *in vivo*

The vast majority of reads from currently published protocols are not chimeric (guide:target hybrid reads). In fact, only ~5% of the reads are guide:target chimeras (Helwak *et al*, 2013; Moore *et al*, 2015). Therefore, even though the intermolecular ligation step of the modified CLASH protocol promotes formation of chimeric reads (Fig. 4.3), the ligation step is sub-optimal. Additionally, the modified CLASH protocol has not been tested before with gut tissue samples, and without crosslinking the RNA to the AGO protein the AGO:guide:target interactions might be weak resulting in the loss of real targets. Nevertheless, even if the guide and target are not ligated due to technical reasons, targets can still be pulled down if the interactions with the guide and the AGO protein are strong and withstand the stringent washes performed in the protocol. Thus, host targets could be identified in a chimeric or non-chimeric read.

4.4.1 exWAGO host targets in a non-chimeric read

To detect mouse host targets that are not in a chimeric read, the following pipeline was applied by Dr. Sujai Kumar (Chapter 2, Section 2.14.2). Briefly, reads shorter than 18 nt and reads mapping perfectly to the *H. bakeri* genome were removed. Removing perfect matches to the nematode genome allows us to exclude nematode-derived sequences including nematode guide RNAs, since we are interested in identification of host targets. Reads mapping to mouse rRNAs were also removed. Any reads mapping to the rest of the mouse genome (C57BL/6J mouse strain) with up to 1 nt mismatch were kept. Then ShortStack was run simultaneously on all samples to define clusters. Mouse miRNA clusters were removed in an attempt to remove host guides. Then I explored the remaining clusters of putative host targets (23,881 clusters) using the online tool Degust (Powell, 2019). The Degust file can be accessed here: https://degust.erc.monash.edu/degust/compare.html?code=3ea6602f802df716e7c3e352ddd3c580#/. A summary of the bioinformatic pipeline described above is summarised in Figure 4.11.

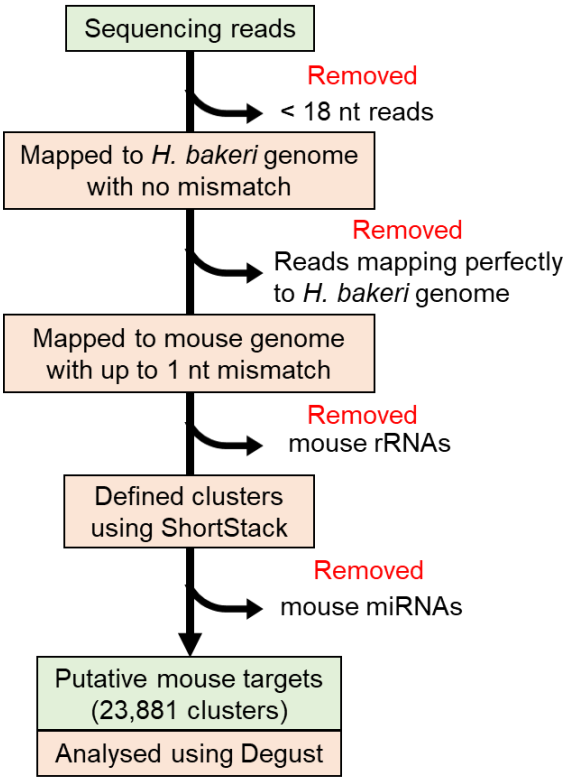


Figure 4.11 | Summary of the bioinformatics pipeline applied to identify mouse targets in non-chimeric reads.

To test whether the *in vivo* gut IP method developed here used in conjunction with the modified CLASH technique have indeed successfully pulled down putative mouse targets in non-chimeric reads, I first analysed the clusters to identify mAGO2-specific mouse genes that are targeted during *H. bakeri* infection. We expect that the mAGO2 will target mainly the 3'UTR and CDS regions of gene transcripts as per the literature. Using the Degust interface, the false discovery rate (FDR) cut-off was set to 0.01 and the fold change was set to average. Clusters enriched in the mAGO2 IP from infected gut tissue but depleted in control samples were selected (Fig. 4.12A). Controls were included to attempt to identify targets specific to mAGO2 during *H. bakeri* infection. The control samples are: (1) exWAGO IP from infected gut tissue – this controls for molecules upregulated during infection that are not specific to mAGO2 but are targeted by exWAGO; (2) naïve rat serum IP from infected gut tissue – this controls for molecules that are upregulated during infection that stick to beads; (3) exWAGO IP from uninfected tissue – this controls for mouse molecules that stick to the beads and/or the exWAGO antibody when exWAGO is not present; (4) exWAGO IP from adult worms performed previously by Dr. Franklin Chow as the libraries generated in Chapter 3 were unavailable at the time of data analysis and controls for molecules

that are most likely nematode guides if they map to both mouse and nematode genome. The selection resulted in the identification of 5,396 unique clusters (Supp. Table 1). The 5,396 clusters range in length from 18-272 nt and most of them are 24 nt long (184 out of 5,396; 3.4%) (Fig. 4.12B). The clusters also have a high number of reads and there is a moderate preference for the clusters to start with the base A (Fig. 4.12C&D).

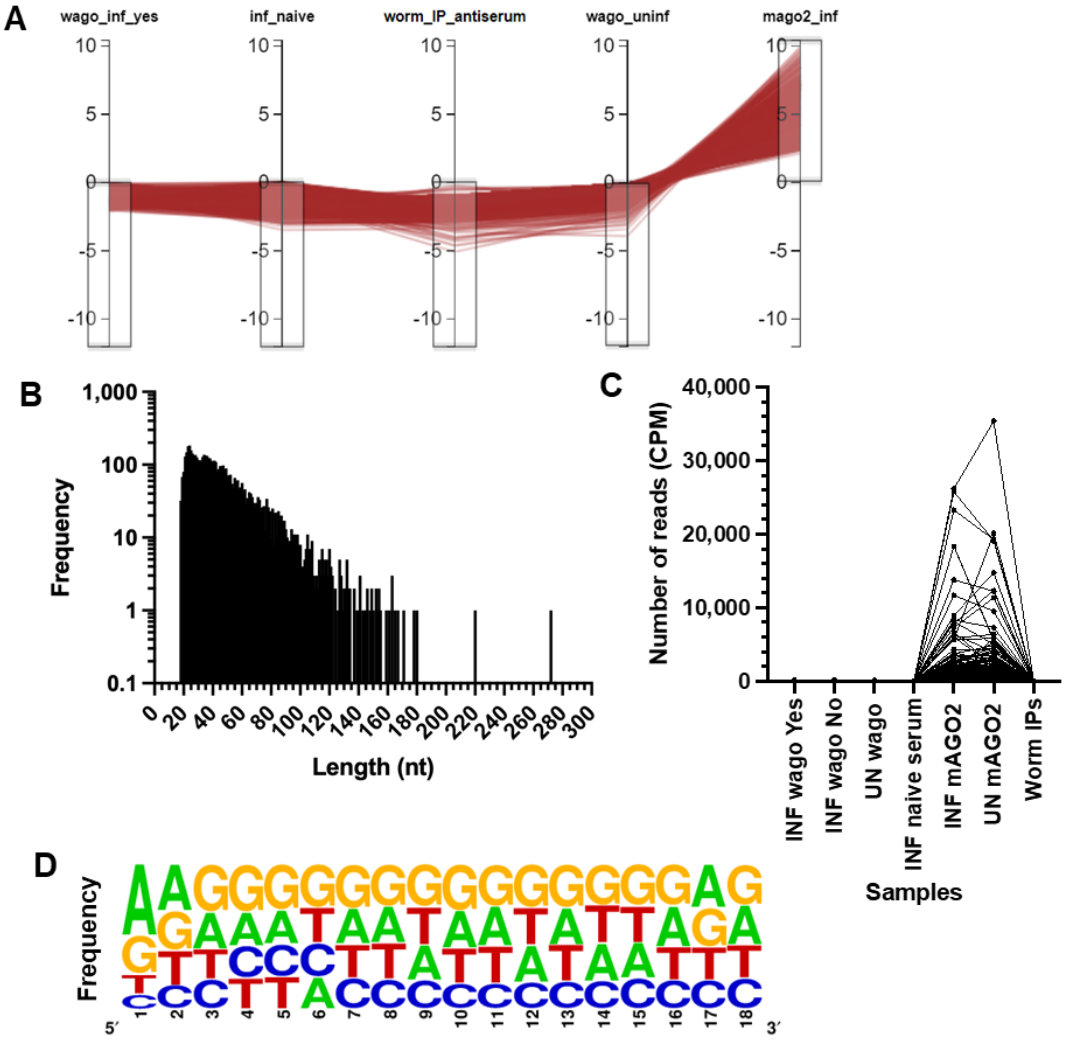


Figure 4.12 | Selection and characterisation of the 5,396 clusters identified as putative mAGO2 targets.

A) Selection of putative mAGO2 clusters on Degust. B) Length distribution of the clusters. C) The number of reads in counts per million (CPM) shown as average per sample per cluster (n=3 for all samples except for INF naïve serum sample where n=1). D) Frequency plot of the first 18 nucleotides of the clusters created using WebLogo (Crooks *et al*, 2004). WAGO = exWAGO, INF = infected gut, UN = uninfected gut, Yes = RNase treated, No = was not RNase treated.

Mapping to canonical and non-canonical transcripts reveals that 94.5% of the clusters (5,100 out of 5,936) map onto annotated genomic regions while only 5.5% of the clusters (296 out of 5,936) map onto non-annotated regions (Fig. 4.13). Most of the classified clusters map onto regions annotated as non-regulatory regions only (3,703 out of 5,936; 68.6%) (Fig. 4.13). Non-regulatory regions include 3'UTR, CDS, introns and 5'UTR. Only 2.3% (125 out of 5,936) of the clusters map exclusively to regulatory regions (Fig. 4.13), which include regions that could be implicated in gene regulation such as promoters, enhancers and transcription factor binding sites. The remaining 23.6% (1,272 out of 5,936) of the clusters map on genomic areas that are classed both as a regulatory and non-regulatory (Fig. 4.13). Further analysis of the biotypes of the clusters annotated as regulatory regions only indicates that 49.6% map onto promoter regions (62 out of 125) (Fig. 4.13). Analysis of clusters annotated as non-regulatory regions only indicates that they map onto CDS (1,076 out of 3,703; 29.1%) and 3'UTR regions (936 out of 3,703; 25.3%) (Fig. 4.13).

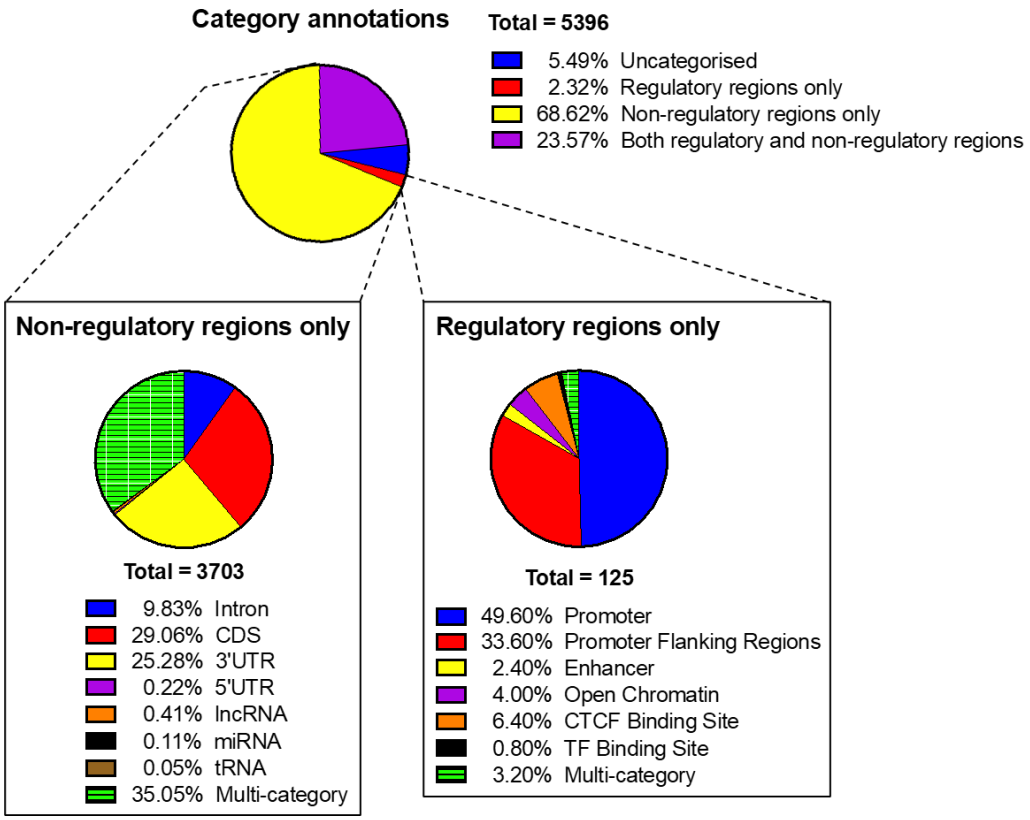


Figure 4.13 | Annotation of the 5,396 clusters identified as putative mAGO2 targets after mapping to canonical and non-canonical regions of the mouse genome.

Analysis of mAGO2 targets reveals that their characteristics are consistent with the literature and function of mAGO2 which is reported to regulate gene expression by targeting the 3'UTR and CDS regions of gene transcripts in the cytoplasm (Sarshad *et al*, 2018). These findings suggest that the method developed and used here is valid. Thus, a similar analysis was carried out with exWAGO, to identify exWAGO-specific mouse genes that are targeted during *H. bakeri* infection. Using the Degust interface, the false discovery rate (FDR) cut-off was set to 0.01 and the fold change was set to average. Clusters enriched in the exWAGO IP from infected gut tissue (libraries with RNase treatment) but depleted in control samples were selected (Fig. 4.14A). Controls were included to attempt to identify targets specific to exWAGO during *H. bakeri* infection. The control samples are: (1) naïve rat serum IP from infected gut tissue – this controls for molecules that are upregulated during infection that stick to beads; (2) exWAGO IP from uninfected tissue – this controls for mouse molecules that stick to the beads and/or the exWAGO antibody when exWAGO is not present; (3) exWAGO IP from adult worms performed previously by Dr. Franklin Chow as the libraries generated in Chapter 3 were unavailable at the time of data analysis and controls for molecules that are most likely nematode guides if they map to both mouse and nematode genome; and (4) mAGO2 IP from infected gut tissue – this controls for molecules upregulated during infection that are not specific to exWAGO but are targeted by mAGO2. The selection resulted in the identification of 911 unique clusters (Supp. Table 2). The 911 clusters range in length from 18-53 nt and most of them are 19 nt long (192 out of 911, 20%) (Fig. 4.14B). The clusters also have a high number of reads and there is a strong preference for the clusters to start with the base G (Fig. 4.14C&D). The first nucleotide preference of the putative exWAGO mouse targets is striking and led us to question whether these are parasite guides rather than mouse targets, as *H. bakeri* sRNA guides also preferentially start with the base G (Chapter 3).

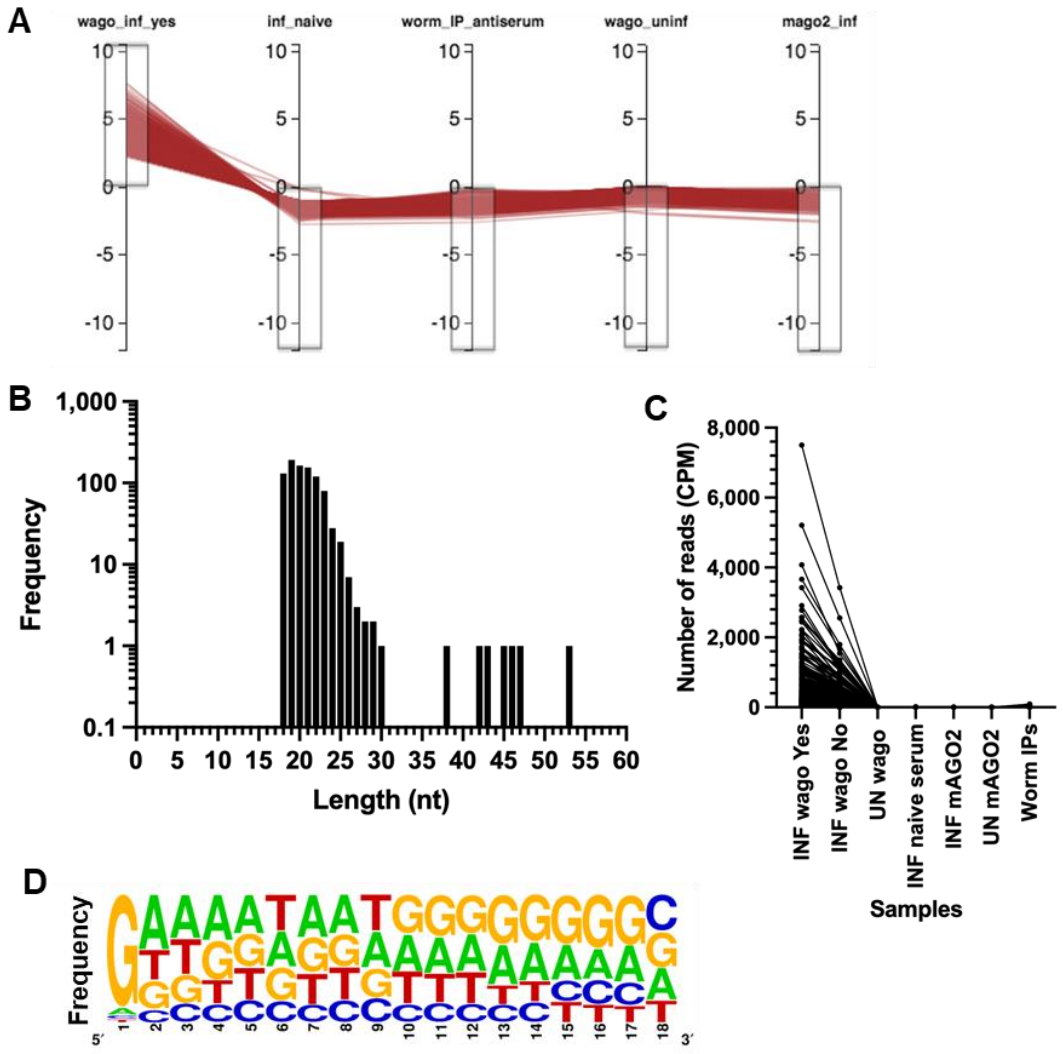


Figure 4.14 | Selection and characterisation of the 911 clusters identified as putative exWAGO targets.

A) Selection of putative exWAGO clusters on Degust. B) Length distribution of the clusters. C) The number of reads in counts per million (CPM) shown as average per sample per cluster (n=3 for all samples except for INF naïve serum sample where n=1). D) Frequency plot of the first 18 nucleotides of the clusters created using WebLogo (Crooks *et al*, 2004). WAGO = exWAGO, INF = infected gut, UN = uninfected gut, Yes = RNase treated, No = was not RNase treated.

To understand whether the reads that make up the clusters identified in exWAGO are truly derived from mouse and are not parasite guides, we explored how the reads map to the mouse genome. Comparison of some of the exWAGO 911 and mAGO2 5,396 clusters by qualitative visual exploration using the Integrative Genomics Viewer (IGV) tool (Robinson *et al*, 2011), reveals that most exWAGO clusters map to genomic features with a mismatch. These mismatches are persistent through the reads that make up these clusters and do not resemble sequencing errors. In contrast, most mAGO2 clusters map perfectly to genomic features and some reads have mismatches which resemble sequencing errors. More precisely, quantitative analysis by Dr. Sujai Kumar identified that out of the 72,681 reads that make up the 911 exWAGO clusters, only 4,455 (6%) of them map perfectly to the C57BL/6J mouse genome (Fig. 4.15). The remaining 94% map with one nucleotide mismatch (Fig. 4.15). In contrast, 85% (1,346,876 out of 1,591,947) of the reads that make up the 5,396 mAGO2 clusters map perfectly (Fig. 4.15). Although these results along with the observation that most exWAGO clusters start with a G suggest that the exWAGO clusters identified as putative mouse targets are in fact not derived from the mouse but rather from the parasite, it is remarkable that these reads map to host targets with near-perfect complementarity. Further quantitative assessment of how the reads map to the genomic loci identified as putative targets of exWAGO and mAGO2 should be performed to understand if the reads map sense or antisense. Reads that map to genomic loci in a sense direction are thought to be sequences that come from that region and are thus considered targets. Reads that map in an antisense direction to genomic loci are possibly sequences that act as guides. This information could aid in determining if the reads detected as putative targets of exWAGO are true mouse targets or parasite sequences.

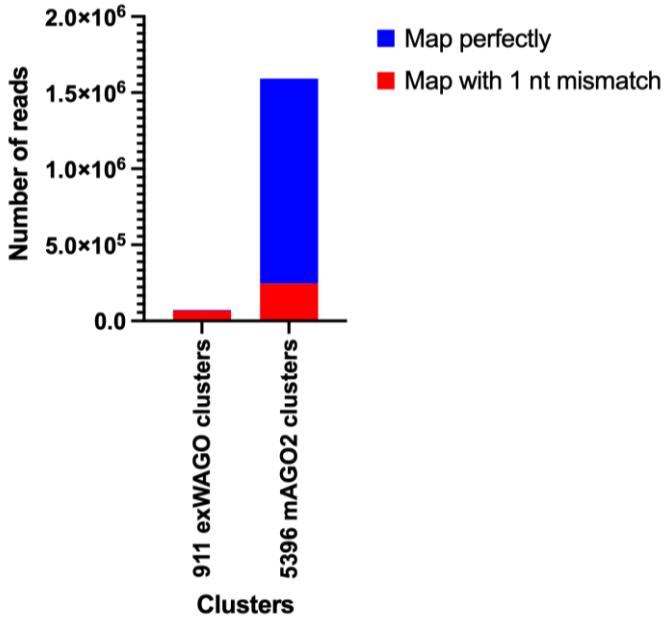


Figure 4.15 | The number of reads that map perfectly or with 1 nucleotide mismatch to the mouse genome from the clusters identified as putative mouse targets of exWAGO and mAGO2.

To continue with, mapping to canonical and non-canonical transcripts reveals that a substantial number of the exWAGO clusters map onto non-annotated genomic regions (400 out of 911, 44%) while the remaining clusters map onto classified regions of the genome (511 out of 911, 56%) (Fig. 4.16). Most of the classified clusters map onto regions annotated as non-regulatory regions only (315 out of 911, 35%) (Fig. 4.16). Regulatory regions make up 8% (70 out of 911) of the clusters (Fig. 4.16). The remaining 14% (126 out of 911) of the clusters map on genomic regions that are classed both as regulatory and non-regulatory (Fig. 4.16). Further analysis of the biotypes of the clusters annotated as regulatory regions only indicates that most map onto promoter flanking regions (26 out of 70, 37%) (Fig. 4.16). Analysis of clusters annotated as non-regulatory regions only indicates that they largely map onto introns (274 out of 315, 87%) (Fig. 4.16). The differences in the characteristics between the putative mAGO2 and exWAGO targets identified suggests that exWAGO functions differently to mammalian AGOs, assuming that the exWAGO clusters are true host targets.

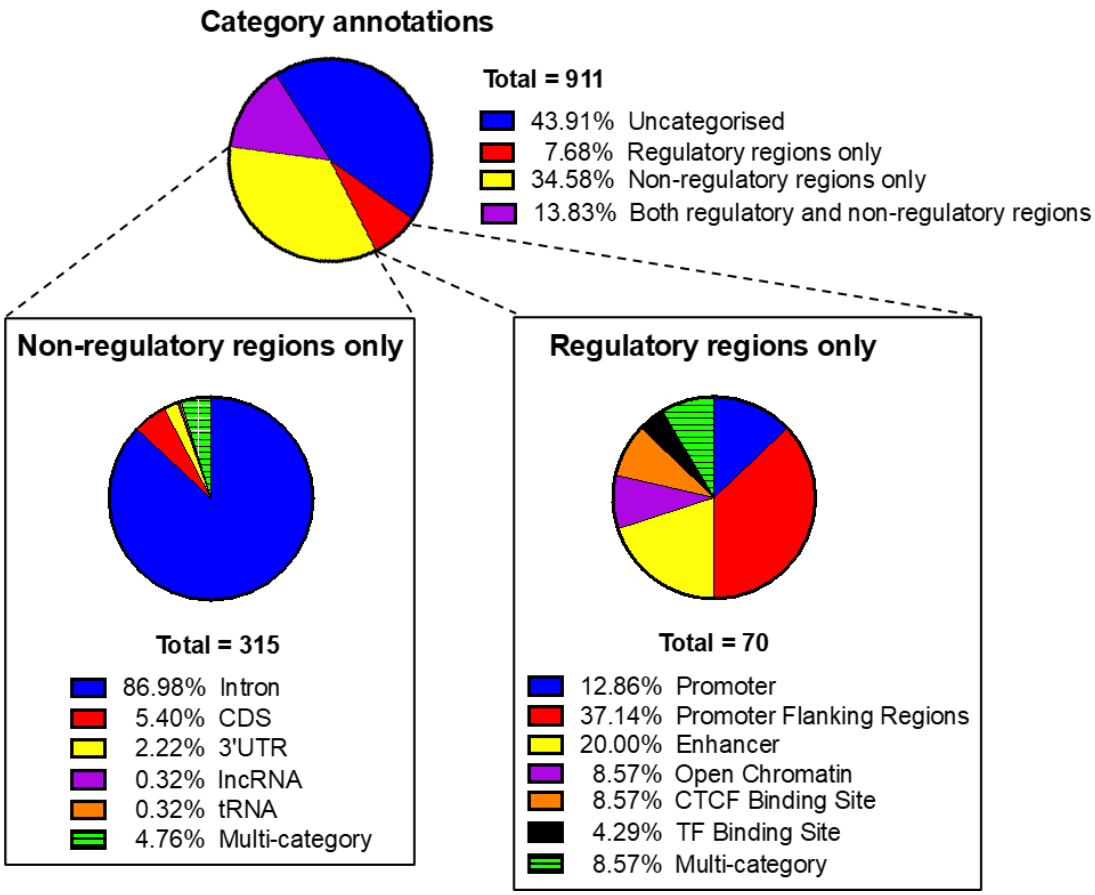


Figure 4.16 | Annotation of the 911 clusters identified as putative exWAGO targets after mapping to canonical and non-canonical regions of the mouse genome.

Network analysis based on the Gene Ontology (GO) biological process terms was performed using the genes identified in the clusters categorised as non-regulatory only and the clusters that map to both non-regulatory and regulatory regions (441 clusters in total). This analysis revealed the enrichment of various terms including processes related to the epithelium, nervous system, and transcriptional regulation (Table 4.2, Supp. Table 4.3). Although we cannot be certain that the clusters identified here are true exWAGO targets, the processes identified enriched in the GO analysis are deemed to be relevant to gastrointestinal worm infections. For example, during an intestinal helminth infection, the worms are in close contact with host epithelial cells and trigger the expansion of intestinal epithelial tuft and goblet cells and increase epithelial permeability (Mckay *et al*, 2017; Sorobetea *et al*, 2018). The increased permeability of the epithelium could allow immune cells (such as macrophages and eosinophils) and immune molecules (such as antibodies) to reach the gut lumen to

attack the parasite (Mckay *et al*, 2017). It could thus be advantageous to the parasite to limit migration of epithelial cells and regulate cell-cell adhesion (Table 4.2, Supp. Table 4.3). Furthermore, gastrointestinal parasites are also known for disrupting the enteric nervous system (Halliez & Buret, 2015). More specifically, gastrointestinal worms can induce nerve remodelling, neuron and axon degradation, alter ion and fluid transport, and can also affect the contractility of the intestinal muscle (Halliez & Buret, 2015). Thus, identification of processes involved in the regulation of ion transmembrane transport, axonogenesis, and synaptic assembly and transmission (Table 4.2, Supp. Table 4.3) could indicate that exWAGO and its sRNA guides might be employed by *H. bakeri* to manipulate the host enteric nervous system to its advantage. It is of interest to validate such findings and understand whether other excreted/secreted parasite molecules affect the aforementioned biological processes.

The GO network analysis performed for the putative targets of exWAGO yielded different GO biological terms compared to the mAGO2 putative targets. Analysis for the mAGO2 clusters was performed similarly to exWAGO, by analysing the clusters categorised as non-regulatory only and the clusters that map to both non-regulatory and regulatory regions (4,975 clusters in total). Interestingly, terms related to immune regulation pathways such as production of immunoglobulins, phagocytosis, B-cell activation and B-cell receptor signalling are found significantly depleted (FDR < 0.05) in the dataset (Supp. Table 4.4). This would suggest that the genes involved in immunity are less targeted by mAGO2, further implying that these genes would not be suppressed by mAGO2, and their expression would be upregulated. Upregulation of immune pathways by the host would be required to fight the parasite.

4.4.2 Analysis of chimeric reads from *in vivo* gut IPs

Bioinformatic analysis of chimeric reads is still under way. We predict that optimisation of the CLASH protocol from gut tissue might be required. This is required to obtain reproducible chimeric reads in biological replicates that are real and have not been produced by random ligation, according to CLASH or CLEAR-CLIP data generated by other lab members.

Table 4.2 | Gene Ontology biological process term analysis of the annotated clusters identified in the 911 putative exWAGO targets. The table is ranked by fold enrichment and only shows the classes identified and over-represented GO terms.

GO biological process	No. of genes	Fold enrichment	FDR (<0.05)
Innate vocalization behaviour (GO:0098582)	3	55.95	1.66E-02
Regulation of pre-synapse assembly (GO:1905606)	6	8.83	1.75E-02
Regulation of establishment of planar polarity (GO:0090175)	6	7.99	2.44E-02
Post-synapse organization (GO:0099173)	10	5.71	4.63E-03
Positive regulation of synapse assembly (GO:0051965)	8	5.59	2.22E-02
Cerebellum development (GO:0021549)	10	5.13	9.07E-03
Cell growth (GO:0016049)	9	4.75	2.54E-02
Learning (GO:0007612)	14	4.3	2.95E-03
Memory (GO:0007613)	12	4.25	9.13E-03
Protein autophosphorylation (GO:0046777)	14	4.04	4.59E-03
Positive regulation of epithelial cell migration (GO:0010634)	11	3.97	2.38E-02
Cell-cell adhesion via plasma-membrane adhesion molecules (GO:0098742)	13	3.77	1.31E-02
Positive regulation of GTPase activity (GO:0043547)	16	3.76	3.26E-03
Positive regulation of synaptic transmission (GO:0050806)	11	3.73	3.53E-02
Locomotory behaviour (GO:0007626)	16	3.58	4.64E-03
Negative regulation of cell projection organization (GO:0031345)	13	3.48	2.24E-02
Axonogenesis (GO:0007409)	18	2.84	1.75E-02
Chemical synaptic transmission (GO:0007268)	17	2.71	3.61E-02
Regulation of ion transmembrane transport (GO:0034765)	24	2.71	4.38E-03
Regulation of neuron projection development (GO:0010975)	26	2.68	3.08E-03
Heart development (GO:0007507)	27	2.64	2.85E-03
Regulation of membrane potential (GO:0042391)	21	2.54	2.08E-02
Regulation of developmental growth (GO:0048638)	18	2.54	4.73E-02
Developmental growth (GO:0048589)	20	2.53	2.70E-02
Positive regulation of organelle organization (GO:0010638)	27	2.52	4.48E-03
Morphogenesis of an epithelium (GO:0002009)	22	2.35	3.97E-02
Actin filament-based process (GO:0030029)	24	2.31	2.72E-02
Tube morphogenesis (GO:0035239)	30	2.27	1.04E-02
Cell migration (GO:0016477)	36	2.24	3.05E-03
Regulation of MAPK cascade (GO:0043408)	28	2.19	2.24E-02
Regulation of secretion by cell (GO:1903530)	26	2.17	3.71E-02
Anatomical structure formation involved in morphogenesis (GO:0048646)	34	1.99	3.17E-02
Intracellular signal transduction (GO:0035556)	48	1.99	3.25E-03
Embryo development (GO:0009790)	41	1.9	1.93E-02
Positive regulation of transcription, DNA-templated (GO:0045893)	51	1.85	7.90E-03
Macromolecule localization (GO:0033036)	68	1.66	8.31E-03
Cellular localization (GO:0051641)	66	1.63	1.76E-02

Discussion

4.5 Summary

In this chapter, I developed and optimised a technique to immunopurify exWAGO and mAGO2 from *H. bakeri*-infected mouse gut (Fig. 4.1). This technique was then employed in conjunction with a modified version of the CLASH protocol, to directly detect sRNA guides and their target mRNAs *in vivo* (Fig. 4.3). Initially we analysed the guides pulled down with exWAGO and mAGO2 from *in vivo* gut tissue. The data indicate that this technique successfully identifies the sRNAs that exWAGO and mAGO2 associate with (Fig. 4.7, 4.8 and 4.9).

The CLASH protocol involves generation of RNA-RNA hybrids through ligation of the guide and its target whilst bound by the AGO protein. Bioinformatic analysis of RNA hybrids is complex and requires building a pipeline specific for these samples. At the present time, the bioinformatics pipeline for analysis of *H. bakeri-Mus musculus* chimeric reads is under development, hence data from chimeric reads are not discussed in this thesis. Nevertheless, since published chimeric datasets only result in up to ~5% chimeras, the other reads sequenced in the dataset are valuable for identifying targets. Analysis of targets in non-chimeric reads of the mammalian mAGO2 suggests that the method is valid since the reads in mAGO2 IPs mainly map to 3'UTR and CDS regions consistent with the literature (Fig. 4.13). Examination of exWAGO targets, however, reveals that they come mostly from intronic regions and have a strong propensity to start with G, reminiscent of exWAGO guide characteristics (Fig. 4.14 and 4.16). Combined with the fact that almost none of the clusters identified map to the mouse genome perfectly (Fig. 4.15), the data suggest that these reads might not be true exWAGO host targets and/or that exWAGO functions in an unconventional way using mechanisms that are not reported to be employed by mammalian AGOs, such as mAGO2.

4.6 exWAGO putative targets

Early studies of miRNA target gene identification in animals demonstrated that target specificity is largely determined by the seed region of miRNAs (Lai, 2002). While one miRNA can target several mRNAs, one gene transcript can be targeted by several miRNAs (Bartel, 2009). The 3'UTRs of mRNA transcripts have been regarded as regions where the presence of binding sites for miRNAs can affect translation by inducing mRNA degradation. For a long time, 3'UTRs were the only mRNA regions considered to be functional targets of miRNAs (Mazumder *et al*, 2003; Mayya &

Duchaine, 2019). Further research, however, proved that there are other mRNA regions that can be targeted by sRNAs to successfully repress gene expression. For example, Lytle *et al*, (2007) showed that 5'UTRs also carry miRNA seed sites. miRNA targeting of the 5'UTR or the 3'UTR of the mRNA transcript represses gene expression with similar efficiency (Lytle *et al*, 2007). The CDS region of genes also contains miRNA-binding sites and these are more efficient at inhibiting translation rather than triggering target degradation (Fang & Rajewsky, 2011; Hausser *et al*, 2013). Groups utilising AGO-CLIP (Cross-linking and Immunoprecipitation) experiments of mammalian AGOs from whole cells or tissue, report that most targets map to 3'UTR or the CDS (Helwak *et al*, 2013; Broughton *et al*, 2016). This is consistent with the results obtained for the mAGO2 IPs in this project (Fig. 4.13).

Examination of putative exWAGO host targets indicates that most reads map to introns (Fig. 4.16). Interestingly, some research groups also detect introns in sRNA chimeric reads (Moore *et al*, 2015; Sarshad *et al*, 2018). Intronic binding by sRNAs is expected to relate to nuclear targeting by Argonaute proteins (Moore *et al*, 2015; Sarshad *et al*, 2018). Indeed, AGO2-CLIP after isolation of the nuclear fraction from human cells indicates that miRNAs associating with nuclear human AGO2 target mainly intronic sites rather than 3'UTRs or CDS miRNA binding sites (Sarshad *et al*, 2018). The fact that other AGO proteins target introns mainly in the nucleus suggests that the putative exWAGO targets identified here might be real and that exWAGO might act mainly in the nucleus. The idea that AGO proteins can target introns to regulate gene expression remains under-researched. However, the Kornblihtt group identified that association of siRNAs and the human AGO1 with introns near an alternative exon or enhancers can control alternative splicing (Alló *et al*, 2009, 2014). Additionally, the splice sites of mammalian introns generally have GU-AG boundaries meaning that introns start with a 5'GU and have an AG at their 3' end (Burset *et al*, 2001). It would thus be of interest to: (1) investigate if the putative exWAGO intronic targets come from the intron splice site which would explain the preference of these reads to start with G, (2) understand if exWAGO regulates host gene expression by affecting alternative splicing, and (3) determine the localisation of exWAGO and hence understand whether exWAGO is localised to the nucleus of host cells.

Overall, the clusters identified as putative targets of exWAGO have odd characteristics, as mentioned before. However, the characteristics of the clusters identified as putative mAGO2 targets are consistent with the literature, suggesting

that the method applied here is valid for enquiring which genes exWAGO targets. Nevertheless, at present we cannot rule out that the reads identified as putative targets of exWAGO derive from the parasite. Currently, there are no (W)AGO-CLIP datasets available for any nematodes to compare with the exWAGO CLASH results and since exWAGO is the first AGO protein reported to be secreted in the host environment by parasitic worms it remains unknown how it would regulate host gene expression. Detection of exWAGO host targets is expected to be more challenging compared to the host AGO2. This is mainly based on the fact that to detect host targets and hence immunopurify exWAGO:guide:host target complexes, we need to purify the exWAGO protein that has been internalised by host cells. However, the gut tissues used in these experiments still have worm parasites. Hence, the majority of exWAGO pulled down in the immunoprecipitation is mostly from adult worms, as indicated in Section 4.3 (Fig. 4.10). We therefore expect that only a very low amount of exWAGO that has been internalised by host cells is pulled down in the modified CLASH experiment. The limited amount of internalised exWAGO immunopurified using these *in vivo* samples and the fact that it remains unknown how a parasite guide would interact with mouse targets render the goal of identifying exWAGO host targets challenging.

4.7 Analysis of exWAGO targets in a non-chimeric read

To identify mouse targets of exWAGO in a non-chimeric read, we removed reads that map perfectly to the nematode genome, but allowed mapping of reads with 1 nt mismatch to the mouse genome. It is important to remember, however, that although the *H. bakeri* genome has recently been improved after re-sequencing and annotation, it is still not fully complete or fully annotated (Chow *et al*, 2019). Some of the annotations could also be incorrect. This suggests that some true nematode sequences might not be detected by mapping the reads to the nematode genome. Additionally, it is possible that some of the reads that map to the mouse genome with 1 mismatch may map equally well to the nematode genome (i.e. with 1 mismatch). Moreover, there is also a complication in our analysis in that the sequenced reads derive from CBA x C57BL/6 F1 mice, compared to the genome used from mapping which is derived from C57BL/6J mice. This could explain why the putative exWAGO host targets map mostly with a nucleotide mismatch, but does not explain why mAGO2 targets map mostly perfectly (Fig. 4.16). The observation that the reads making up the mAGO2 targets map perfectly to the C57BL/6J genome could be explained by the fact that miRNA binding sites are expected to be conserved within

the same species. In the future, using *H. bakeri*-infected guts from a mouse strain that has an intermediate or slow rate of rejecting *H. bakeri* infection and whose genome is sequenced would control the issue of detecting mismatches due to the genome used for mapping not being from the same strain of mice as the mouse material used in the immunoprecipitation. Such mouse strains could be the C57BL/6NJ or the CBA/J strains (Keane *et al*, 2011; Reynolds *et al*, 2012).

By removing reads mapping to mouse miRNAs and nematode RNAs (including siRNAs) in this analysis, we removed guides to enrich the dataset with mouse targets (Fig. 4.11). But what if the exWAGO:parasite guides do not target mouse mRNA transcripts? Could exWAGO:parasite complexes act as sponges sequestering host miRNAs and hence preventing host miRNAs from regulating their targets? The filtering applied to the sequencing data does not enable us to test this alternative hypothesis of how exWAGO might regulate host gene expression.

4.8 Future work

Using the dataset generated here, we can also ask whether exWAGO can bind host guides and whether parasite guides can hijack the host mAGO2. According to data from the gel shift assays with mAGO2 and the 5'PPP nc16320 sequence (Chapter 3, Section 3.2.1) it is unlikely that mAGO2 will be binding 5'PPP guides, but it is possible that mAGO2 might bind the miRNAs secreted by the parasite, since those carry a 5'P feature. *In vitro*, exWAGO can bind 5'P guides with high affinity (Chapter 3, Section 3.2.2). However, as exWAGO is exported from the parasite pre-loaded with 5'PPP guides, it might be less likely that exWAGO would operate in host cells using host guides. Moreover, we can also use the datasets generated here to detect which sRNAs and targets associated with mAGO2 are upregulated during infection with *H. bakeri*. This could be explored by comparing the reads detected in the mAGO2 IPs from *H. bakeri*-infected tissue to those detected in the mAGO2 IPs from uninfected tissue.

Furthermore, data in this chapter clearly show that the *in vivo* gut IP method can be used as a technique to identify AGO guides from parasite species that cannot be cultured in the lab assuming the molecular tools (i.e. antibodies against AGO protein of interest) are available. As exWAGO is highly conserved amongst Clade V parasitic nematodes (Chapter 1, Table 1.1), this particular method employing the rat anti-exWAGO polyclonal antibody can be implemented to characterise the sRNA guides of the exWAGO orthologues present, for example, in *Heligmosomoides*

polygyrus which infects wild wood mice. In this way one could study the evolution of the sRNAs generated by these closely related nematodes.

Factors secreted by *H. bakeri* alter host gene expression. For example, work from our lab shows that incubation of total HES, EVs, and EV-depleted HES suppresses gene expression and protein production of alternative activation markers in Bone Marrow-Derived Macrophages (Coakley *et al*, 2017). Moreover, recent work by the Maizels lab shows that HES also affects gene expression of small intestinal organoids, even though the organoids were treated with HES from the basolateral side rather than from the apical side (Drurey *et al*, 2022). HES products inhibit tuft and goblet cell hyperplasia and downregulate the activity of genes that respond to the cytokines IL-4 and IL-13 (involved in differentiation of naïve T-cells to Th2 cells and induce B-cells to generate parasite-specific antibody isotypes) (Drurey *et al*, 2022). We hypothesise that some of these gene expression changes are driven by sRNAs and exWAGO. To test this hypothesis and identify which host genes are regulated by exWAGO:sRNA complexes we could employ a method that does not require extensive development of bioinformatic pipelines for data analysis. For example, we can deplete the non-vesicular exWAGO from HES or EV-depleted HES by exWAGO IP. The unbound fraction of the exWAGO IP could undergo several rounds of exWAGO IP to deplete the exWAGO protein. Then the exWAGO-depleted material can be incubated with mouse intestinal epithelial cells or intestinal organoids and then we can examine the changes in the expression of transcripts using RNA-seq (RNA sequencing). As a control, we can compare gene expression to cells or organoids that have been treated with total or EV-depleted HES that has been processed after naïve rat serum IPs. This is an indirect way to detect which transcripts are likely to be targeted by the non-vesicular exWAGO and its associated guides.

Although I have not yet identified exWAGO target genes with confidence, I have established a new technique that can be utilised to investigate other research questions to elucidate the function of exWAGO. In the next chapter I investigate the mechanism of action that exWAGO may employ to manipulate host gene expression. Amongst other techniques, I employ the method of exWAGO immunopurification from gut tissue developed here, to identify the protein-protein interactors of exWAGO *in vivo*.

Chapter 5: Functional analysis of exWAGO

Introduction

AGO proteins can function in various ways to regulate gene expression as discussed in Chapter 1, Section 1.7. Some AGOs silence genes by endonucleolytic cleavage (i.e. slicing) of the targets. The slicing mechanism involves an active site of specific amino acid motifs DD(D/H) and DED(D/H), often referred to as the catalytic triad and tetrad respectively (Park & Shin, 2014). However, not all AGO proteins which possess the catalytic motif exhibit slicer activity (Tolia & Joshua-Tor, 2007). Slicing also requires the presence of Mg^{2+} ions (Rivas *et al*, 2005), or other divalent metal ions (Kaya *et al*, 2016) and cleavage occurs in the target sequence between the 10^{th} and 11^{th} nucleotide position (counting from the 5' end of the guide sRNA that forms a duplex with the target) (Elbashir *et al*, 2001a, 2001b). Additionally, slicer activity requires the guide to be fully complementary to the target (beyond just the seed pairing) (Gebert & Macrae, 2019). However, most interactions of mRNA transcripts with miRNAs do not exhibit perfect complementary and are therefore not sliced but regulated via recruitment of other proteins by the AGO (Peters & Meister, 2007).

AGO proteins operate within molecular complexes that can involve tens of other proteins. Interaction with some proteins such as the heat-shock protein chaperone HSP90 and Dicer (involved in generation of sRNAs) is related to loading the guides in AGOs (Czech & Hannon, 2010; Iwasaki *et al*, 2010). More specifically, AGOs are part of the RNA-induced silencing complex (RISC) which functions in the cytoplasm. AGO is thought to form a minimal RISC-loading complex by interacting with Dicer (Ketting *et al*, 2001; Czech & Hannon, 2010) and the TAR RNA-binding protein (TRBP) during loading of the small RNA (MacRae *et al*, 2008). The proteins dissociate after AGO loading (MacRae *et al*, 2008). It is worth noting that the nematode WAGO proteins (secondary AGOs) are not expected to interact with Dicer. WAGOs bind 5'PPP RNAs which are generated by RNA-dependent RNA polymerases rather than Dicer. Moreover, AGOs also interact with other proteins relating to gene silencing, such as the GW182. The human AGO2 interacts with GW182 to achieve efficient translational repression (Lian *et al*, 2009). GW182 in turn recruits the CCR4:NOT and the PABP (polyadenylate binding protein 1) deadenylase complexes to mediate destabilisation of the mRNA transcript target, hence inducing its destruction (Behm-Ansmant *et al*, 2006; Jinek *et al*, 2010). Furthermore, several other proteins have been reported to interact with AGOs including helicases, chaperones, and LIM

domain proteins. Some of the proteins interacting with AGOs have been studied further to understand their function in gene silencing (Peters & Meister, 2007; Meister, 2013).

As reported from studies in *Schizosaccharomyces pombe*, AGO proteins also form the RITS complex (RNA-Induced Transcriptional Silencing complex) which acts in the nucleus to regulate gene expression at the epigenetic/transcriptional level (Volpe *et al*, 2002). In *S. pombe*, AGO1 interacts with the heterochromatin-associated protein Chp1 (chromodomain protein 1) which has a chromodomain that binds methylated histones, and the Tas3 protein (Targeting complex subunit 3) which is required for H3K9 methylation (Motamedi *et al*, 2004; Verdel *et al*, 2004). Interestingly, although AGO1 is a highly conserved protein among other organisms, Chp1 and Tas3 are not. Nevertheless, the mechanistic basis of these interactions in *S. pombe* seems to be conserved with other organisms, as the Tas3 protein possesses GW repeats (Bhattacharjee *et al*, 2019). This motif is present on other protein interactors of AGOs including the GW182 protein family which is known to interact with AGOs in mammalian cells as mentioned above (Pfaff & Meister, 2013). Thus, some of the protein interactors of AGOs carry similar features despite lacking protein homology. The aforementioned examples clearly indicate that AGO proteins operate in complexes and interactions of AGOs with particular proteins shape the mechanism(s) of gene regulation, and hence the functional outcome of the AGO:guide:target associations.

AGO protein interactors have been identified by employing immunoprecipitation experiments followed by proteomic analysis (Motamedi *et al*, 2004; Verdel *et al*, 2004; Meister *et al*, 2005; Friend *et al*, 2012; Frohn *et al*, 2012; Fabián Flores-Jasso *et al*, 2013) or via genetic screens (Kim *et al*, 2021), while functional studies demonstrate how these interactions dictate the mechanism(s) of action of AGOs. Proteomic analysis can be qualitative or quantitative. Most quantitative methods require labelling at the protein or peptide level. Labelling of proteins can be achieved by SILAC (Stable Isotope Labelling using Amino acids in Cell culture) where heavy ^{13}C or ^{15}N isotopes are metabolically incorporated during protein synthesis in the cultured cells (Jazurek *et al*, 2016). In systems where this technique is not feasible, chemical labelling using isotopic or isobaric tags to label peptides, rather than the protein during protein synthesis, can be applied (Jazurek *et al*, 2016). One such technique is dimethyl labelling. Dimethyl labelling is a cheap and fast labelling method where primary

amines (i.e. the *N*-terminus of the ε-group of lysines) are converted to dimethylamines (Boersema *et al*, 2009). Label-free analysis is also an option for quantitative proteomics and allows for reduced loss of material in comparison to dimethyl labelling for example, and also allows comparison of theoretically an unlimited number of samples (Bantscheff *et al*, 2007). However, as samples are analysed sequentially on the mass spectrometry instruments rather than simultaneously like in metabolic or isotopic/isobaric labelling, comparability between samples is more difficult and the run conditions (eg. temperature) per sample could differ slightly. To increase the confidence in results, replicates or additional controls are required (Bantscheff *et al*, 2007).

We hypothesise that exWAGO is internalised by host cells and interacts with host RNAi machinery to manipulate host gene expression. To determine if this hypothesis stands, in this chapter I test whether exWAGO has slicer activity and employ LC-MS/MS (Liquid Chromatography tandem Mass Spectrometry) analysis following immunoprecipitation of exWAGO to identify the protein interactors of exWAGO inside the host and inside adult worms. Understanding how exWAGO operates whilst still inside the parasite, may shed light as to whether it functions like other known AGOs.

Results

5.1 exWAGO does not possess slicer activity

Protein sequence alignment of exWAGO with other AGOs known to possess slicer activity shows that exWAGO lacks the catalytic motif (Fig. 5.1A). This suggests that exWAGO does not possess endonucleolytic activity and that it would not silence target transcripts through cleavage. To test this using biochemical techniques too, I exploited the ability of exWAGO to bind the miR-100 guide strongly *in vitro* (Kd = 3.26 nM ± 0.42 S.E.M) (Chapter 3, Fig. 3.14), and assessed whether it can cleave the fully complementary target of miR-100 in the presence of Mg^{2+} ions (Fig. 5.1B). The mAGO2, known for its ability to cleave mRNA targets, was used as a positive control in this assay. I used conditions were the miR-100 guide (25 nM) is bound by both exWAGO and the mAGO2 (25 nM). The data clearly indicate that exWAGO does not possess slicer activity, consistent with the absence of the cleavage motif in contrast to the mAGO2 which cleaves the target (Fig. 5.1C).

As the endogenous exWAGO mainly binds secondary siRNAs, I also tested if it could slice a target that I designed to be fully complementary to one of the most abundant siRNAs found associated to exWAGO in IPs (which derives from a locus termed nc16320) (Fig. 5.1D). exWAGO binds the 5'PPP nc16320 with high affinity (Kd = 9.20 nM ± 1.03 S.E.M) in contrast to the mAGO2 (Kd = 56.69 nM ± 8.53 S.E.M). As expected, exWAGO failed to induce cleavage. The mAGO2 also failed to induce cleavage (Fig. 5.1E). This is thought to be due to the low affinity mAGO2 has for the 5'PPP nc16320 guide and the fact that mAGO2 is only expected to bind less than 10% of the guide when used at 25 nM, as shown in Chapter 3, Fig. 3.12.

The above results indicate that exWAGO does not function by slicing mRNA targets. To investigate how it might otherwise manipulate host gene expression, I set out to identify the protein interactors of exWAGO using immunoprecipitation methods followed by proteomic analysis.

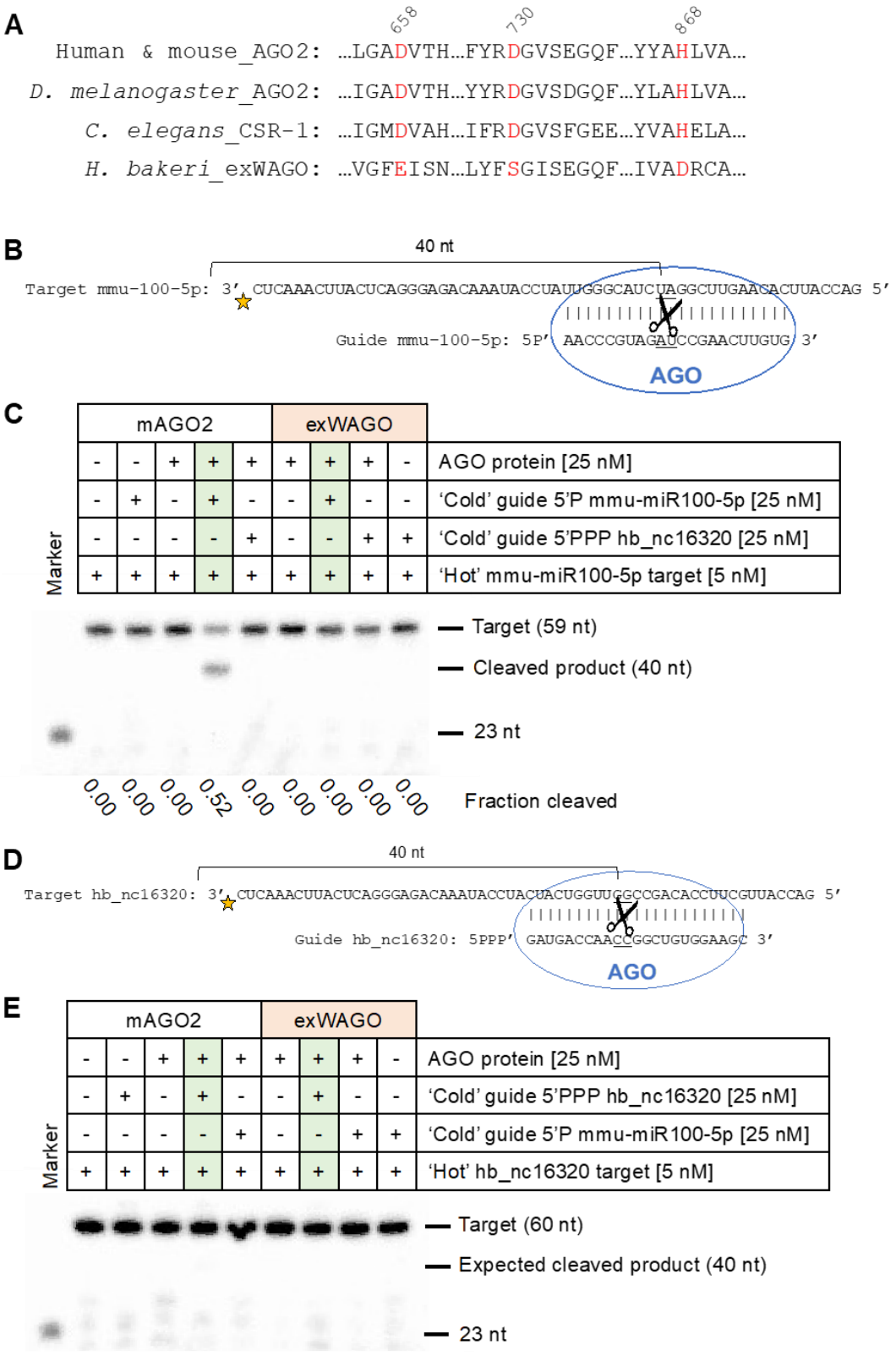


Figure legend on next page.

Figure 5.1 | exWAGO does not possess slicer activity

A) Protein sequence alignment showing catalytic residues of AGO proteins known to possess slicer activity (indicated in red). The numbers above the cleavage motif indicate the amino acid position based on the human and mouse AGO2 proteins.

B) Schematic representation of the mmu-miR100-5p target sequence radiolabelled at its 3' end (denoted by the star), and its fully complementary guide. Underlined nucleotides denote where cleavage is expected (represented by scissors) if sliced by an AGO. If slicing occurs, a 40 nt molecule will be generated. mmu denotes *Mus musculus.*

C) mAGO2 and exWAGO (25 nM) were loaded with either non-radiolabelled 5'P mmu-miR-100-5p or non-radiolabelled hb_nc16320 (25 nM) or no guide sequence and incubated with 3' end radiolabelled mmu-miR100-5p target (5 nM). The reactions were resolved on 15% (0.5x) TBE-Urea Sequagel and a radiolabelled 23 nt sRNA was used as a marker. The green columns represent the reactions where cleavage is expected. The fraction cleaved was calculated as described in methods chapter (Chapter 2). mmu denotes *Mus musculus*, hb denotes *Heligmosomoides bakeri*.

D) Schematic representation of the hb_nc16320 target sequence radiolabelled at its 3' end (denoted by the star), and its fully complementary guide. Underlined nucleotides denote where cleavage is expected (represented by scissors) if the AGO proteins tested here, exWAGO and mAGO2, possess slicer activity. If slicing occurs, a 40 nt molecule will be generated. hb denotes *Heligmosomoides bakeri*.

E) mAGO2 and exWAGO (25 nM) were loaded with either non-radiolabelled hb_nc16320 or non-radiolabelled mmu-miR-100-5p (25 nM) or no guide sequence and incubated with 3' end radiolabelled hb_nc16320 target (5 nM). The reactions were resolved on 15% (0.5x) TBE-Urea Sequagel and a radiolabelled 23 nt sRNA was used as a marker. The green columns represent the reactions where cleavage is expected. mmu denotes *Mus musculus*, hb denotes *Heligmosomoides bakeri*.

5.2 Identifying the exWAGO protein interactome in *H. bakeri* adult worms

Prior to understanding how exWAGO functions inside host cells, we sought to get an idea of the proteins it interacts with in the worm, which could relate to sRNA biogenesis/loading, gene silencing and also possibly export out of the parasite. Adult worms were harvested from CBA x C57BL/6 F1 mice after 14 days of infection and kept for 1 day in culture. The worms were extensively washed with PBS and then lysed by bead beating in the presence of RNase inhibitors. RNase inhibitors are used to avoid RNA degradation and hence reduce disruption of RNA-dependent protein interactions. exWAGO was immunoprecipitated from adult worm lysates (16.7-17.0 mg of worms by weight) (three replicates from worms harvested from F1 mice) for 45 mins using rat anti-exWAGO polyclonal antibody or naïve rat serum as a negative control (30 µl), conjugated on Protein G beads (125 µl). High salt washes were omitted to capture weaker protein-protein interactions. A small portion of the IP (10%) was eluted in denaturing buffer for quality control analyses using western blot and silver staining. Western blot analysis shows that exWAGO was successfully pulled down from worm lysates (Fig. 5.2). The rest of the IP material (90%) was processed for label-free quantitative LC-MS/MS as described in Chapter 2, Section 2.16.4 in the EdinOmics Facility at The University of Edinburgh by Ms Lisa Imrie. The peptides identified were mapped to the *H. bakeri* proteome. Supplementary table 5.1 contains all the peptides detected in this experiment. A summary of the significantly enriched peptides detected in the experimental samples is found in Table 5.1 and visualised using a volcano plot (Fig. 5.3).

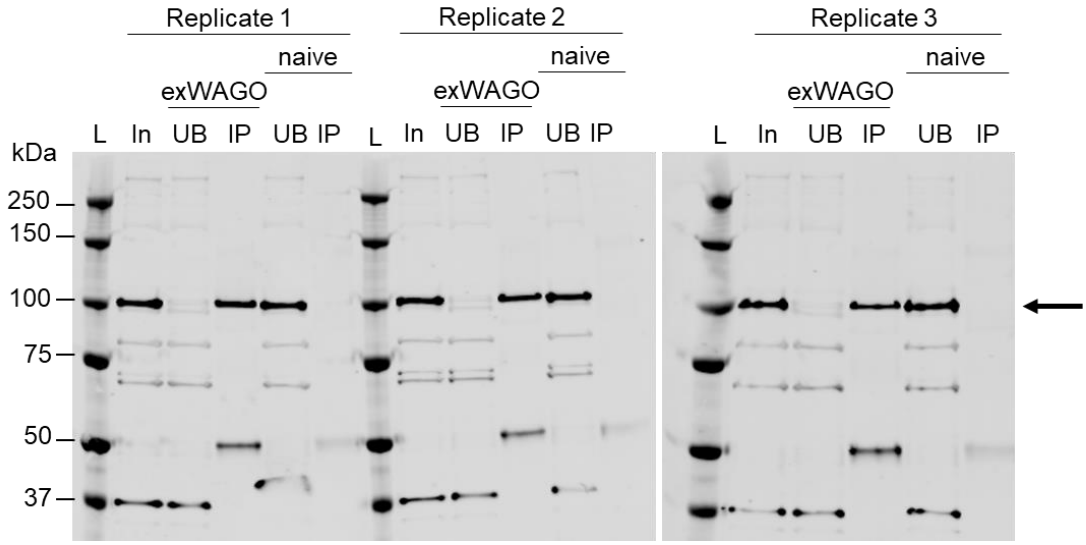


Figure 5.2 | Western blot analysis of exWAGO IPs from adult worm lysate using rat anti-exWAGO serum or naïve rat serum. Percentage of fraction analysed: 1% input (In); 1% unbound (UB); 4% eluate (IP). L = ladder. The blots were probed with 1:2,000 rabbit anti-exWAGO antibody, followed by 1:10,000 goat anti-rabbit IgG Dylight 800. The arrow indicates exWAGO (102 kDa).

Table 5.1 | Summary of significantly enriched proteins identified by label-free LC-MS/MS from anti-exWAGO or naïve serum IPs from adult worm lysate. Where no *H. bakeri* protein annotation was available, the *C. elegans* homologous gene(s) based on sequence similarity is/are reported here, following protein BLAST analysis (Basic Local Alignment Search Tool) on WormBase (Angeles-Albores *et al*, 2016, 2018). Only significant hits are reported with p-value < 0.05.

Accession	*C.elegans* homologous gene(s)	p-value	Fold change	Spectral counts					
				naïve IP			exWAGO IP		
HPOL_0001586401-mRNA-1	cey-2, cey-3	3.32E-08	Infinity	0	0	0	1	1	1
HPOL_0000298801-mRNA-1	exWAGO2 (lowly expressed)	8.33E-07	Infinity	0	0	0	0	1	0
HPOL_0000298601-mRNA-1	exWAGO	4.78E-06	31.16	0	0	1	92	125	106
HPOL_0002162301-mRNA-1	cbd-1	8.58E-05	14.35	0	0	0	8	12	15
HPOL_0000563401-mRNA-1	Y54G2A.36, cpg-3	0.000247	8.40	0	0	0	1	1	1
HPOL_0002115101-mRNA-1	cht-3	0.000263	7.29	0	0	0	0	1	0
HPOL_0000787001-mRNA-1	F48E3.4	0.000280	11.55	0	0	0	3	1	1
HPOL_0000207401-mRNA-1	pab-1	0.000297	40.96	0	0	0	1	1	1
HPOL_0002142001-mRNA-1	egg-1, egg-2	0.000502	128.18	0	0	0	1	1	2
HPOL_0000043801-mRNA-1	ipla-2	0.000547	2.88	0	0	2	1	0	0
HPOL_0000298901-mRNA-1	exWAGO-like	0.000955	6160.00	0	0	0	0	1	1
HPOL_0000013901-mRNA-1	ttr-46	0.003299	27.32	0	0	0	1	1	2
HPOL_0001122201-mRNA-1	cyc-2.1	0.003447	27.76	0	0	0	1	1	1
HPOL_0000232001-mRNA-1	gpd-3	0.023239	4.86	0	0	0	1	0	0
HPOL_0000100601-mRNA-1	perm-4	0.039878	46.26	0	0	0	0	1	0

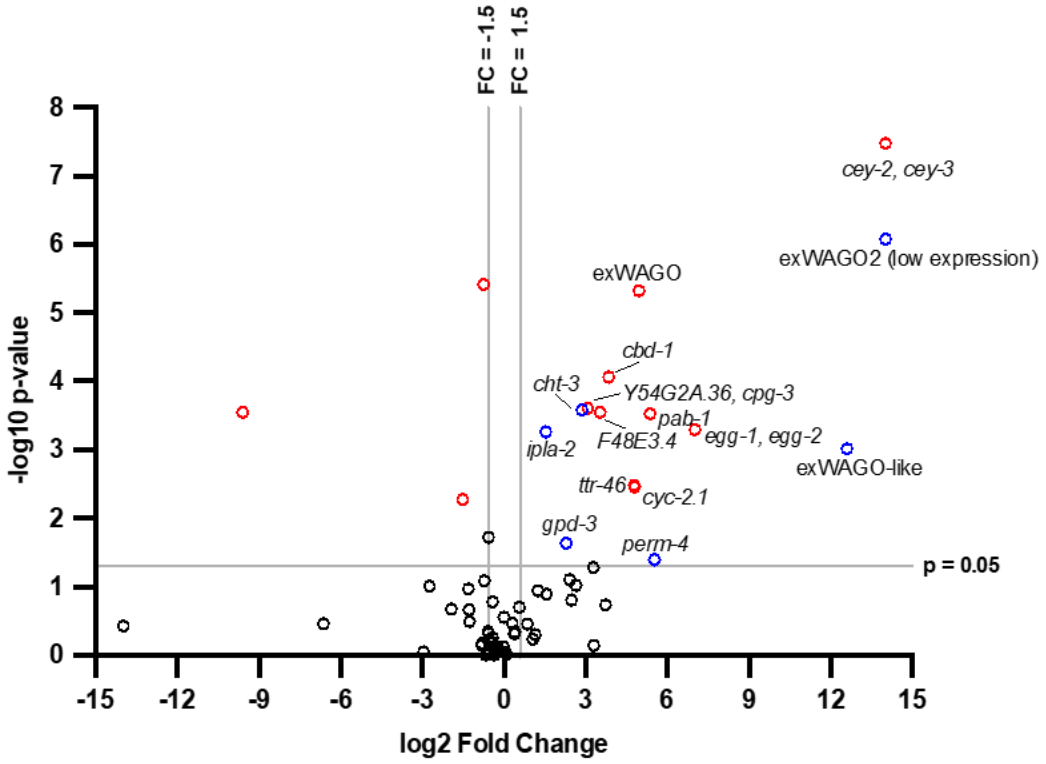


Figure 5.3 | Volcano plot of the proteins detected by label-free LC-MS/MS from anti-exWAGO or naïve serum IPs from adult worm lysate. Where no *H. bakeri* protein annotation was available, the *C. elegans* homologous gene(s) based on sequence similarity is reported here, following protein BLAST analysis (Basic Local Alignment Search Tool) on WormBase (Angeles-Albores *et al*, 2016, 2018). In red are significant hits identified in 3 out of 3 experimental samples with p-value ≤ 0.05 and -1.5>Fold Change>1.5. In blue are significant hits identified in 1 or 2 experimental samples with p-value ≤ 0.05 and -1.5>Fold Change>1.5. In black are non-significant hits with p-value > 0.05 or -1.5<Fold Change<1.5. Fold change (FC) represents the abundance of the proteins identified in the experimental samples compared to the control samples.

Gene Ontology enrichment analysis of the proteins found to be significantly enriched in the experimental samples was carried out using the WormBase interface (Fig. 5.4A, Table 5.2A) (Angeles-Albores *et al*, 2016, 2018). This analysis revealed over-representation of categories under the terms mRNA 3'-UTR binding and P-body (q-value cut off = 0.05), among others (Fig. 5.4A, Table 5.2A). P-bodies are ribonucleoprotein granules which contain transcriptionally repressed transcripts and proteins related to mRNA decay (Luo *et al*, 2018). Mammalian and nematode AGO proteins are reported to localise at P-bodies and other granules (Liu *et al*, 2005b; Meister *et al*, 2005; Sundby *et al*, 2021). The protein responsible for the identification of P-bodies and mRNA 3'-UTR binding terms as enriched is the *H. bakeri* protein with accession number HPOL_0000207401. This is a protein homologous to the *C. elegans* PAB-1. PAB-1 binds the poly-A tail of mRNAs and has been previously shown to interact with AGOs and GW182 proteins to induce translational repression (Kuzuoğlu-Öztürk *et al*, 2012; Huntzinger *et al*, 2013; Zhang *et al*, 2015). Additionally, *C. elegans* PAB-1 is known to be involved in transposon and transcriptional silencing in the germline and soma respectively (Vastenhouw *et al*, 2003; Grishok *et al*, 2005).

Individual examination of the proteins found to be enriched in the exWAGO IP compared to IP with naïve rat serum indicates that exWAGO interacts with the *H. bakeri* protein with accession number HPOL_0001586401. This is a homologue of the *C. elegans* proteins CEY-2 and CEY-3 which are homologues of the mammalian YBX-1 proteins (Arnold *et al*, 2014). YBX-1 is required for sorting miRNAs in extracellular vesicles as shown by a cell-free *in vitro* assay and for the secretion of miRNAs in mammalian EVs (Shurtleff *et al*, 2016). In *C. elegans*, CEY-2 interacts with the RDE-10 AGO protein (supplementary material of Zhang *et al*, (2012)) which is in turn involved in the regulation of gene silencing by promoting amplification of secondary siRNAs (Zhang *et al*, 2012). Moreover, in *C. elegans* CEY-2 interacts with PAB-1 and this interaction is RNA-dependent (Arnold *et al*, 2014). Identification of CEY-2/3 and PAB-1 as protein interaction partners of exWAGO inside adult worms shows that exWAGO interacts with proteins previously identified to interact with other known AGO proteins. This strongly suggests that exWAGO is implicated in gene regulation inside adult worms, as expected of AGO proteins according to the literature.

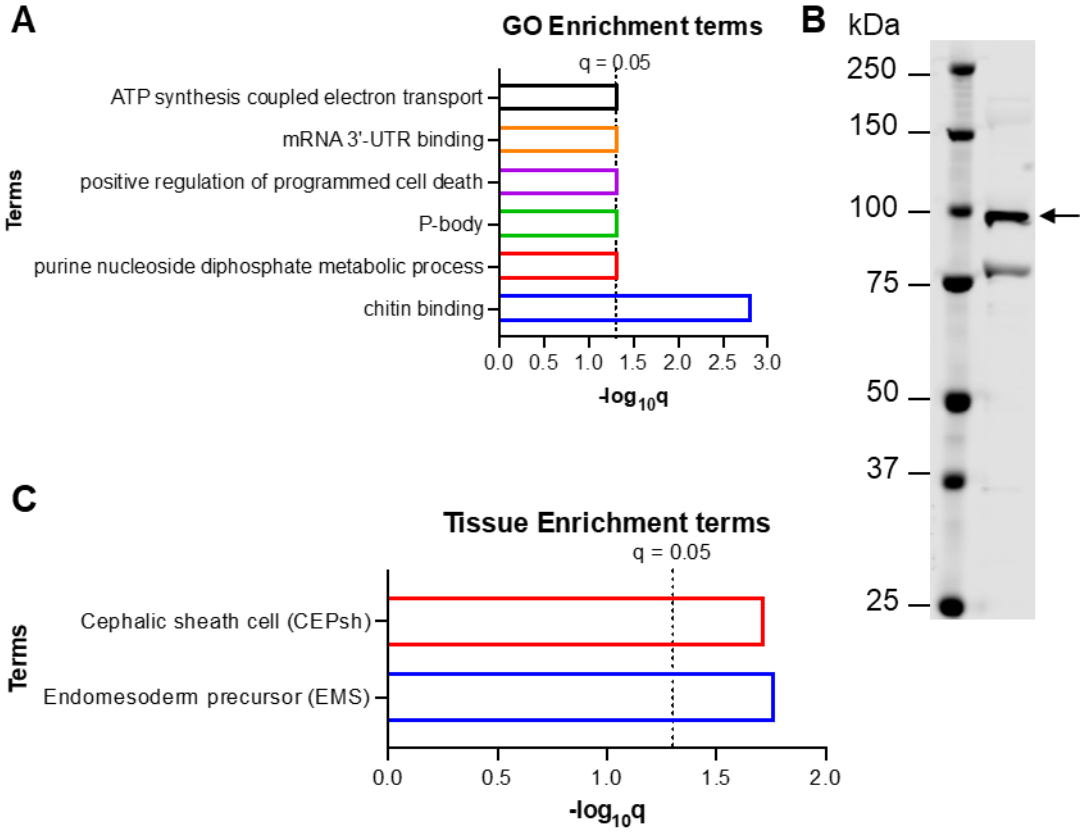


Figure 5.4 | Enrichment analyses of the proteins identified significantly enriched by label-free LC-MS/MS from anti-exWAGO or naïve serum IPs from adult worm lysate. (A) Gene Ontology enrichment analysis (q-value cut off = 0.05). (B) Western blot analysis of *H. bakeri* egg lysate (3 µg). The blot was probed with 1:2,000 rabbit anti-exWAGO antibody, followed by 1:10,000 goat anti-rabbit IgG Dylight 800. The arrow denotes exWAGO (102 kDa). (C) Tissue enrichment analysis (q-value cut off = 0.05). Enrichment analyses were executed on WormBase (Angeles-Albores *et al*, 2016, 2018).

Table 5.2 | The terms identified enriched following A) gene ontology (GO) and B) Tissue enrichment analysis of significant protein hits identified by label-free LC-MS/MS from anti-exWAGO or naïve serum IPs from adult worm lysate. Where no *H. bakeri* protein annotation was available, the *C. elegans* homologous gene(s) based on sequence similarity is/are reported here, following protein BLAST analysis (Basic Local Alignment Search Tool). GO, Tissue enrichment and BLAST analyses were performed using WormBase (Angeles-Albores *et al*, 2016, 2018). Only significant hits are reported with q-value ≤ 0.05.

A

Protein	WormBase ID	WormBase GO term
cyc-2.1	WBGene00017121	positive regulation of programmed cell death GO:0043068
pab-1	WBGene00003902	P-body GO:0000932
cbd-1	WBGene00010351	chitin binding GO:0008061
cht-3	WBGene00016084	chitin binding GO:0008061
pab-1	WBGene00003902	mRNA 3'-UTR binding GO:0003730
gpd-3	WBGene00001685	purine nucleoside diphosphate metabolic process GO:0009135
cyc-2.1	WBGene00017121	ATP synthesis coupled electron transport GO:0042773

B

Protein	WormBase ID	WormBase anatomy term
egg-1	WBGene00015083	EMS WBbt:0006876
cbd-1	WBGene00010351	EMS WBbt:0006876
cey-2	WBGene00000473	cephalic sheath cell WBbt:0008406
F48E3.4	WBGene00018605	cephalic sheath cell WBbt:0008406
ipla-2	WBGene00009801	cephalic sheath cell WBbt:0008406
perm-4	WBGene00016638	cephalic sheath cell WBbt:0008406
egg-1	WBGene00015083	cephalic sheath cell WBbt:0008406

Furthermore, mass spectrometry analysis identified *H. bakeri* proteins which are homologues of the *C. elegans* proteins EGG-1/2, CBD-1, PERM-4, CHT-3, and Y54G2A.36/CPG-3. These proteins have been identified as components of the *C. elegans* apical extracellular matrix of the eggshell encompassing the developing embryo, making up the vitelline layer, chitin layer and the chondroitin proteoglycan layer (Cohen & Sundaram, 2020). Although physical interactions of an AGO protein with eggshell-related proteins have not been described, the Barr lab reports the detection of eggshell-related proteins (EGG-1/2, PERM-4, CBD-1 and CPG-1/2), RNA metabolism proteins like CEY-2/3, and AGO proteins including the exWAGO orthologue PPW-1 in EVs secreted by *C. elegans* (Nikonorova *et al*, 2021). The

interaction of exWAGO with these proteins may indicate the germline as one of the sources of exWAGO. This is consistent with the fact that exWAGO is expressed in *H. bakeri* eggs using western blot analysis (Fig. 5.4B). It would be of interest to investigate in the future whether exWAGO can be released from *H. bakeri* eggs, although the current proteomics dataset available did not detect exWAGO in the egg ES products (Hewitson *et al*, 2013).

Tissue enrichment analysis using the WormBase interface (Angeles-Albores *et al*, 2016, 2018), identified enrichment of the Cephalic sheath cell (CEPsh) tissue (q-value cut off = 0.05) (Fig. 5.4C, Table 5.2B). Interestingly, CEPsh cells are glial cells which are important players of the nervous system. CEPsh cells are involved in various aspects of central synapses including the development and activity of synapses, and physically interact with the male-specific cephalic neuron CEP in *C. elegans* (Oikonomou & Shaham, 2011; Stout *et al*, 2014). Cilia on sensory neurons including cilia of the CEP neuron secrete EVs with the EV-marker PKD-2. EV preparations enriched with PKD-2 vesicles contain proteins that bind nucleic acids and are involved in RNA interference (Nikonorova *et al*, 2021). These findings suggest that exWAGO might be produced and secreted from the CEPsh cells or from cilia of cephalic neurons in *H. bakeri*. Furthermore, the tissue enrichment analysis also identified the Endomesoderm precursor (EMS) term (q-value cut off = 0.05) (Fig. 5.4C, Table 5.2B). EMS is one of the cells of the 4-cell stage *C. elegans* embryo. The EMS cell divides into the cell E which will later develop into the midgut (intestine) and the cell MS which will form part of the worms' muscles and the posterior half of the pharynx (Maduro, 2009). This suggests that exWAGO might be produced in the intestine, a finding that links to results from Buck *et al*, (2014) who reports detection of intestinal proteins in the EV population secreted by *H. bakeri* which also carry exWAGO.

All in all, identification of protein interactors of exWAGO inside adult worms has generated the hypothesis that exWAGO could be secreted from different locations, with the cephalic sheath and the intestine of adult worms and eggs (germline) being putative locations of exWAGO expression and secretion. The idea of exWAGO functioning and being secreted from different locations also links to the findings of the vesicular and non-vesicular exWAGO binding unique sRNAs and sRNAs of different abundance (Chapter 3). These observations could thus be explained by distinct exWAGO complexes functioning in different body compartments of the worm.

5.3 Identifying the protein interaction partners of exWAGO inside the host *in vitro*

To elucidate the function of exWAGO inside host cells and to build a model for how it could be involved in regulation of gene expression, I set out to find the mouse proteins that interact with exWAGO. Several different methods were employed to identify host proteins associated with exWAGO in host cells or tissues by LC-MS/MS.

5.3.1 IP with recombinant cell line identifies SFPQ as a putative interactor of exWAGO

Our lab has previously established a mouse intestinal epithelial cell line (MODE-K immortalised cell line) which stably expresses N-terminal 3XFLAG-tagged exWAGO (Fig. 5.5). MODE-K cells are enterocyte cells (Vidal *et al*, 1993), which are the most abundant class of epithelial cells that line the gut (Hooper, 2015). Epithelial cells are in contact with *H. bakeri* during infection, making the MODE-K cells a logical starting point for the identification of exWAGO:host protein interactions.

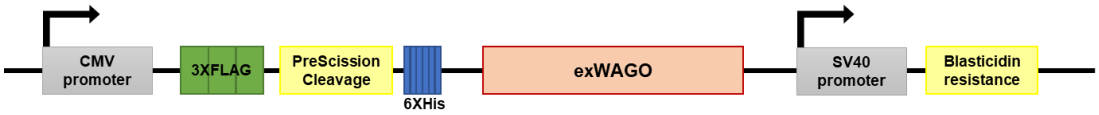


Figure 5.5 | Schematic representation of the plasmid region expressing recombinant exWAGO. exWAGO has a 3XFLAG-tag followed by a PreScission Cleavage site and a 6xHis-tag at the N-terminus. exWAGO-expressing MODE-K cells are selected using blasticidin. The plasmid region is not drawn to scale.

To optimise immunoprecipitation of exWAGO:protein complexes from MODE-K cells, different washing conditions were tested. exWAGO IPs from $6x10^7$ cells were performed using high salt washes (800 mM NaCl) with rat polyclonal anti-exWAGO antiserum (30 µl) conjugated on Protein G beads (50 µl) or using low or high stringency washes (150 mM or 800 mM of NaCl respectively) with anti-FLAG beads (50 µl) per IP. The protocol using low stringency wash buffer with FLAG-beads was provided by Dr. Rob van Nues (Elizabeth Bayne lab, The University of Edinburgh), and is referred to as the van Nues' protocol in this thesis. The samples were incubated with the antibody-conjugated beads overnight and one biological replicate was used per condition. The parental MODE-K cell line was used as a negative control to identify non-specific proteins pulled down in the IP. The protein complexes were eluted in denaturing buffer. Part of the eluate was used for quality control analysis (western blots) while 60% of the eluate was analysed using qualitative LC-MS/MS

analysis by Dr Dominic Kurian at the Proteomics and Metabolomics Facility at The Roslin Institute.

The results indicate that the exWAGO IPs successfully captured the exWAGO protein according to western blot analysis (Fig. 5.6A). Silver staining indicates that the profile of the proteins pulled down from the MODE-K cells expressing exWAGO was similar to that of the parental MODE-K (Fig. 5.6B), suggesting that not many unique protein interactors were pulled down with exWAGO. Moreover, the profile of the exWAGO IP using low salt washes was more complex compared to when high salt washes were performed, as expected (Fig. 5.6B).

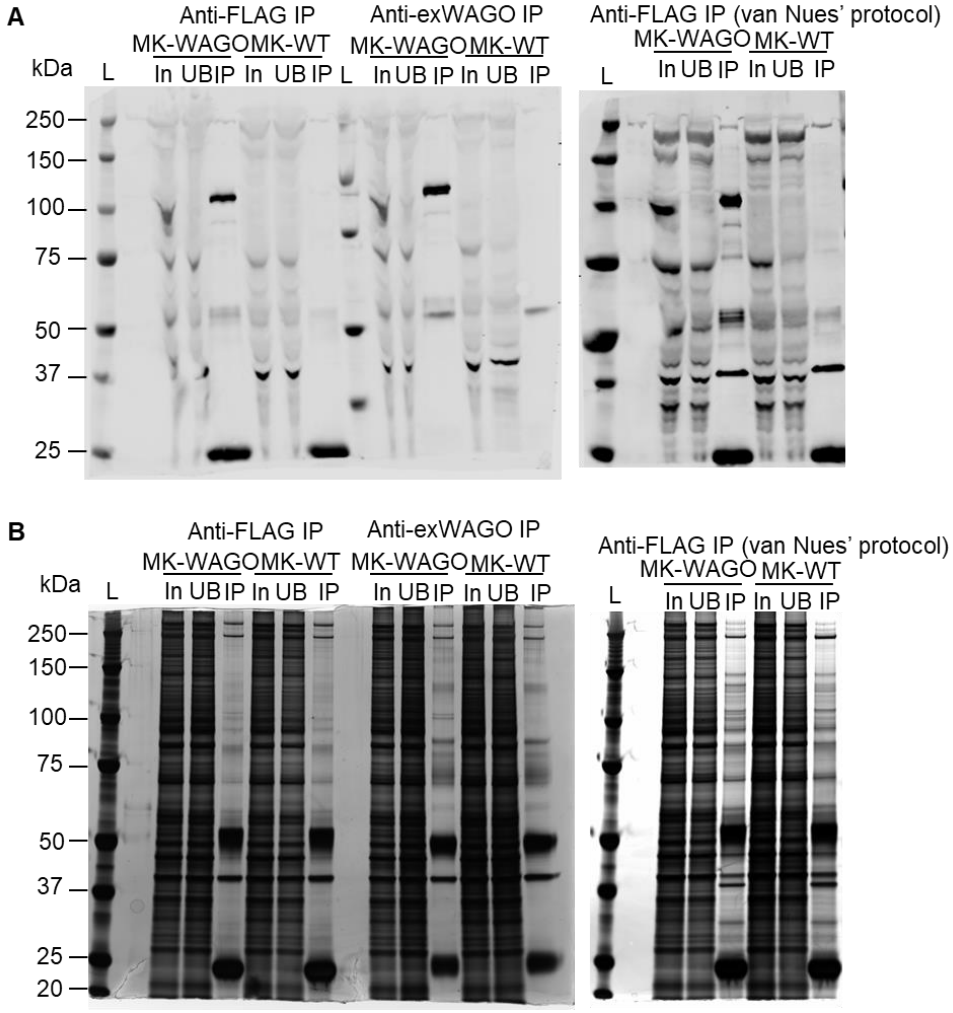


Figure 5.6 | Quality control analyses of exWAGO IPs from MODEK cells expressing FLAG-tagged exWAGO prior to proteomic analysis. (A) Western blot analyses where the blots were probed with 1:4,000 mouse anti-FLAG antibody, followed by 1:10,000 anti-mouse IgG AlexFluor 680. Percentage of fraction analysed: 1% input (In), 1% unbound (UB), 5% eluate (IP). L = ladder, MK-WAGO = exWAGO-expressing MODE-K cells, MK-WT = parental MODE-K cells. (B) Silver staining analyses. Percentage of fraction analysed: 0.0125% input (In), 0.0125% unbound (UB), 5% eluate (IP). L = ladder MK-WAGO = exWAGO-expressing MODE-K cells, MK-WT = parental MODE-K cells.

Mass Spectrometry analysis identified exWAGO in all the experimental IPs and not in the controls (Supp. Table 5.2). However, no other proteins were found to be common to all three experimental samples without also being identified in the corresponding negative controls (Supp. Table 5.2). Certain proteins were only pulled down under low stringency conditions, suggesting that they transiently interact with exWAGO or form weak interactions (Supp. Table 5.2). To capture weak exWAGO:host protein interactions we decided to only use low salt buffers in future mass spectrometry experiments.

To increase reliability of our results, I repeated the anti-FLAG exWAGO IPs from exWAGO-expressing MODE-K cells and parental cells ($6x10^7$ cells per IP, sample incubation on beads overnight) with low stringency washes as per van Nues' protocol (n = 3 per condition). Additionally, to reduce the amount of the capture antibody detected by LC-MS/MS due to the elution of the sample in denaturing buffer and avoid reduction of material due to protein purification from a gel, the exWAGO:protein complexes were left bound on the beads. 70% of the IP material was passed on to the Proteomics and Metabolomics Facility at The Roslin Institute which processed the samples using on-bead tryptic digestion, as described in Chapter 2, Section 2.16.2, and then performed qualitative LC-MS/MS analysis. The remaining 30% of the IP was eluted in denaturing buffer to perform western blot and silver stain quality control analyses. The results indicate that the IPs were successful however the profile of the IPs from exWAGO-expressing cells looks similar to that of the parental cells (Fig. 5.7). The proteins identified in the experimental samples are summarised in Table 5.3, and the full list of proteins detected is found in Supplementary Table 5.3. Only two proteins were identified in all three experimental samples, exWAGO and a protein with accession number A0A0A6YW67 which is homologous to ubiquitin (Table 5.3, Supp. Table 5.3).

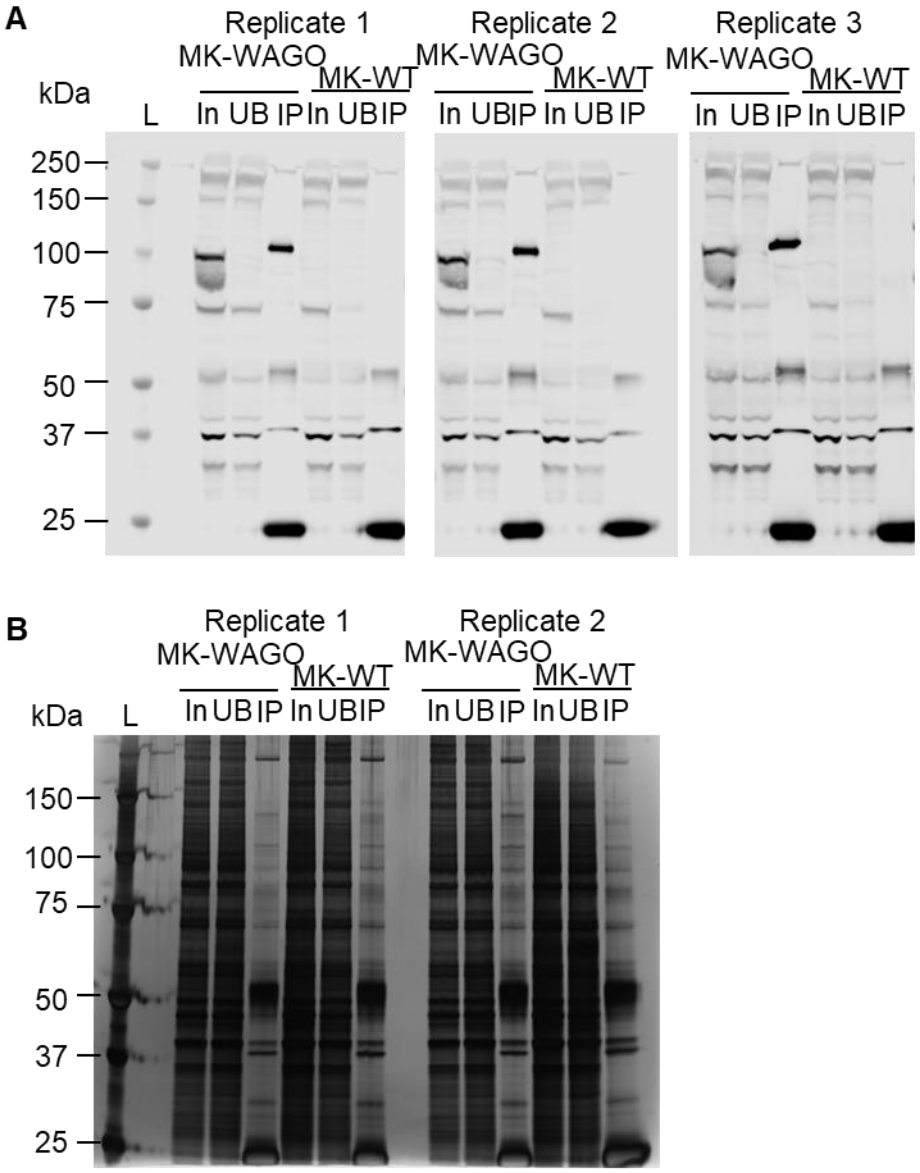


Figure 5.7 | Quality control analyses of FLAG IPs from MODEK cells expressing FLAG-tagged exWAGO samples prior to proteomic analysis. (A) Western blot analyses where the blots were probed with 1:4,000 mouse anti-FLAG antibody, followed by 1:10,000 anti-mouse IgG AlexFluor 680. Percentage of fraction analysed: 1% input (In), 1% unbound (UB), 4.5% eluate (IP). L = ladder, MK-WAGO = exWAGO-expressing MODE-K cells, MK-WT = parental MODE-K cells. (B) Silver staining analyses for two of the replicates. Percentage of fraction analysed: 0.002% input (In), 0.002% unbound (UB), 5% eluate (IP). L = ladder, MK-WAGO = exWAGO-expressing MODE-K cells, MK-WT = parental MODE-K cells.

Table 5.3 | The proteins identified from qualitative proteomic analysis following *in vitro* anti-FLAG IP under low salt condition. Only proteins identified in ≥ 2 experimental samples (WAGO1/2/3) only or were identified in all 3 experimental samples and in ≤1 control sample (WT-1/2/3) are shown here. Significance was determined as $p \leq 0.05$. WAGO = exWAGO-expressing MODE-K cells. WT = parental MODE-K cells. 1-3 = replicates.

Protein Accession Code	Gene name	Protein description	Count of significant peptides					
			WAGO-1	WAGO-2	WAGO-3	WT-1	WT-2	WT-3
HPOL_000029860	exWAGO	mRNA-1 peptide HPOL_0000298601-mRNA-1 protein_coding	1	103	103	0	0	0
A0A0A6YW67	Gm8797	MCG23377 isoform CRA_b	1	1	1	0	0	0
G3UYJ6	Gtf2i	General transcription factor II-I (Fragment)	0	4	2	0	0	0
Q9Z2V5	Hdac6	Histone deacetylase 6	0	25	19	0	0	0
J3QM80	Hnrnpf	Heterogeneous nuclear ribonucleoprotein F (Fragment)	0	1	1	0	0	0
Q8C2Q7	Hnrnph1	Heterogeneous nuclear ribonucleoprotein H	0	1	1	0	0	0
A6BLY7	Krt28	Keratin type I cytoskeletal 28	0	1	3	0	0	0
Q923A8	Map3k7	Map3k7 protein	0	8	6	0	0	0
A2AP92	Map3k7	Mitogen-activated protein kinase kinase kinase 7	0	4	3	0	0	0
P61514	Rpl37a	60S ribosomal protein L37a	0	1	1	0	0	0
Q80ZX8	Spag1	Sperm-associated antigen 1	0	1	1	0	0	0
D3Z216	Tab2	TGF-beta-activated kinase 1 and MAP3K7-binding protein 2 (Fragment)	0	2	5	0	0	0
Q571K4	Tab3	TGF-beta-activated kinase 1 and MAP3K7-binding protein 3	0	5	4	0	0	0
Q9ESX4	Zcchc17	Nucleolar protein of 40 kDa	0	1	1	0	0	0
Q9JJW6-2	Alyref2	Isoform 2 of Aly/REF export factor 2	5	9	6	0	0	9
V9GXQ2	Gm17087	Uncharacterized protein	1	2	1	0	1	0
O08638	Myh11	Myosin-11	52	54	49	52	0	0
Q8VIJ6	Sfpq	Splicing factor proline- and glutamine-rich	1	2	2	1	0	0
A0A087WRX8	Srrm2	Serine/arginine repetitive matrix protein 2 (Fragment)	1	6	3	0	2	0
D3Z6I8	Tpm3	Tropomyosin alpha-3 chain	13	5	6	19	0	0

To identify whether the proteins pulled down with the anti-FLAG exWAGO IP have physical and functional associations with other known AGO proteins, I performed a protein-protein association network analysis using the STRING database (Szklarczyk *et al*, 2019). Therefore, in this initial analysis I used the proteins mouse AGO1-4 and the proteins that were detected in more experimental samples than were detected in the negative controls after LC-MS/MS. The data reveal associations of mouse AGO proteins with the molecular chaperone heat-shock protein HSP90ab1 (HSP90beta), ribosomal proteins RPS3, RPS18, and RPL11, the LSM14a protein essential for P-body formation, and the splicing factor SFPQ (Fig. 5.8, Supp. Table 5.3).

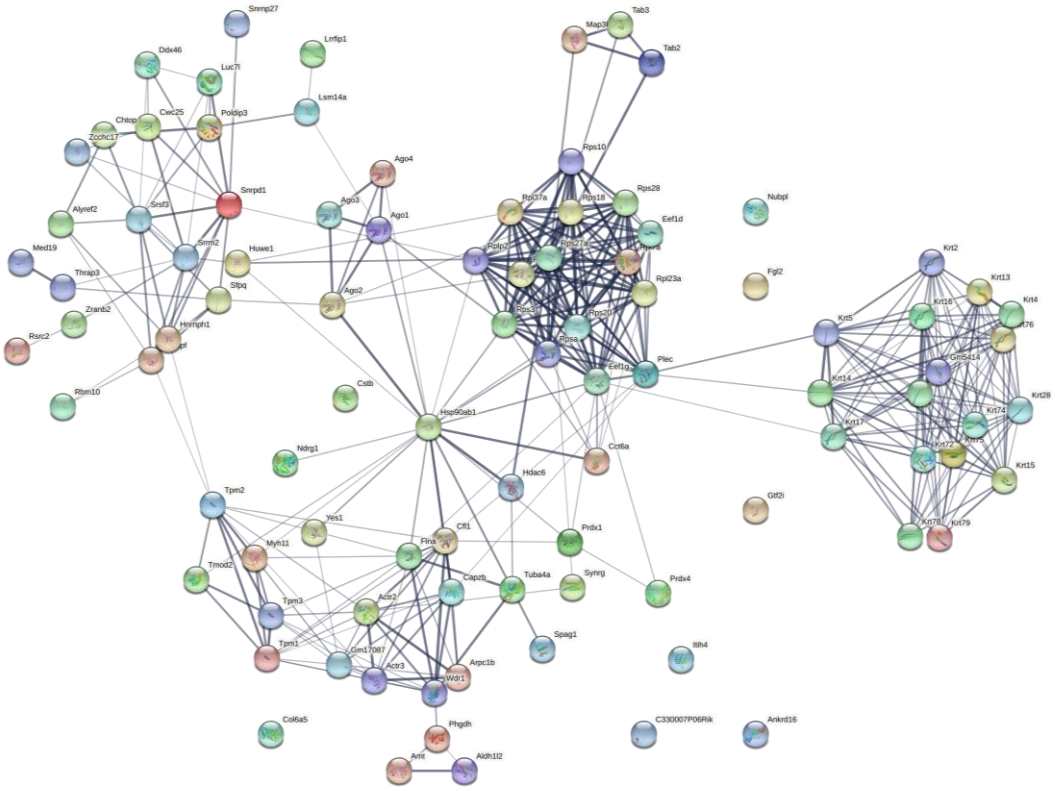


Figure 5.8 | Protein-protein association analysis of proteins identified as putative exWAGO interactors in mass spectrometry analysis after FLAG IPs under low salt conditions from FLAG-tagged exWAGO-expressing MODEK cells. Analysis was performed on STRING v11.5 (Szklarczyk *et al*, 2019), using a medium confidence interaction score. The thickness of the line connections indicates the strength of the data support and both functional and physical protein associations are shown in this network. Only proteins that were identified in more experimental replicates compared to the negative controls were used in this analysis. The mouse AGO1-4 proteins were not detected in the mass spectrometry data but were 'spiked' in the STRING analysis to identify known/predicted interactions of the proteins identified in the exWAGO IPs with mammalian AGOs.

Literature search shows that indeed HSP90ab1 physically interacts with mouse AGO2 (Frohn *et al*, 2012). This protein is localised in the cytoplasm and has been identified to capture and stabilise the conformation of empty AGO2 in *Drosophila* after it is primed 'open' by the chaperone HSP70, so that AGO2 can bind the siRNA duplex (Miyoshi *et al*, 2010; Tsuboyama *et al*, 2018). Although identification of HSP90ab1 as a protein interactor of exWAGO could be valid, it seems unlikely that this interaction would be based on loading sRNAs inside the host *in vivo*. This is because our findings from Chapter 3 suggest that exWAGO is secreted already bound to specific sRNAs, and hence we hypothesise that loaded exWAGO would be internalised by host cells whilst still bound to the parasite 5'PPP secondary siRNAs.

Furthermore, identification of the protein LSM14a is intriguing. Along with other proteins, LSM14a is involved in the formation of the cytoplasmic bodies known as P-bodies (Ayache *et al*, 2015). AGO proteins have been reported by several groups to localise at P-bodies (Liu *et al*, 2005b; Sen & Blau, 2005). P-bodies are implicated in mRNA processing including mRNA degradation and gene silencing (Luo *et al*, 2018). Additionally, following proximity-dependent biotin identification analysis of human cells, Youn *et al*, (2018) report identification of the LSM14a protein as a protein interactor of the CCR4-NOT deadenylase complex, which also interacts with the AGO-RISC complex. These findings suggest that exWAGO is localised in the cytoplasm of host cells and associates with proteins that would typically interact with AGO proteins. Additionally, these findings generate the hypothesis that exWAGO might operate in special compartments of the host cell such as P-bodies and may be involved in gene silencing of targets at the post-transcriptional level through removal of the poly(A) tail of mRNA target via the CCR4-NOT deadenylase complex.

Furthermore, the ribosomal proteins RPS3 and RPS18 detected to interact with exWAGO have previously been reported to interact with mAGO2 (Frohn *et al*, 2012). However, ribosomal proteins identified in proteomics analysis are usually considered contaminants due to the high abundance of ribosomes. The ribosomal proteins RPS3 and RPS18 have been reported to be some of the most frequent proteins pulled down non-specifically and detected by mass spectrometry following affinity purification (Mellacheruvu *et al*, 2013). This suggests that further validation experiments are necessary to confirm interactions of ribosomal proteins with the protein of interest. Nevertheless, Singh *et al* (2021) reports that the *C. elegans* AGO protein CSR-1 interacts with ribosomal proteins directly in the germline and that this interaction is

RNA-independent. The authors show that CSR-1 is involved in the biogenesis of secondary siRNAs generated in phase with translating ribosomes in the cytosol (Singh *et al*, 2021). Functional assays should be employed to understand if the interaction of exWAGO with mammalian ribosomal proteins is true rather than an artifact of the high abundance of ribosomal proteins.

The putative interaction between exWAGO and SFPQ (Splicing Factor Proline/Glutamine rich) is also intriguing for several reasons. Bottini *et al* (2017), report that SFPQ interacts with the mouse and human AGO2 in an RNA-dependent manner and binds long 3' UTRs of mRNAs (Bottini *et al*, 2017). SFPQ then aggregates onto the long 3'UTRs and alters the configuration of the target mRNA and accessibility of the sRNA binding sites, optimising positioning of guide miRNAs, leading to post-transcriptional gene silencing (Bottini *et al*, 2017). Additionally, knocking out SFPQ in mouse and human cells rescues its targets indicating that certain mRNAs are silenced in an SFPQ-dependent manner (Bottini *et al*, 2017), thus implicating SFPQ in the mechanism of gene silencing. Additionally, SFPQ is a nuclear protein (Yarosh *et al*, 2015) and has been identified to interact with the GW182 protein, a well-known protein interactor of AGO proteins, following anti-GW182 IPs from nuclear extracts of human cells (Hicks *et al*, 2017). These findings suggest that exWAGO might also be localised in the nucleus. SFPQ also interacts with the proteins Paraspeckle 1 (Pspc1) and Non-POU domain containing octamer-binding protein (NoNO) (Yarosh *et al*, 2015). Although these proteins have not been identified in this particular dataset, NoNO and SFPQ had previously been identified in the preliminary LC-MS/MS experiment to interact with exWAGO under low salt conditions only (Supp. Table 5.2). Hence, I set out to validate SFPQ as a putative interactor of exWAGO.

5.3.2 Analysis of SFPQ as a putative protein interactor of exWAGO

Western blot analysis using material from the IPs reported above with high salt washes or low salt washes, shows that SFPQ interacts with exWAGO following anti-FLAG IP using low stringency conditions only (Fig. 5.9A). The interaction is lost under high salt conditions (Fig. 5.9A). To further validate this interaction using a reverse IP method, I immunopurified SFPQ using anti-SFPQ antibody (5 µg) or exWAGO using the rat anti-exWAGO polyclonal antibody (6 µl) conjugated on Protein G beads (50 µl) from approximately $8x10^6$ cells per IP. The cell lysate was incubated with the antibody-conjugated beads overnight and the IP was performed under low salt washes only. Western blot analysis indicates that the IPs were successful (Fig.

5.9B). However, exWAGO is not detected in the SFPQ IP (Fig. 5.9B). This could be an issue of the low amount of material used in these IPs which is almost 10 times less than the number of cells used for proteomic analysis, hence these experiments should be repeated with greater numbers of cells in the future.

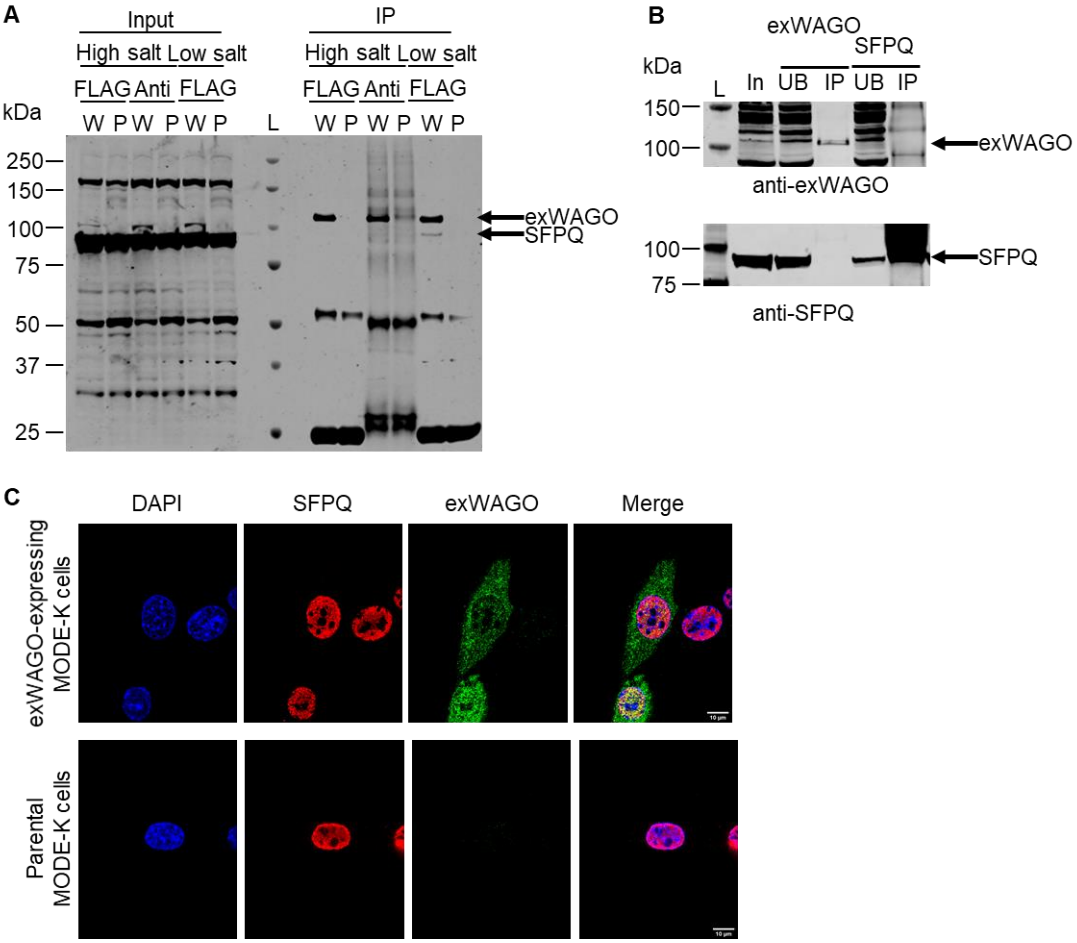


Figure 5.9 | Validation of SFPQ as a putative interactor of exWAGO in MODEK cells expressing FLAG-tagged exWAGO. (A) Western blot analysis of immunopurified exWAGO under different conditions shows that SFPQ (95 kDa) is pulled down only under low salt conditions. The blot was probed with 1:1,500 rabbit anti-SFPQ antibody followed by 1:10,000 goat anti-rabbit IgG Dylight 800 and re-probed with 1:4,000 mouse anti-FLAG antibody followed by 1:10,000 goat anti-mouse IgG AlexFluor 680. FLAG = IP utilising FLAG-beads, Anti = IP using anti-exWAGO rat serum antibody conjugated on Protein G beads, W = exWAGO-expressing MODE-K cells, P = parental MODE-K cells. Loaded 5% of the eluate fraction (IP) and 0.3% of the input. The arrows indicate FLAG-tagged exWAGO and SFPQ (B) Western blot analysis of exWAGO IP and SFPQ IP. The top blot was probed with 1:2,000 anti-exWAGO rabbit antibody followed by 1:10,000 goat anti-rabbit IgG Dylight 800 and the bottom blot was probed with 1:1,500 rabbit anti-SFPQ antibody followed by 1:10,000 goat anti-rabbit IgG Dylight 800. Percentage of fraction analysed: 2% input (In), 2% unbound (UB), 16.7% eluate (IP). (C) Immunofluorescence of exWAGO-expressing and parental MODE-K cells. Cells were stained with the nuclear stain DAPI, with 1:100 rabbit anti-SFPQ followed by 1:2,000 goat anti-rabbit IgG AlexaFluor 546 and 1:50 rat epoxy cleaned anti-exWAGO antibody followed by goat anti-rat IgG AlexFluor 647. Scale bar = 10 μm.

To understand whether exWAGO and SFPQ co-localise I performed immunofluorescence on the MODE-K cells expressing exWAGO and used the parental cell line as a negative control. Immunofluorescence data show that SFPQ is a nucleoplasmic protein (Fig. 5.9C), consistent with published literature (Yarosh *et al*, 2015). The exWAGO signal detected suggests that exWAGO is nuclear but also cytoplasmic (Fig. 5.9C). Nuclear fractionation experiments should provide additional evidence regarding the localisation of exWAGO.

Furthermore, immunofluorescence experiments allow us to examine exWAGO expression at the single cell level compared to western blotting which allows assessment of the global cell population. This led to the discovery that rather few cells of the recombinant FLAG-tagged exWAGO MODE-K cell line express exWAGO (Fig. 5.9C and data not shown). We hypothesise that the transgene is largely silenced since it is controlled by the CVM promoter, known to be silenced in mammalian cells (Hsu *et al*, 2010). The antibiotic resistance cassette blasticidin, used to select for cells expressing exWAGO, is under a different promoter (Simian Virus 40 promoter; SV40) suggesting that blasticidin resistant cells are selected regardless of whether exWAGO is expressed. This finding may explain why despite the use of tens of millions of cells in the IP that was then used in the mass spectrometry analysis we have not been able to identify highly reproducible protein interactors. It is recommended that a new recombinant cell line stably expressing exWAGO is generated for use in future experiments. exWAGO should be under the control of a promoter that does not get methylated in mammalian cells, such as the eukaryotic translation elongation factor 1 alpha (EF-1α) (Teschendorf *et al*, 2002).

5.4 Identifying the protein interaction partners of exWAGO inside the host *in vivo*

Working with mammalian cell lines in the above work provides an infinite source of material, however there are several limitations. In particular, the environment where exWAGO functions is more complex than what we mimic *in vitro*, as *H. bakeri* resides in the gastrointestinal tract which is composed of at least 7 different epithelial cell types that function in a consorted manner along with immune cell populations that infiltrate the gut to mount an immune response.

5.4.1 *In vivo* gut IPs identify Ubap2l and Dab2 as putative interactors of exWAGO

To identify the protein interactors of exWAGO in a more physiologically relevant context, I employed the *in vivo* exWAGO IP that I developed (described in Chapter 4, Section 4.1). exWAGO was immunoprecipitated from infected gut tissue and uninfected tissue was used as a negative control. IPs were carried out with three biological replicates per condition using the rat polyclonal anti-exWAGO antibody (25 µl) conjugated on Protein G beads (75 µl) per IP. The gut lysate was incubated with the antibody-conjugated beads (45 min) and only low salt washes were performed to capture weaker protein interactions. As previously, part of the IP was eluted in denaturing buffer (24%) to perform western blot and silver stain quality control analyses. The results indicate that the IPs were successful, and the IP profiles from infected and uninfected tissues have some noticeable differences (Fig. 5.10). The rest of the IP material (76%) was passed on to the Proteomics and Metabolomics Facility at The Roslin Institute which processed the samples using on-bead tryptic digestion, as described in Chapter 2, Section 2.16.2, and then performed qualitative LC-MS/MS.

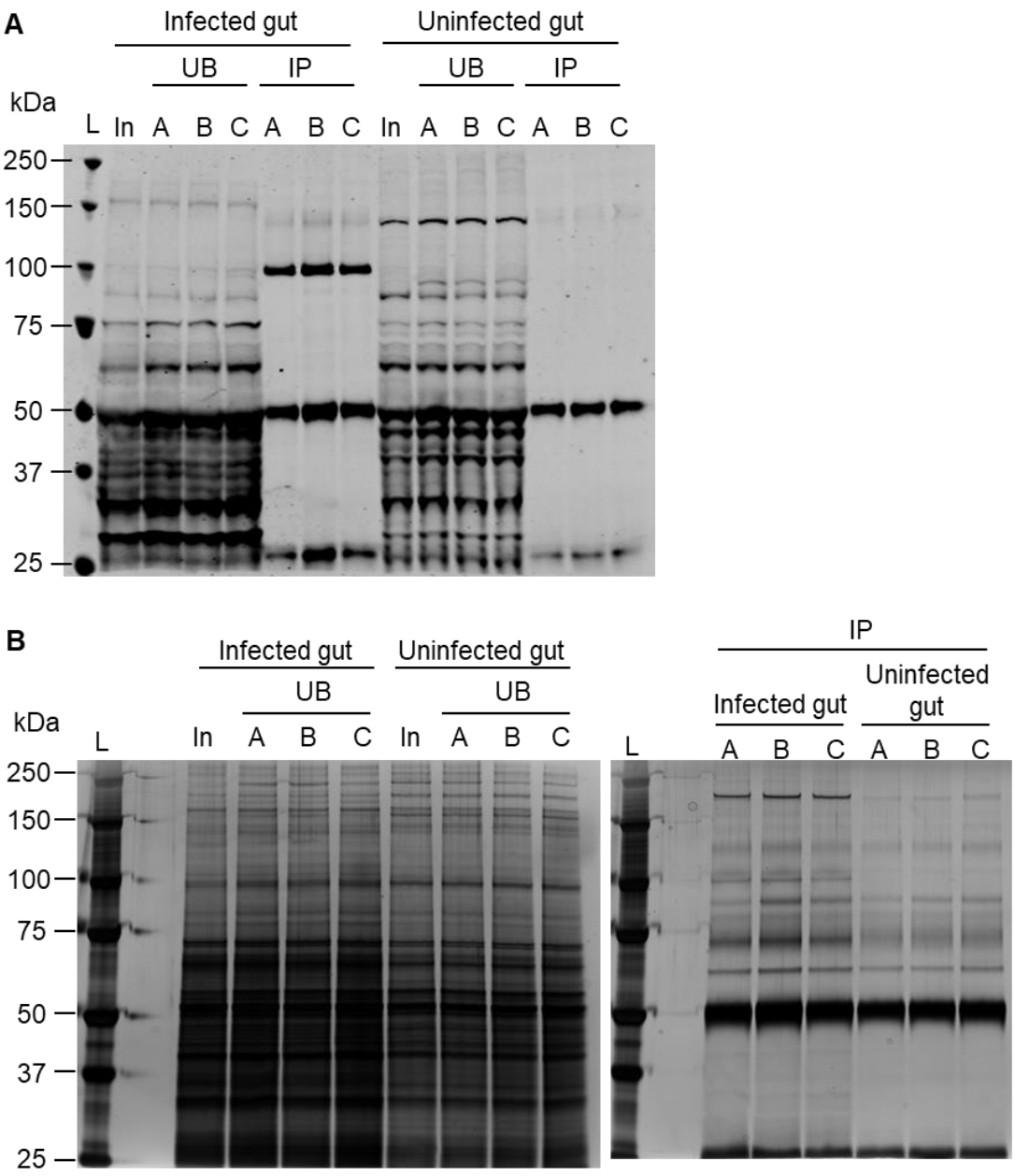


Figure 5.10 | Quality control analyses of immunopurified exWAGO from *in vivo* samples prior to proteomic analysis. (A) Western blot analysis of exWAGO IPs from *H. bakeri*-infected and uninfected gut tissue. The blot was probed with 1:4,000 rabbit anti-exWAGO antibody followed by 1:10,000 goat anti-rabbit Dylight 800. Percentage of fraction analysed: 1.5% input (In), 1.5% unbound (UB), 4% eluate (IP). A-C denote replicates. L = ladder. (B) Silver stain analysis of the samples analysed by western blot in (A). Percentage of fraction analysed: 0.019% input (In) for infected gut and 0.038% for uninfected gut, 0.019% unbound (UB) for infected gut and 0.038% for uninfected gut, 3.3% eluate (IP). A-C denote replicates. L = ladder.

Qualitative LC-MS/MS analysis identified 93 proteins unique to the experimental samples including exWAGO (Supp. Table 5.4). One of these proteins is the DEAD-box RNA helicase protein, Ddx41. DEAD-box proteins have previously been reported to interact with mammalian AGOs (Peters & Meister, 2007) and interestingly, Ddx41 is involved in the translational repression of the cyclin-dependent kinase inhibitor p21 transcript (Peters *et al*, 2017). Protein association network analysis using STRING (Szklarczyk *et al*, 2019) identified that most of the proteins detected in this dataset are ribosomal proteins (Fig. 5.11). Other proteins involved in cell migration regulation and response to wounding and wound healing were also identified. Some of these proteins, including the proteins Clca1 (Calcium-activated chloride channel regulator), Muc2 (Mucin-2), Tff2 (Trefoil factor 2), Tff3 (Trefoil factor 2) and Fcgbp (Fc fragment of IgG binding protein), are upregulated during the Type 2 immune response and contribute to the mucosal barrier (Sharpe *et al*, 2018). These proteins are also major products of goblet cells, which expand during worm infections leading to increased mucus secretions (Kim & Ho, 2010; Nyström *et al*, 2018; Sharpe *et al*, 2018). The proteins identified here are most likely highly abundant proteins induced during infection with *H. bakeri*. Unfortunately, therefore, the negative control used in this experiment (i.e. uninfected mouse gut immunoprecipitated with anti-exWAGO antiserum) will not have the same background of proteins that could be pulled down non-specifically. Hence, to identify true exWAGO protein interactors *in vivo* we require further controls such as IPs from infected mouse gut immunoprecipitated with naïve rat antiserum.

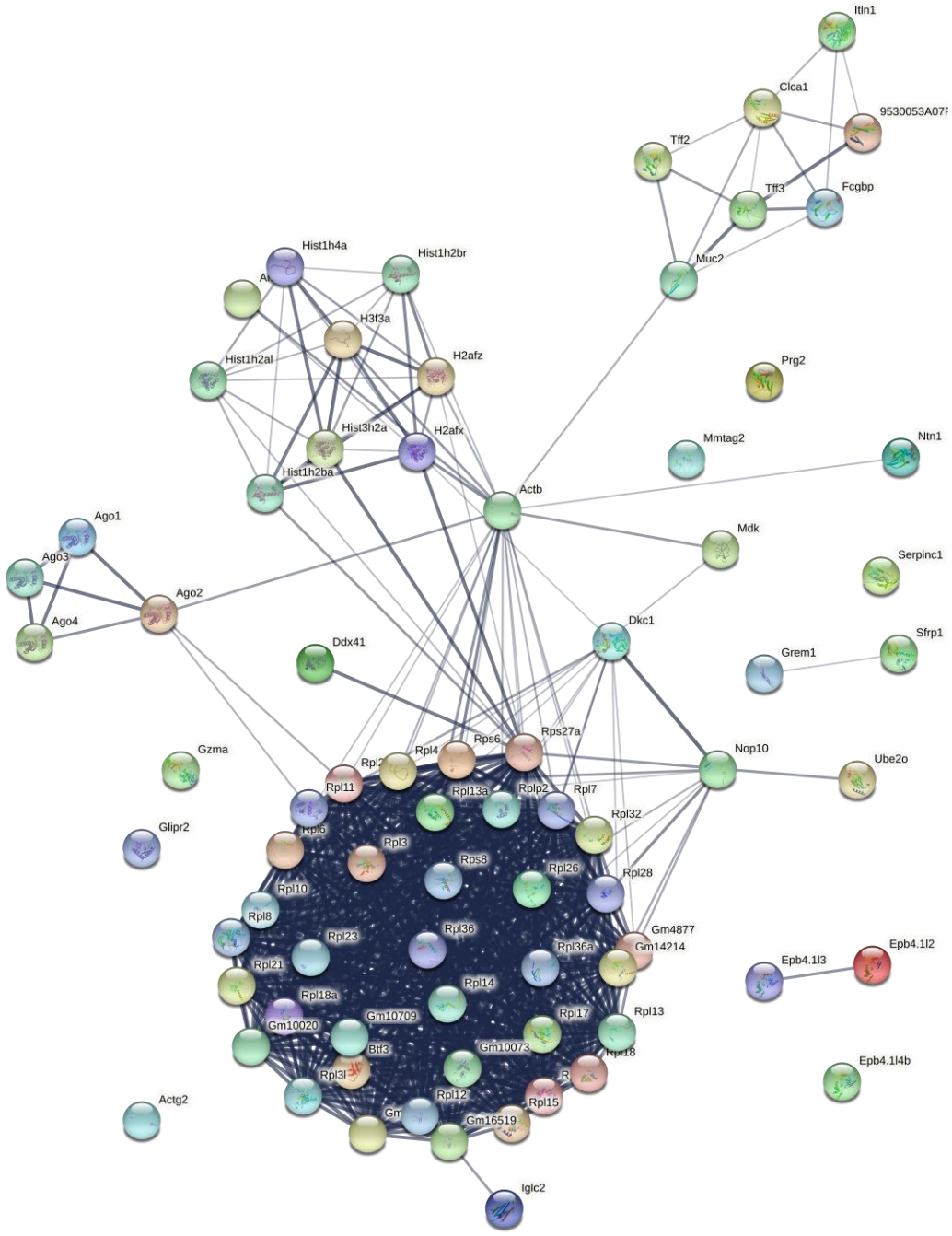


Figure 5.11 | Protein-protein association analysis of proteins identified as putative exWAGO interactors in mass spectrometry analysis of *in vivo* gut samples. Analysis was performed on STRING v11.5 (Szklarczyk *et al*, 2019), using a medium confidence interaction score. The thickness of the line connections indicates the strength of the data support and both functional and physical protein associations are shown in this network. Only proteins that were exclusively identified in all experimental replicates were used here. The mouse AGO1-4 proteins were 'spiked' in the analysis to identify known/predicted interactions of the proteins identified in the exWAGO IPs with mammalian AGOs.

To account for the upregulation of proteins induced due to the infection, I immunopurified exWAGO from *H. bakeri*-infected mouse gut using the rat anti-exWAGO polyclonal antibody or naïve rat serum as a negative control and performed proteomic analysis. For these IPs I used the aforementioned antibodies (25 µl) conjugated to Protein G beads (75 µl) per IP. The gut lysate was incubated with the antibody-conjugated beads (45 min) and only low salt washes were performed to capture weaker protein interactions. Only one biological sample was generated in this experiment. Western blot and silver stain quality control analysis show that the IPs were successful (Fig. 5.12). 85% of the IPs was used for quantitative LC-MS/MS. The protein peptides were labelled using stable isotope dimethyl labelling to allow data quantitation (Boersema *et al*, 2009), by Dr. Dominic Kurian at The Roslin Institute as described in Chapter 2, Section 2.16.3.

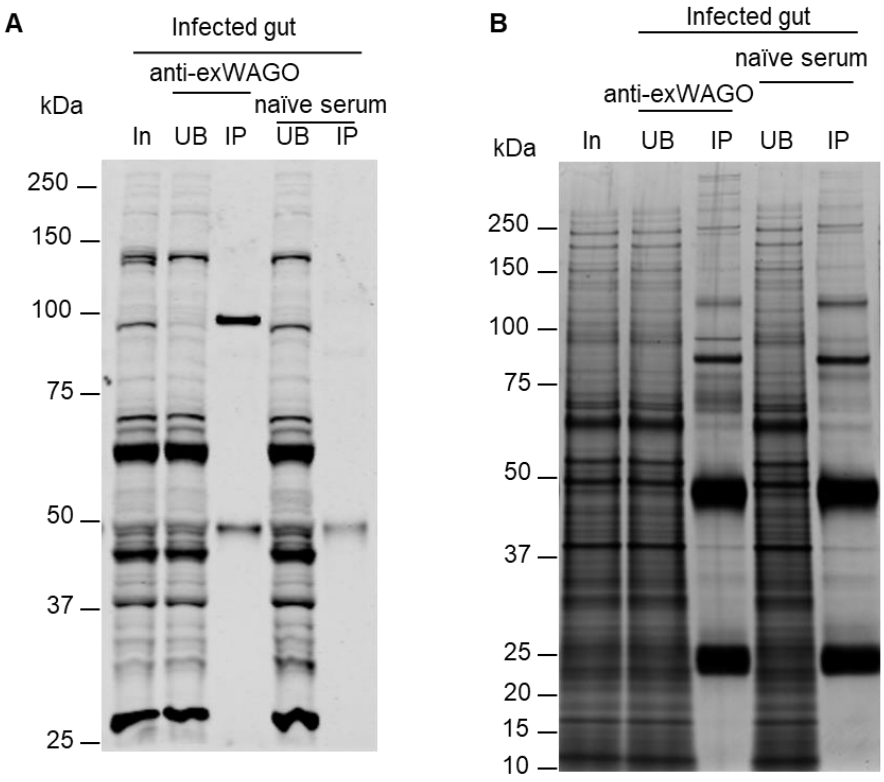


Figure 5.12 | Quality control analyses of immunopurified exWAGO from *in vivo* samples prior to proteomic analysis. (A) Western blot analysis of exWAGO or naïve serum IPs from *H. bakeri*-infected tissue. The blot was probed with 1:4,000 rabbit anti-exWAGO antibody followed by 1:10,000 goat anti-rabbit IgG Dylight 800. Percentage of fraction analysed: 1% input (In), 1% unbound (UB), 5% eluate (IP). (B) Silver stain analysis of the samples tested by western blot in (A). Percentage of fraction analysed: 0.025% input (In), 0.025% unbound (UB), 5% eluate (IP).

Proteomic analysis shows that exWAGO is successfully detected to be enriched in the experimental samples compared to the negative control (Table 5.4). The other proteins detected enriched in the exWAGO IP correspond to the mouse proteins Ubap2l (Ubiquitin-associated protein 2-like) and Dab2 (Disabled homolog 2) (Table 5.4). Ubap2l is involved in the formation of stress granules although is not implicated in translational inhibition induced by stress (Cirillo *et al*, 2020). Stress granules are cytosolic RNA-protein compartments in cells, thought to be involved in gene regulation, where AGO proteins can localise (Leung *et al*, 2006; Buchan & Parker, 2009). Stress granules form in response to acute stress conditions and during infection (Campos-Melo *et al*, 2021). However, there is no published literature examining the implication of stress granules during nematode infection specifically. Interestingly, Pare *et al* (2009) show that the localisation of the human AGO2 (hAGO2) is regulated by the heat shock protein chaperone HSP-90. Inhibition of HSP-90 leads to reduction in hAGO2 recruitment to stress granules and was also associated with decreased small RNA-dependent suppression of genes (Pare *et al*, 2009). This leads to the hypothesis that exWAGO might interact with a heat shock protein, as identified from the proteomic experiment from exWAGO-expressing MODE-K (Supp. Table 5.3), to localise to cellular compartments like stress granules. Moreover, identification of Ubap2l as a host protein interactor of exWAGO suggests that exWAGO is internalised by host cells.

But how does exWAGO enter host cells to begin with? There are various proposed routes regarding vesicle uptake by host cells (Drurey *et al*, 2020), and several groups show that *H. bakeri* and other helminth EVs can be internalised by host cells *in vitro* (Marcilla *et al*, 2012; Buck *et al*, 2014; Coakley *et al*, 2017; Eichenberger *et al*, 2018b; De La Torre-Escudero *et al*, 2019) Thus, the vesicular exWAGO could be taken up while encapsulated in EVs. Research is, however, lacking at studying how helminth non-vesicular proteins are internalised by host cells. Hence the question remains; how does the non-vesicular exWAGO enter host cells? The other protein identified in this experiment, Dab2, could hold the key to the answer.

Dab2 is an adaptor protein of the clathrin-mediated endocytosis pathway facilitated through interaction of the phosphotyrosine-binding (PTB) domain of the Dab2 protein with the NxxY motif (where x is any amino acid) of the internalised protein (Finkielstein & Capelluto, 2016). Dab2 is involved in the regulation of immune cells; it is highly expressed in M2 alternatively activated macrophages and it plays a role in

macrophage polarisation and activation (Figliuolo da Paz *et al*, 2020). To understand if Dab2 and Ubap2l are true protein interactors of exWAGO, the mass spectrometry findings need to be validated.

Table 5.4 | Proteins identified after *in vivo* exWAGO gut IP from quantitative proteomic analysis.

Accession	-10lgP	Fold Change (Experimental /Control)	Coverage (%)	No. Peptides	No. of unique peptides	Description
sp|Q80X50|UBP2L_MOUSE	56.62	41.91	4	1	1	Ubiquitin-associated protein 2-like OS=Mus musculus OX=10090 GN=Ubap2l PE=1 SV=1
sp|Q80X50-4|UBP2L_MOUSE	56.62	41.91	4	1	1	Isoform 4 of Ubiquitin-associated protein 2-like OS=Mus musculus OX=10090 GN=Ubap2l
sp|Q80X50-2|UBP2L_MOUSE	56.62	41.91	4	1	1	Isoform 2 of Ubiquitin-associated protein 2-like OS=Mus musculus OX=10090 GN=Ubap2l
sp|Q80X50-5|UBP2L_MOUSE	56.62	41.91	4	1	1	Isoform 5 of Ubiquitin-associated protein 2-like OS=Mus musculus OX=10090 GN=Ubap2l
sp|Q80X50-3|UBP2L_MOUSE	56.62	41.91	4	1	1	Isoform 3 of Ubiquitin-associated protein 2-like OS=Mus musculus OX=10090 GN=Ubap2l
A0A0G2JDV6|A0A0G2JDV6_MOUSE	56.62	41.91	4	1	1	Ubiquitin-associated protein 2-like OS=Mus musculus OX=10090 GN=Ubap2l PE=1 SV=1
A0A0H2UH17|A0A0H2UH17_MOUSE	56.62	41.91	4	1	1	Ubiquitin-associated protein 2-like OS=Mus musculus OX=10090 GN=Ubap2l PE=1 SV=1
HPOL_0000298601-mRNA-1_exWAGO	125.85	11.8	2	1	1	exWAGO
F6TQN9|F6TQN9_MOUSE	39.42	9.99	6	1	1	Disabled homolog 2 (Fragment) OS=Mus musculus OX=10090 GN=Dab2 PE=1 SV=1
sp|P01789|HVM20_MOUSE	172.48	3.42	11	2	2	Ig heavy chain V region M603 OS=Mus musculus OX=10090 PE=1 SV=1
A0A075B5M3|A0A075B5M3_MOUSE	111.64	3.42	19	1	1	Immunoglobulin kappa variable 4-58 OS=Mus musculus OX=10090 GN=Igkv4-58 PE=1 SV=1
sp|P01867|IGG2B_MOUSE	189.96	2.5	4	2	1	Ig gamma-2B chain C region OS=Mus musculus OX=10090 GN=Igh-3 PE=1 SV=3
sp|P01867-2|IGG2B_MOUSE	189.96	2.5	5	2	1	Isoform 2 of Ig gamma-2B chain C region OS=Mus musculus OX=10090 GN=Igh-3
A0A0A6YVP0|A0A0A6YVP0_MOUSE	225.68	2.2	4	3	1	Immunoglobulin heavy constant gamma 2B (Fragment) OS=Mus musculus OX=10090 GN=Ighg2b PE=1 SV=2
A0A075B5P3|A0A075B5P3_MOUSE	225.68	2.2	5	3	1	Immunoglobulin heavy constant gamma 2B (Fragment) OS=Mus musculus OX=10090 GN=Ighg2b PE=1 SV=1
A0A0B4J1H7|A0A0B4J1H7_MOUSE	33.35	1.9	11	1	1	Immunoglobulin kappa variable 1-135 (Fragment) OS=Mus musculus OX=10090 GN=Igkv1-135 PE=4 SV=1
sp|P18527|HVM56_MOUSE	33.87	1.22	20	1	1	Ig heavy chain V region 914 OS=Mus musculus OX=10090 PE=1 SV=1
A0A140T8P6|A0A140T8P6_MOUSE	86.07	1.04	11	2	2	Immunoglobulin kappa variable 12-46 (Fragment) OS=Mus musculus OX=10090 GN=Igkv12-46 PE=1 SV=2
A0A0A6YWR2|A0A0A6YWR2_MOUSE	230.34	0.82	4	3	1	Ig gamma-1 chain C region secreted form (Fragment) OS=Mus musculus OX=10090 GN=Ighg1 PE=1 SV=1
A0A075B5P4|A0A075B5P4_MOUSE	230.34	0.82	5	3	1	Ig gamma-1 chain C region secreted form (Fragment) OS=Mus musculus OX=10090 GN=Ighg1 PE=1 SV=1
sp|P01837|IGKC_MOUSE	96.64	0.75	33	1	1	Immunoglobulin kappa constant OS=Mus musculus OX=10090 GN=Igkc PE=1 SV=2
sp|E9Q414|APOB_MOUSE	337.02	0.71	2	4	4	Apolipoprotein B-100 OS=Mus musculus OX=10090 GN=Apob PE=1 SV=1
sp|P03987-2|IGHG3_MOUSE	253.56	0.64	3	1	1	Isoform 2 of Ig gamma-3 chain C region OS=Mus musculus OX=10090
sp|P03987|IGHG3_MOUSE	253.56	0.64	3	1	1	Ig gamma-3 chain C region OS=Mus musculus OX=10090 PE=1 SV=2
sp|Q8BFZ3|ACTBL_MOUSE	37.48	0	3	1	1	Beta-actin-like protein 2 OS=Mus musculus OX=10090 GN=Actbl2 PE=1 SV=1

5.4.2 Validation of Dab2 and Ubap2l as protein interactors of exWAGO *in vivo*

To validate the findings of LC-MS/MS, the peptide identified for Upab2l (TAQALAQLAAQHSQSGSTTTSSWDMGSTTQSPSLVQYDLK) and Dab2 (SSANDLLASDIFASEPPGQMSPTGQPAVPQSNFLDLFK) were first checked for their specificity to the proteins reported, using the online tool BLASTP 2.13.0+ against the *Mus musculus* proteome (Altschul *et al*, 1997, 2005). Then, I performed reverse IPs where I immunopurified Upab2l, Dab2 and exWAGO using rabbit anti-Ubap2l polyclonal antibody (5 µg), mouse anti-Dab2 monoclonal antibody (5 µg) and rat anti-exWAGO polyclonal antiserum (6 µl) respectively. The antibodies were conjugated on Protein G beads (25 µl, 2 h) per IP. Infected gut tissue lysate was prepared and incubated with the antibody-conjugated beads (45 min) and the beads were washed only with low salt buffer. The eluted complexes were analysed using western blot analysis. Figure 5.13A shows that exWAGO IP successfully pulls down exWAGO. Ubap2l also seems to be pulled down, however, the band detected is around 85 kDa while Ubpa2l is 104-117 kDa (Fig. 5.13A). A strong band thought to correspond to Dab2 is also detected after exWAGO IP (Fig. 5.13A), but further experiments deemed that band to be non-specific as the Dab2 antibody detects a band at the right size (93 kDa) following IP of exWAGO from uninfected mouse gut (Fig. 5.13B). Anti-Ubap2l IP results suggest that Ubap2l was successfully pulled down, but it is inconclusive whether exWAGO interacts with it (Fig. 5.13A). It is clear, however, that Ubap2l does not interact with Dab2 (Fig. 5.13A). IP against Dab2 was also successful and consistent with the data from Ubap2l IP, we can deduce that Dab2 and Ubap2l do not interact (Fig. 5.13A). The western blot also shows that exWAGO is pulled down following Dab2 IP. This validates the LC-MS/MS findings. However, these results are preliminary and further validation experiments are required as the Dab2-exWAGO interaction is not always reproducible (data not shown). Experiments troubleshooting these findings indicate that the starting material (possibly related to the level of infection in the mouse gut tissue) used dictates whether exWAGO will be identified after IP with Dab2 using western blot analysis (data not shown).

Although Dab2 has never been reported in trafficking of AGO proteins, unpublished data from work led by Dr. Katrina Gordon in our lab implicates mAGO2 interacting with Dab2 *in vitro*. To test if I can recapitulate these findings using *in vivo* gut IPs, I immunopurified Dab2 using mouse anti-Dab2 monoclonal antibody (5 µg) from infected tissue and used IgG IP as a negative control (mouse IgG antibody, 5 µg).

The antibodies were conjugated on Protein G beads (50 µl, 6 h) per IP. Infected gut tissue lysate was prepared and incubated with the antibody-conjugated beads (overnight) and the beads were washed only with low salt buffer. The eluted complexes were analysed using western blot analysis. Figure 5.13C shows that Dab2 was successfully pulled down in the Dab2 IP but not in the IgG IP. Consistent with *in vitro* data from our lab (Gordon *et al*, in prep), mAGO2 is pulled down by Dab2 IP but not by the IgG control (Fig. 5.13C). Validation of the mAGO2-Dab2 interaction *in vivo* suggests that Dab2 interacts with AGO proteins, and it could facilitate AGO trafficking inside cells. Additionally, the findings indicate that the *in vivo* gut IP method can be used to validate protein interactions *in vivo*.

Identification of mAGO2 as an interactor of Dab2 and future investigation of this association could aid in understanding how exWAGO may interact with Dab2. As mentioned above, internalisation of proteins via Dab2 is facilitated by interaction between the PTB domain of Dab2 and the NxxY motif of the internalised protein (Finkielstein & Capelluto, 2016). To investigate whether exWAGO and mAGO2 possess such a motif that could support the interaction of them with Dab2, I analysed the protein sequence of exWAGO and mAGO2 using the online Eukaryotic Linear Motif resource (Kumar *et al*, 2020). Motif analysis identified that both exWAGO and mAGO2 have one motif that can be bound by Dab-like PTB domains, such as the domains that Dab2 possesses (Table 5.5). Interestingly, other endocytic motifs are also detected in exWAGO and mAGO2. These are tyrosine-based motifs that require interaction with the µ subunit of the Adaptor Protein complex, another clathrin-mediated endocytic mechanism.

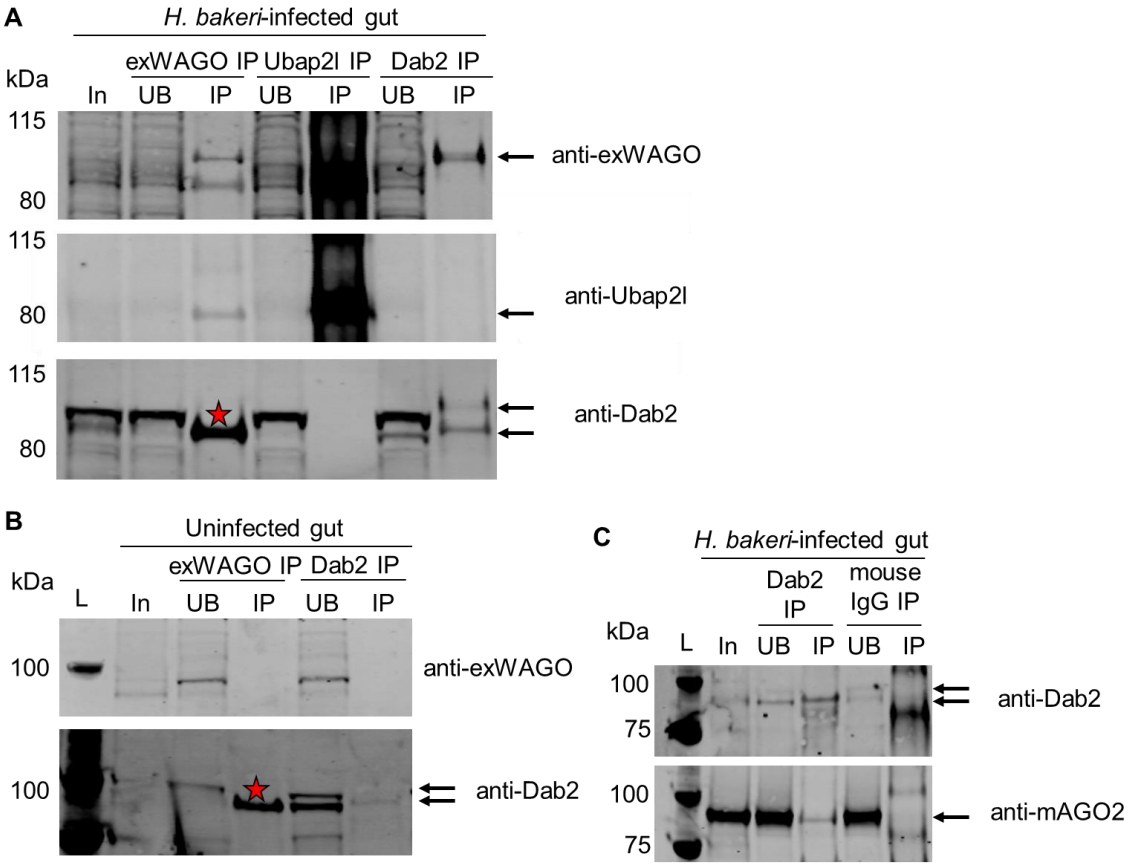


Figure 5.13 | Validation of Ubap2l and Dab2 as protein interactors of exWAGO and mAGO2 *in vivo.* (A) Western blot analysis of reverse IPs from infected gut tissue using Ubap2l and Dab2 to validate association with exWAGO. The blots were probed with 1:4,000 rabbit anti-exWAGO antibody followed by 1:10,000 goat anti-rabbit Dylight 800 (top blot), 1:2,000 rabbit anti-Ubap2l antibody followed by 1:10,000 goat anti-rabbit Dylight 800 (middle blot) and 1:2,000 mouse anti-Dab2 antibody followed by 1:10,000 goat anti-mouse AlexFluor 680 (bottom blot). The star denotes non-specific band detected in the exWAGO IP with the Dab2 antibody. Ubpab2l is expected to be 104-107 kDa, Dab2 is expected to be 93 and 96 kDa, and exWAGO is expected at 102 kDa as denoted by the arrows. Percentage of fraction analysed: 2% input (In), 2% unbound (UB), 16.7% eluate (IP). (B) Western analysis of reverse IPs from uninfected gut tissue. The blots were probed with 1:4,000 rabbit anti-exWAGO antibody followed by 1:10,000 goat anti-rabbit Dylight 800 (top blot), and 1:2,000 mouse anti-Dab2 antibody followed by 1:10,000 goat anti-mouse AlexFluor 680 (bottom blot). The star denotes non-specific band detected in the exWAGO IP with the Dab2 antibody. Dab2 is expected to be 93 and 96 kDa as denoted by the arrows. Percentage of fraction analysed: 2% input (In), 2% unbound (UB), 16.7% eluate (IP). L = ladder. (C) Western blot analysis of reverse IPs using Dab2 to validate association with mAGO2. The blots were probed with 1:2,000 mouse anti-Dab2 antibody followed by 1:10,000 goat anti-mouse AlexFluor 680 (top blot) and 1:4,000 mouse anti-mAGO2 antibody followed by 1:10,000 goat anti-mouse AlexFluor 680 (bottom blot). Dab2 is expected to be 93 and 96 kDa and mAGO2 is expected at 97 kDa as denoted by the arrows. Percentage of fraction analysed: 1% input (In), 1% unbound (UB), 16.7% eluate (IP). L = ladder.

Table 5.5 | exWAGO and mAGO2 possess various endocytic motifs. The table shows some of the endocytic motifs identified by the Eukaryotic Linear Motif (ELM) resource online tool (Kumar *et al,* 2020). The LIG_PTB_Apo_2 eukaryotic linear motif (.[^P].NP.[FY].)|(.[ILVMFY].N..[FY].) is interpreted as two patterns separated by the symbol "|". The first pattern is: any amino acid followed by any amino acid except P, followed by any amino acid, followed by NP, followed by any amino acid, followed by either F or Y, followed by any amino acid. The second pattern is: any amino acid followed by either I or L or V or M or F or Y, followed by any amino acid, followed by N, followed by any two amino acids, followed by either F or Y, followed by any amino acid. The TRG_ENDOCYTIC_2 eukaryotic linear motif Y..[LMVIF] is interpreted as follows: Y followed by any two amino acids, followed by either L or M or V or I or F.

Protein	ELM Name	Matched Sequence	Positions	ELM Description	Cell Compartment	Pattern	Pattern probability
exWAGO	LIG_PTB_Apo_2	PVINEFYL	804-811	These phosphorylation-independent motifs bind to Dab-like PTB domains.	cytosol, internal side of plasma membrane, integrin, receptor complex, cytoplasmic membrane-bounded vesicle	(.[^P].NP.[FY].)|(.[ILVMFY].N..[FY].)	3.108e-04
mAGO2		DMRNKQFH	437-444				
exWAGO	TRG_ENDOCYTIC_2	YKHL YKDV YMGI YPGM YYPV YLMF YQIV YNVM YVYI YSLL	92-95 99-102 233-236 364-367 379-382 558-561 580-583 620-623 688-691 826-829	Tyrosine-based sorting signal responsible for the interaction with mu subunit of AP (Adaptor Protein) complex	plasma membrane, clathrin-coated endocytic vesicle, cytosol	Y..[LMVIF]	2.587e-03
mAGO2		YTAM YPHL YAGL	102-105 323-326 513-516				

The proteomic data presented in this chapter suggest that exWAGO is involved in gene silencing inside the host, although further work is required to experimentally validate the interactions reported. Identification of Dab2 as an interactor of exWAGO from *in vivo* gut IPs also suggests that we discovered a potential mechanism for trafficking of AGO proteins and thus a potential mechanism of how the non-vesicular exWAGO may enter host cells.

But are non-vesicular proteins secreted by parasitic worms internalised by host cells? Although several groups study extracellular non-vesicular helminth proteins and report their immunomodulatory effects on host cells (Maizels *et al*, 2018), only a few studies investigate the mechanism(s) of uptake of the parasite protein(s) of interest. Robinson *et al* (2012) discovered that the *F. hepatica* protein HDM-1 (Helminth Defence Molecule 1) is taken up by endocytosis in macrophages after interacting with lipid rafts on the cell surface of the host. The ω-1 protein, a protein which possesses RNase activity, is secreted by the eggs of the human-infective worm *Schistosoma mansoni* and it is internalised by dendritic cells through association of the mannose receptor with the Le^x (Lewis X) motif of the ω-1 protein (Everts *et al*, 2012). Internalisation of the protein supresses activation of dendritic cells and downstream T-cell responses by affecting protein synthesis in the host, by degrading rRNA and mRNA transcripts (Everts *et al*, 2012). The best-studied non-vesicular extracellular worm proteins of *H. bakeri* act at the interface between the parasite and the host or in necrotic host cells, rather than inside live host cells. Examples include the *H. bakeri* TGF-β mimic (Transforming Growth Factor beta) which interacts with the mammalian TGF-β receptor to induce a signalling cascade but it is not itself expected to be internalised (Johnston *et al*, 2017). The proteins *H. bakeri* BARI (Binds Alarmin Receptor and Inhibits) and BARI_Hom2 bind the IL-33 receptor (ST2), thus blocking the receptor without being internalised, and hence control activation of Type 2 immune responses (Vacca *et al*, 2020). Moreover, the protein *H. bakeri* ARI (Alarmin Release Inhibitor) which inhibits the release of the IL-33 cytokine, traffics inside necrotic cells through their permeabilised membrane rather than live cells (whose membranes are impermeable) (Osbourn *et al*, 2017). It is thus important to understand if and how the non-vesicular exWAGO is internalised by live host cells and generally whether non-vesicular extracellular *H. bakeri* proteins are internalised by live host.

5.5 Non-vesicular proteins can be internalised by host cells *in vitro*

Identification of Dab2 as a putative interactor with exWAGO that could mediate its internalisation into host cells led us to question whether we can detect the non-vesicular exWAGO inside host cells. To begin to investigate this, I incubated MODE-K cells with EV-depleted HES (50 µg/ml) for 4 h and 24 h, and then with the help of Dr. Tom Fenton looked for exWAGO by immunofluorescence. We could not detect uptake of the non-vesicular exWAGO using the EVOS M7000 instrument (ThermoFisher) (data not shown). This could be a sensitivity issue and future work in the lab will test additional antibodies and methods of detection.

Despite not detecting exWAGO after incubation of MODE-K cells with EV-depleted HES, I wanted to determine whether non-vesicular proteins can be internalised by live host cells. For this, I incubated MODE-K cells with Cy-5 labelled EV-depleted HES (10 µg/ml or 50 µg/ml) or Cy-5 labelled PBS (equal volume as required for the 10 µg/ml or 50 µg/ml of EV-depleted HES) for 4 h and 24 h. These preliminary data show immunofluorescence in MODE-K cells and the signal appears mainly in the cytoplasm (Fig. 5.14), suggesting that non-vesicular proteins secreted by *H. bakeri* are internalised by mouse cells *in vitro*. However, further work is required to prove that the non-vesicular proteins have actually been internalised and that they do not remain at the cell surface.

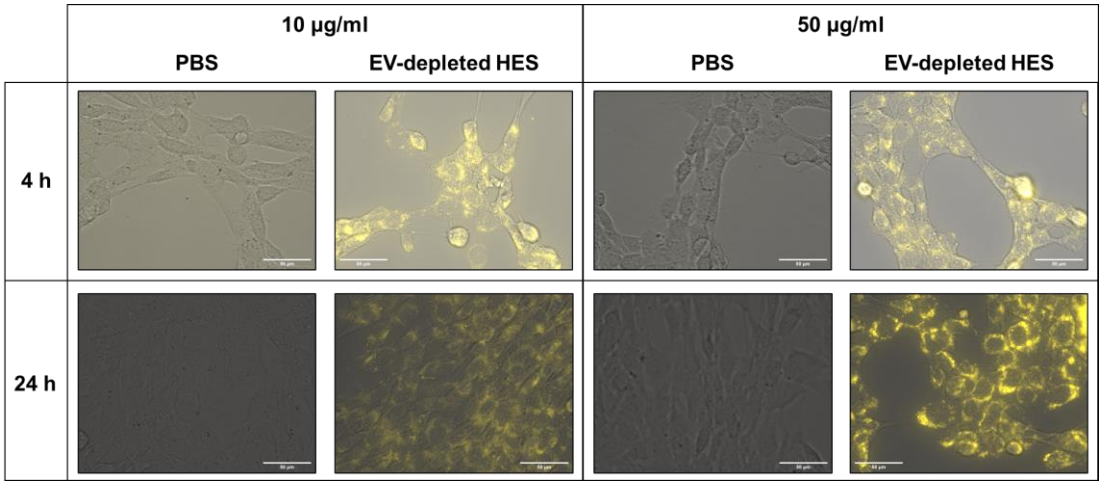


Figure 5.14 | Non-vesicular proteins secreted by *H. bakeri* are internalised by MODE-K cells *in vitro*. Scale bar = 50 µm.

Discussion

5.6 Summary

In this chapter, I set out to understand the mechanism of action of exWAGO in gene silencing. Using slicer assays I was able to show that exWAGO does not possess endonucleolytic activity, consistent with the lack of slicer motif (Fig. 5.1). To understand if exWAGO is directly implicated in gene silencing, I set out to identify whether exWAGO interacts with proteins that other AGOs associate with. First, I identified the protein interactors of exWAGO inside adult worms. The results show that exWAGO interacts with known protein partners of AGOs such as CEY proteins (YBX-1) and PAB-1, implying that exWAGO functions in gene silencing (Fig. 5.2, 5.3, Table 5.1). These data also suggested that exWAGO functions in specialised compartments (P-bodies) and that it is expressed and likely secreted from the worm intestine, cephalic sheath, and eggs (Fig. 5.4, Table 5.2). To determine if exWAGO interacts with host proteins that may be implicated in gene regulation (in line with our hypothesis that exWAGO manipulates host gene expression), I carried out immunoprecipitation of exWAGO from a recombinant cell line expressing exWAGO or from mouse gut tissue followed by LC-MS/MS. The data from *in vitro* experiments revealed that exWAGO may interact with some host proteins involved in gene silencing such as SFPQ but further validations of these interactions are required (Fig. 5.9, Supp. Table. 5.3). These results along with immunofluorescence experiments suggest that exWAGO might operate in the nucleus consistent with localisation of SFPQ, but also in the cytoplasm (Fig. 5.9). Proteomic data from *in vivo* material indicates that exWAGO might specifically localise to stress granules after identification of Ubap2l as a protein interactor (Table 5.4). Additionally, identification of the adaptor protein Dab2 as a putative interactor of exWAGO could be an important finding, as Dab2 is involved in internalisation of proteins, and hence its putative association with exWAGO prompts us to hypothesise that Dab2 could be involved in trafficking of AGO proteins. This further implies that the non-vesicular exWAGO could be internalised by mouse host cells (Table 5.4). However, we do not yet have any evidence of non-vesicular exWAGO uptake inside host cells. As a preliminary investigation, I was able to demonstrate that helminth non-vesicular extracellular proteins are taken up by live host cells *in vitro* (Fig. 5.14).

5.7 The silencing ability of WAGOs despite the lack of slicing activity

The biochemical confirmation that exWAGO does not possess slicer activity is not unexpected (Fig. 5.1). As shown in Chapter 3, exWAGO binds 5'PPP molecules, which is reminiscent of the cargo loaded in the *C. elegans* WAGO proteins. AGO proteins of the WAGO clade in *C. elegans* do not possess a catalytic motif and therefore are not expected to have slicer activity (Yigit *et al*, 2006), with the exception of the WAGO protein CSR-1 (Aoki *et al*, 2007). So how do WAGO proteins induce gene silencing if not by cleaving the mRNA target(s)? Understanding how WAGO proteins are involved in gene silencing without possessing slicer activity can shed light on how exWAGO may be involved in RNAi.

The specific mechanism(s) by which most WAGOs supress target expression remain largely unknown. Evidence, however, indicates that WAGOs can be involved in both epigenetic/transcriptional and post-transcriptional gene silencing. WAGO-9/HRDE-1 in the germline and WAGO-12/NRDE-3 in the soma are implicated in epigenetic gene regulation by promotion of the repressive histone mark Histone 3 Lysine 9 trimethylation (H3K9me3) on targets in the nucleus (Guérin *et al*, 2014; Ketting & Cochella, 2020). The nuclear WAGO-9/HRDE-1 interacts with the SUMOylated histone deacetylase HDAC1 to initiate epigenetic gene silencing, which in turn interacts with the histone methyl-transferase MET-2 (establishes H3K9me2/3 silencing marks) and the histone demethylase SPR-5 (removes the active histone marks H3K4me2/3) (Kim *et al*, 2021). WAGO-9/HRDE-1 also interacts with the heterochromatin-like protein HPL-2 which is involved in maintaining heritable epigenetic gene silencing (Kim *et al*, 2021). Interestingly, studies in *C. elegans* by Reed *et al*, (2020) and Kim *et al*, (2021) suggest that AGO proteins (the primary AGO protein, PRG-1) and other WAGO-pathway factors known as mutator proteins (involved in synthesis of the 22G sRNAs) are involved in promoting transition of spermatogenesis to oogenesis in hermaphrodite worms, by interacting with SUMOylated HDAC1. This might link to the findings that exWAGO is expressed in the eggs and that it interacts with proteins that make up the eggshell (Fig. 5.4, Table 5.2). Furthermore, WAGO-12/NRDE-3 is involved in epigenetic gene silencing by associating with other NRDE factors and the nascent transcript targeted in the nucleus. The NRDE factors repress elongation of RNA polymerase II and promote deposition of H3K9me3 repressive marks (Guang *et al*, 2010). Additionally, along with protein HPL-2, NRDE-3 directs compaction of chromatin (Fields & Kennedy, 2019).

WAGO-1 is involved in post-transcriptional gene silencing and is important for gene silencing in the germline as it targets transposons, pseudogenes and specific genes (Gu *et al*, 2009). Another WAGO, WAGO-4, is expressed in the germline and involved in post-transcriptional gene silencing of mainly protein-coding genes. WAGO-4 also mediates RNAi inheritance across multiple *C. elegans* generations by associating with 22G sRNAs that have been uridylated by the poly(U) polymerase CDE-1 (Xu *et al*, 2018). Heritable RNAi of exogenous or endogenous sRNAs across generations allows the descendant to carry these sRNAs which act as an immune defence memory mechanism against transposons and viruses for example that had previously affected the parent organism and could negatively affect the genome of the descendant (Xu *et al*, 2018).

The *C. elegans* orthologues SAGO-1, SAGO-2 (expressed in the soma (Seroussi, unpublished)) and PPW-1 (expressed in the germline, (Tijsterman *et al*, 2002)) also act to supress gene expression in a post-transcriptional manner, although the exact mechanism by which they silence genes remains to be determined (Yigit *et al*, 2006). PPW-1 and SAGO-2 bind guides targeting genes related to *C. elegans* immune response genes (Seroussi, unpublished). As SAGO-1, SAGO-2 and PPW-1 (iSAGOs) are phylogenetically the closest *C. elegans* AGO proteins to exWAGO (Chow *et al*, 2019), it is of interest to understand what role the iSAGOs play in gene silencing inside the worm itself, but also understand if and how iSAGOs are implicated in extracellular communication. This is based on the recent findings reporting that AGO proteins, including PPW-1, are secreted by *C. elegans* in EVs (Nikonorova *et al*, 2021). This information will help us determine the specific function of exWAGO inside adult worms and how it might be implicated in intraspecies (worm-to-worm) communication. However, as *C. elegans* is a non-parasitic nematode, data on the exWAGO orthologues are deemed unlikely to inform us on the interspecies function of exWAGO with the mouse host.

5.8 Protein interactors of Argonaute proteins

When investigating the functional mechanism of exWAGO inside host cells, we hypothesise that exWAGO will interact with the host RNAi machinery to operate, rather than utilise other worm proteins to silence host genes. Thus, we set up our experiments to identify mouse protein interactors of exWAGO inside host cells rather than testing if the extracellular forms of exWAGO are found as part of a protein complex. Our proteomics data suggest that exWAGO might interact with host proteins

involved in gene silencing and gene regulation, such as SFPQ, LSM14a, Ubap2l and HSP90ab1 (Supp. Table 5.3, Table 5.4). However, the data produced and presented in this chapter require further experimental validation to determine whether interactions with exWAGO are bona fide and reproducible. Identification of Dab2 as an interactor of exWAGO *in vivo* (Fig. 5.13, Table 5.4) is exciting as it provides one possible mechanism by which vesicle-free parasite-derived proteins can be internalised into live host cells. Additionally, further experiments showing that the mammalian Ago2 also interacts with Dab2 *in vivo* (Fig. 5.13C), supports a hypothesis that Dab2 could enable mammalian AGO proteins to traffic between cells. It is of interest to understand where exWAGO is localised inside host cells. Nuclear localisation of exWAGO would suggest transcriptional or co-transcriptional gene silencing whereas localisation in the cytoplasm implies involvement in post-transcriptional gene silencing.

The proteomics data obtained from adult worms provides further evidence that exWAGO interacts with known AGO proteins including homologues of YBX-1 and PAB-1 (Fig. 5.3, Table 5.1). This finding supports the idea that exWAGO functions in gene silencing inside *H. bakeri* adult worms. Tissue enrichment analysis reveals that exWAGO interacts with proteins derived from the intestine and glial cells (Fig. 5.4C, Table 5.2B). These findings link to the idea of where exWAGO could be secreted from, as EVs released by *H. bakeri* carry intestinal proteins (Buck *et al*, 2014), and ciliary EVs secreted by *C. elegans* contain AGO proteins (Nikonorova *et al*, 2021), including the orthologue of exWAGO, PPW-1, along with other eggshell-related proteins also identified in this chapter (Fig. 5.3, Table 5.1) (Nikonorova *et al*, 2021). It is worth noting however that the *C. elegans* EVs examined by Nikonorova *et al*, (2021) could contain EVs derived from other tissues too, indicating that further investigations are required to determine from where exWAGO is secreted.

5.9 Future work

The lack of reproducibility in the mass spectrometry results obtained from exWAGO IPs indicate that the method employed here requires further optimisation to identify bona fide protein interactors of exWAGO. For example, crosslinking can be employed to capture protein-protein interactions that are weak or transient (Lenz *et al*, 2021). Additionally, for quantitative analysis, label-free quantitation should be employed as the dimethyl labelling of peptides results in loss of material. Additionally, more sensitive mass spectrometry instruments could be employed to detect lowly

expressed proteins or aid in identification of proteins when the material is limiting. Regarding the current datasets available, extensive validation experiments including reverse IPs are required to confirm that the proteins identified by LC-MS/MS analyses interact with exWAGO.

To understand if exWAGO directly mediates gene silencing in mouse host cells several experiments could be employed. For example, we could load recombinant exWAGO with a 5'PPP guide that targets a reporter gene such as GFP and transfect the loaded exWAGO into a GFP-expressing cell line. We can then detect by various means (qPCR, western blot, and imaging analyses) if the levels of the GFP have been altered compared to transfection of exWAGO loaded with a scrambled sequence. This experiment will indicate the ability of exWAGO to supress gene expression, although we need to consider what region of the target gene would be deemed physiologically relevant. For example, data from Chapter 4 suggest that exWAGO might target intronic regions. Thus, the experiment should be designed in such a way where we direct exWAGO to introns of the gene target. Another suggestion is to incubate MODE-K cells or intestinal organoids with HES that is depleted of exWAGO and perform RNA sequencing to detect if the expression levels of certain transcripts are altered in comparison to treating cells or organoids without exWAGO depleted from the starting material. This experiment will help define the ability of the native exWAGO to silence host genes and would also identify host genes that are specifically targeted by the parasite guides in association with exWAGO. Furthermore, it could be important to distinguish and understand whether exWAGO actually interacts with host machinery involved in gene silencing or whether exWAGO is already in a complex with worm-derived RNAi proteins when it is secreted from the parasite.

In the next chapter, we seek to test if vaccination with exWAGO confers protection against infection with *H. bakeri*. This will help us understand whether exWAGO could be considered a vaccine candidate for other gastrointestinal nematodes that secrete exWAGO orthologues. Moreover, immunisation with exWAGO may help us understand the importance of exWAGO in infection.

Chapter 6: Vaccine trials with exWAGO

Introduction

It is clear that the prevalence, impact on human health, and agricultural losses caused by gastrointestinal worm infections need to be addressed. This is recognised by organisations such as the WHO, COMBAR, LiHRA, and STAR-IDAZ International Research Consortium who have developed roadmaps and set targets with the ultimate goal of eliminating human helminth infections and controlling livestock worm infections (COMBAR, 2020; WHO, 2020). One of the strategies proposed to achieve this goal is the use of therapeutic interventions (COMBAR, 2020; WHO, 2020) and development of vaccines (COMBAR, 2020). Anthelmintic drugs can be used to resolve parasitic worm infection including flatworm and roundworm infections (McKellar & Jackson, 2004). However, the wide use of these drugs has led to anthelmintic resistance to all available drugs in the worms that infect livestock (Wit *et al*, 2021). Some cases of anthelmintic resistance in human-infective worms have also been reported and this is expected to rise as more mass anthelmintic drug administration programmes are established (Orr *et al*, 2019; Wit *et al*, 2021).

Despite the implementation of mass drug administration programs and provision of improved water, sanitation and hygiene (WASH) conditions in different communities, the prevalence of helminths is still high in some areas (Perera & Ndao, 2021). This is attributed to low efficacy of drugs, lack of additional control strategies, and re-infections (Perera & Ndao, 2021). The aforementioned reasons along with the widespread anthelmintic drug resistance in ruminants, highlight the need for multiple strategies to be used to tackle helminth infections (COMBAR, 2020; Perera & Ndao, 2021), including the need to develop anti-helminth vaccines (COMBAR, 2020; Perera & Ndao, 2021). Vaccines can provide lasting protection against infection and do not contribute to chemical residues in the environment and in animal products (Claerebout & Geldhof, 2020). Vaccination of humans and animals is expected to induce immune memory against a future infection with the parasite (Claerebout & Geldhof, 2020).

Vaccine development against helminths is tricky due to the complex life cycle of helminths and the evasion mechanisms helminths exhibit to manipulate the host immune system (Perera & Ndao, 2021). The antigen-mimicry and ability of helminths to avoid immune responses need to be taken into consideration when developing vaccines. Hence, the so-called correlates of immunity need to be well understood for

development of promising anti-helminthic vaccines (Perera & Ndao, 2021). This involves understanding the immune responses that are required to be triggered by the vaccine candidate to confer protection against the parasite. Currently there are no approved anti-helminth vaccines for human use (Perera & Ndao, 2021). However, there are vaccine candidates in clinical trials (Zawawi & Else, 2020). For example, there are two vaccines against the human hookworm *N. americanus* consisting of Na-APR-1 (aspartic protease-1) and Na-GST-1 (glutathione-S-transferase-1) in phase I of clinical trials (Diemert *et al*, 2017; Bottazzi & Diemert, 2019). To produce effective vaccines, it is predicted that more than one parasite antigen might be required, thus vaccines with multiple antigens have also been investigated, including a vaccine combining both Na-APR-1 and Na-GST-1 antigens (Adegnika *et al*, 2021). Furthermore, a pan-anthelmintic vaccine has been proposed to tackle the three major human-infective gastrointestinal helminths: *A. lubricoides, T. trichiura*, and *N. americanus* (Zhan *et al*, 2014). This is currently under development and the concept is to combine *Ascaris*, *Trichuris* and *Necator* antigens to generate one vaccine. The rationale behind this is based on the fact that co-infections with multiple gastrointestinal nematodes are very common in children (Silva *et al*, 2003; Zhan *et al*, 2014). The same approach could be employed for control of animal-infective helminths to tackle co-infections.

For animal use there are four licensed anti-helminth vaccines which control for the bovine lungworm nematode *D. viviparus,* the sheep nematode *H. contortus,* the sheep- and goat-infective cestode *Echinococcus granulosus* (Claerebout & Geldhof, 2020), and the pig- & human-infective *T. solium* (Ouma *et al*, 2021). There are various strategies that serve as a basis for development of vaccines. These include radiation-attenuated whole parasites, proteins (in adjuvant), nucleic acid-based and viral vector-based technologies *(Perera & Ndao, 2021)*. The *D. viviparus* vaccine uses irradiated L3 stage larvae to provide high protection levels against infection (Claerebout & Geldhof, 2020). However, the use of irradiated larvae was unsuccessful in young lambs infected with *H. contortus,* hence other vaccine platforms were employed to control haemonchosis (Nisbet et al, 2016). As a result, research in *Haemonchus* has focused on identification of proteins/antigens on the cuticular surface of nematodes or gut-derived antigens which are not detected by the host immune system during infection as they are ‘hidden’ (Nisbet et al, 2016). This led to development of the current vaccine for *H. contortus* which is comprised of two *H. contortus* gut membrane proteins, H11 and H-gal-GP (*Haemonchus* galactose

containing glycoprotein complex) (Smith et al, 2001; Kebeta et al, 2021).The ES (excretory/secretory) products of helminths have also been an attractive target for vaccine development against helminth infections (Nisbet et al, 2016). For example, the approved vaccine *for E. granulosus* utilises the secreted protein EG95, which is highly expressed in oncospheres (larva) (Rosenzvit et al, 2006; Claerebout & Geldhof, 2020).

The principle of using ES antigens for vaccination is based on the hypothesis that immune responses (such as antibodies) block the putative immunomodulatory function of these secreted molecules (Hewitson et al, 2015). The other possibility is that the antigens used in vaccination generate immunity that is directed to the worms themselves, with antibodies, for example, binding and attacking the worm. This was shown to be the case for the *S. mansoni* Sm-TSP-2 protein which elicits significant reduction in worm and egg burdens following vaccination with adjuvant (Tran et al, 2006). Sm-TSP-2 is a tetraspanin protein expressed on the worm tegument and found on the membrane of *S. mansoni* extracellular vesicles (EVs) (Tran et al, 2006; Sotillo et al, 2016). Does the immunity generated by the vaccine block the function of EVs carrying this tetraspanin or does it target the worm? Although it cannot be ruled out that immune responses are directed to EVs and block EV function, antibodies against the Sm-TSP-2 protein also bind the tegument of the worms (Tran et al, 2006; Pearson et al, 2012). EVs can be secreted from the tegument (Drurey et al, 2020), thus it is not surprising that vaccination with some ES antigens elicits immune responses that not only may neutralise the function of the secreted molecule in the extracellular environment, but also can directly attack helminth parasites. Another example of a vaccine antigen directing immune responses to the worms is shown by the *N. americanus* protein Na-APR-1. The Na-APR-1 protein is found anchored on the gut surface and is involved in proteolysis of haemoglobin and thus it plays a role in parasite feeding (Loukas et al, 2005). The Na-APR-1 vaccine is currently in clinical trials for human nematode infections and it has been shown that antibodies generated during vaccination with Na-APR-1 are ingested and bind to the gut of the parasite (Loukas et al, 2005). Recognition of Na-APR-1 by antibodies neutralises the enzymatic function of the protein in vitro and thus it is proposed that the protective effects of Na-APR-1 vaccination are attributed to impairment of parasite feeding by inhibiting digestion of host haemoglobin (Pearson et al, 2010). It is important to note however that this protein has not been detected in the ES products of the worm in

recent mass spectrometry analysis (Logan et al, 2020), hence it is not considered a secreted protein.

Several vaccination studies have also been performed in mice with antigens collected from the ES products of the model organisms for intestinal nematode infections, *H. bakeri* and *N. brasiliensis*. In particular, work from the Maizels group showed that immunisation with total HES confers sterile immunity against larval challenge (Hewitson et al, 2015). Vaccination with a cocktail of purified native VAL proteins (Venom-allergen proteins) which are highly abundant in HES, and vaccination with HES depleted of VAL proteins were also shown to effectively protect against *H. bakeri* infection (Hewitson et al, 2015). Collaborative work with our lab shows that *H. bakeri* EVs as well as EV-depleted HES also generate strong protection against infection (Coakley et al, 2017). Moreover, partial protection against *H. bakeri* challenge is also evoked after immunisation of mice with a cocktail of recombinant apyrase proteins which are secreted by *H. bakeri* and catalyse hydrolysis of nucleoside triphosphates and diphosphates *in vitro* (Berkachy et al, 2021). The study by Berkachy et al, (2021) is the first study to show protection against *H. bakeri* by immunisation with (a cocktail of) recombinant proteins (these were expressed in yeast). Vaccination with recombinant proteins is considered challenging since in its recombinant form the antigen might not adopt the correct conformation. Successful use of recombinant proteins in nematode vaccines is not a novel finding (Geldhof et al, 2007; Claerebout & Geldhof, 2020). For example, partial immunity has been observed in rats vaccinated with recombinant *N. brasiliensis* acetylcholinesterase (expressed in yeast) which is extracellular, following infection with *N. brasiliensis* (Ball et al, 2007). Additionally, immunisation of sheep with a cocktail of 8 recombinant proteins also generates partial protection against the nematode *T. circumcincta* (Nisbet et al, 2013). Nevertheless, the findings of Berkachy et al, (2021) are exciting because they suggest that *H. bakeri* can be used to test recombinant vaccines against gastrointestinal nematodes.

As exWAGO is (1) hypothesised to possess immunomodulatory abilities (when complexed with sRNAs), (2) an extracellular protein secreted by *H. bakeri*, (3) is found in the non-vesicular fraction of HES hence presumably the absence of the EV lipid bilayer makes exWAGO vulnerable to antibodies, and (4) since it may derive from the intestine where immune responses might have access during parasite feeding, we set out to evaluate whether exWAGO can be a vaccine candidate for *H. bakeri* infection. To test this hypothesis, we vaccinated mice with recombinant exWAGO (which was

expressed in baculovirus-insect cells). We hypothesise that the acquired immunity induced by the vaccine will prime the mice so that once they encounter exWAGO during *H. bakeri* infection they will be able to neutralise it. As well as investigating the possibility of exWAGO as a target, immunisation with exWAGO will also (indirectly) assess the functional importance of exWAGO by examining the effect on parasite survival *in vivo* during its absence, since *H. bakeri* is a non-genetically tractable organism and gene knockdown is currently not feasible (Lendner et al, 2008). If exWAGO immunisation confers protection against *H. bakeri* infection we reason that exWAGO could be a vaccine candidate for other Clade V parasitic gastrointestinal nematodes that also express and secrete highly conserved exWAGO orthologues (Chow et al, 2019) (Table 1.1).

Results

6.1 Experimental design for mice vaccination with exWAGO

To assess whether exWAGO could be a vaccine candidate we sought to examine its effect on parasite survival *in vivo* during its absence. We immunised C57BL/6 female mice with an intraperitoneal (i.p.) injection of recombinant exWAGO in Imject Alum (ThermoFisher Scientific) or QuilA (Croda Europe Limited) adjuvant and compared its effects to immunisation with PBS (negative control) or HES (positive control) in the appropriate adjuvant. Adjuvants are used to boost or direct immune responses in the immunised subject (McNeil & DeStefano, 2018). Alum (aluminum hydroxide) adjuvants are routinely used in human vaccines, however the exact mechanism by which they function remains unclear (Moyer *et al*, 2020; Wang & Xu, 2020). Evidence suggests that alum generally induces weak cellular immune responses but strong humoral responses (i.e. antibody-mediated responses) such as IgG1 (Kensil, 2006; Moyer *et al*, 2020). The strong humoral responses are induced by enhancing B cell activation via increased antigen processing and presentation once the antigen decorated with alum particles is taken up by B cells (Moyer *et al*, 2020). On the other hand, QuilA is a purified saponin extract derived from the tree *Quillaja saponaria*. It is used in veterinary studies but it is toxic to humans (Garçon *et al*, 2013). For this reason QuilA is used in a modified form for use in human applications (Garçon *et al*, 2013). QuilA has been reported to elicit both cellular and humoral responses. It can induce production of both IgG1 and IgG2a antibodies thus it generates mixed Th1 (characterised by IgG2a) and Th2 (characterised by IgG1) cell responses (Kensil, 2006). As QuilA is often used for vaccine development in ruminants, we decided to test if exWAGO in QuilA adjuvant may confer protection against infection with *H. bakeri*. This is based on the fact that exWAGO is highly conserved in other Clade V gastrointestinal nematodes which infect ruminants (such as *T. circumcincta*) and exWAGO could be a vaccine candidate for these animal-infective nematodes too.

Briefly, mice were primed with 10 µg of protein and then received two booster doses of 2 µg of protein each. After vaccination, all the mice were challenged with 200 *H. bakeri* L3 stage larvae. We performed three separate vaccination experiments which are referred to as Vac01, Vac02, and Vac03 in this thesis. Vac03 was interrupted by the SARS-CoV-2 outbreak and lockdown, which meant that mice received the first boost after 4 months since priming, when normally they would have received it after 1 month. The immunisation timeline is shown in Figure 6.1 and the experimental

design is detailed in Table 6.1. Several samples and measurements were taken during the experiments, including serum collection for investigating antibody responses and collection of stools for faecal egg counts to monitor the infection from 14 days post-challenge (p.c.) onwards, when the adult worms have emerged in the gut lumen, mated, and the females have started producing eggs. Mice were sacrificed at days 14, 28 and/or 35 p.c. for enumeration of worm burdens. The design of the experiment, sample collection plan and analysis were carried out by the author but required support from other colleagues (see Contributions section, pg. iii).

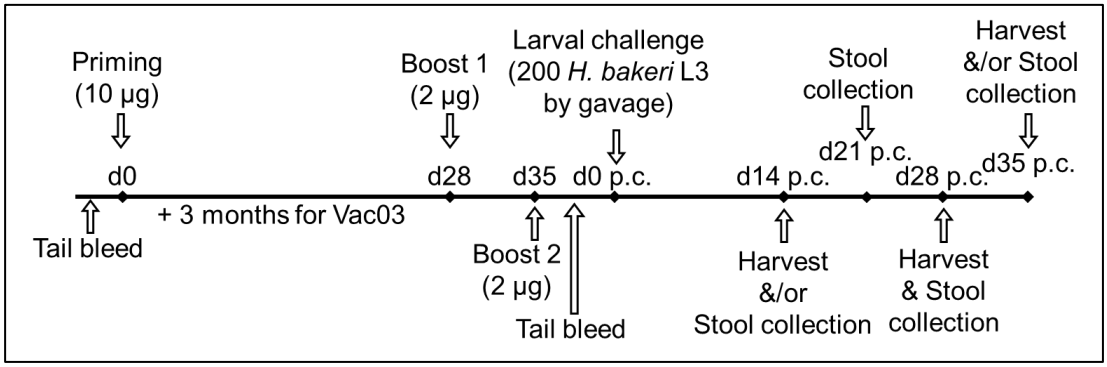


Figure 6.1 | Vaccination timeline. d = day; p.c. = post-challenge; Vac03 = third independent experiment which was interrupted by the SARS-CoV-2 outbreak and lockdown.

Table 6.1 | Details of the three independent vaccination experiments (Vac01, Vac02 and Vac03). d = day; p.c. = post-challenge

		Experiment		
		Vac01	Vac02	Vac03
First boost with 2 µg delayed by 3 months		-	-	✓
Tail and/or cardiac bleed	Pre-immunisation & Pre-challenge	-	-	✓
	Post-immunisation & Pre-challenge	-	-	✓
	Post-immunisation & Post-challenge	✓	✓	✓
Groups	PBS Alum	✓	✓	✓
	exWAGO Alum	✓	✓	✓
	HES Alum	✓	✓	✓
	PBS QuilA	-	-	✓
	exWAGO QuilA	-	-	✓
Harvest day	d14 p.c.	✓	-	-
	d28 p.c.	✓	✓	✓
	d35 p.c.	-	✓	✓
Stool collection	d14 p.c.	✓	✓	✓
	d21 p.c.	✓	✓	✓
	d28 p.c.	✓	✓	✓
	d35 p.c.	-	✓	✓

6.2 Vaccination with exWAGO confers partial protection against challenge with *H. bakeri* larvae

To measure the effect of PBS, exWAGO and HES immunisation on *H. bakeri* infection, we counted the number of eggs present in faeces over the course of the infection and also the number of adult worms within the intestine of mice sacrificed at the indicated times. The egg counts provide a proxy to assess infection levels in a non-invasive manner. As expected, vaccination with HES in alum confers sterile immunity as seen by a dramatic drop in the number of eggs per gram of faeces at 14 days p.c. (post-challenge) and clearance of worms by day 28 p.c. compared to mice vaccinated with PBS in alum (Fig. 6.2). These results are in agreement with the published literature (Hewitson *et al*, 2015). Moreover, vaccination with exWAGO in alum adjuvant confers partial protection to *H. bakeri* larval challenge as described in detail below. The author would like to point to the reader that the data presented have been pooled from the three experiments performed during this project. Although our data shows that overall vaccination with exWAGO in alum confers partial protection, exWAGO vaccination in the Vac02 experiment did not yield in a significant decrease of egg or worm burdens to that of PBS vaccination, in contrast to exWAGO vaccination in the Vac01 and Vac03 experiments (discussed further in discussion). The averages of all three experiments are described here and missing data was not included in the analyses.

Vaccinating with exWAGO in alum adjuvant leads to a 66.7% decrease in worm burden compared to vaccination with PBS in alum by day 28 p.c. (Fig. 6.2A, Table 6.2). This is observed at day 35 p.c. too where exWAGO-vaccinated (with alum) mice carry on average 62% less worms compared to the PBS-vaccinated (with alum) mice (Fig. 6.2A, Table 6.2). Consistent with the reduction in adult worms, exWAGO vaccination (in alum) leads to a 47.6% decrease in the number of eggs per gram of faeces compared to PBS-vaccinated (with alum) mice by day 14 p.c. (Fig. 6.2B, Table 6.2). This decrease is observed throughout the experimental timeline as days 21, 28 and 35 p.c. show 59.0%, 49.5% and 30.8% decrease in egg counts respectively when comparing the number of eggs per gram of faeces in exWAGO-vaccinated (with alum) mice to PBS-vaccinated (with alum) mice (Fig. 6.2B, Table 6.2). These data show that immunisation of mice with exWAGO provides partial immunity against larval challenge and suggests that exWAGO is functionally important during infection. Identification of adult worms in the exWAGO-vaccinated mice at day 14 p.c. also suggests that

vaccination did not affect the worm lifecycle and the development of larvae to adult worms, since the presence of adult worms in the gut lumen indicates that the larvae were able to penetrate the gut wall, develop to adult worms and then emerge in the gut lumen (i.e. this is not sterile immunity, as observed with HES in Hewitson *et al*, (2015). In turn, this might suggest that exWAGO plays a role in the ability of the parasite to maintain chronic infection in the host rather than being required to establish the infection.

Vaccination of mice with exWAGO in QuilA adjuvant did not reduce the number of worms or the number of eggs compared to mice vaccinated with PBS in QuilA (Fig. 6.2, Table 6.2). This indicates that exWAGO in QuilA adjuvant does not confer protection against *H. bakeri* infection, in contrast to vaccination with exWAGO in alum adjuvant. Interestingly, the worm and egg burdens of PBS-vaccinated mice with QuilA adjuvant are lower by a factor of 2.4 and 2.7 respectively at day 35 p.c., compared to mice vaccinated with PBS in alum adjuvant (Fig. 6.2). This suggests that the QuilA adjuvant has some protective effect on its own, but the presence of exWAGO in QuilA does not induce further protection. Although the mode of action of each adjuvant remains unclear, we hypothesise that vaccination with QuilA adjuvant induces mixed type 1 and type 2 immune responses while alum adjuvant induces type 2 immune responses (Kensil, 2006), which are required to clear gastrointestinal worm infections (Allen & Sutherland, 2014). Further studies will be required to understand why vaccination with exWAGO in QuilA adjuvant does not generate immunity against larval challenge compared to vaccination with exWAGO in alum.

Table 6.2 | The percentage reduction of worm and egg burdens in exWAGO-vaccinated mice compared to PBS-vaccinated mice. Pooled data are calculated using the average worm or egg numbers across Vac01, Vac02 and Vac03. Negative values indicate that there was an increase in the worm or egg burdens in the exWAGO-vaccinated mice compared to PBS-vaccinated mice. The asterisk (*) denotes that only data from Vac01 experiment were considered. d = day; p.c. = post-challenge; n.d. = no data; p.d. = partial data derived from PBS-vaccinated mice only.

% reduction of worm burdens					
Adjuvant	Day p.c.	Vac01	Vac02	Vac03	Pooled data
Alum	d14	-49.6	n.d.	p.d.	-49.6*
	d28	58.4	41.2	82.5	66.7
	d35	n.d.	15.8	86.3	62.0
QuilA	d14	n.d.	n.d.	n.d.	n.d.
	d28	n.d.	n.d.	-69.1	-69.1
	d35	n.d.	n.d.	5.8	5.8
% reduction of egg burdens					
Adjuvant	Day p.c.	Vac01	Vac02	Vac03	Pooled data
Alum	d14	-85.9	5.9	76.8	47.6
	d21	85.0	9.8	73.3	59.0
	d28	90.5	-9.7	81.8	49.5
	d35	n.d.	7.1	46.6	30.8
QuilA	d14	n.d.	n.d.	-29.3	-29.3
	d21	n.d.	n.d.	-21.1	-21.1
	d28	n.d.	n.d.	-77.1	-77.1
	d35	n.d.	n.d.	-55.9	-55.9

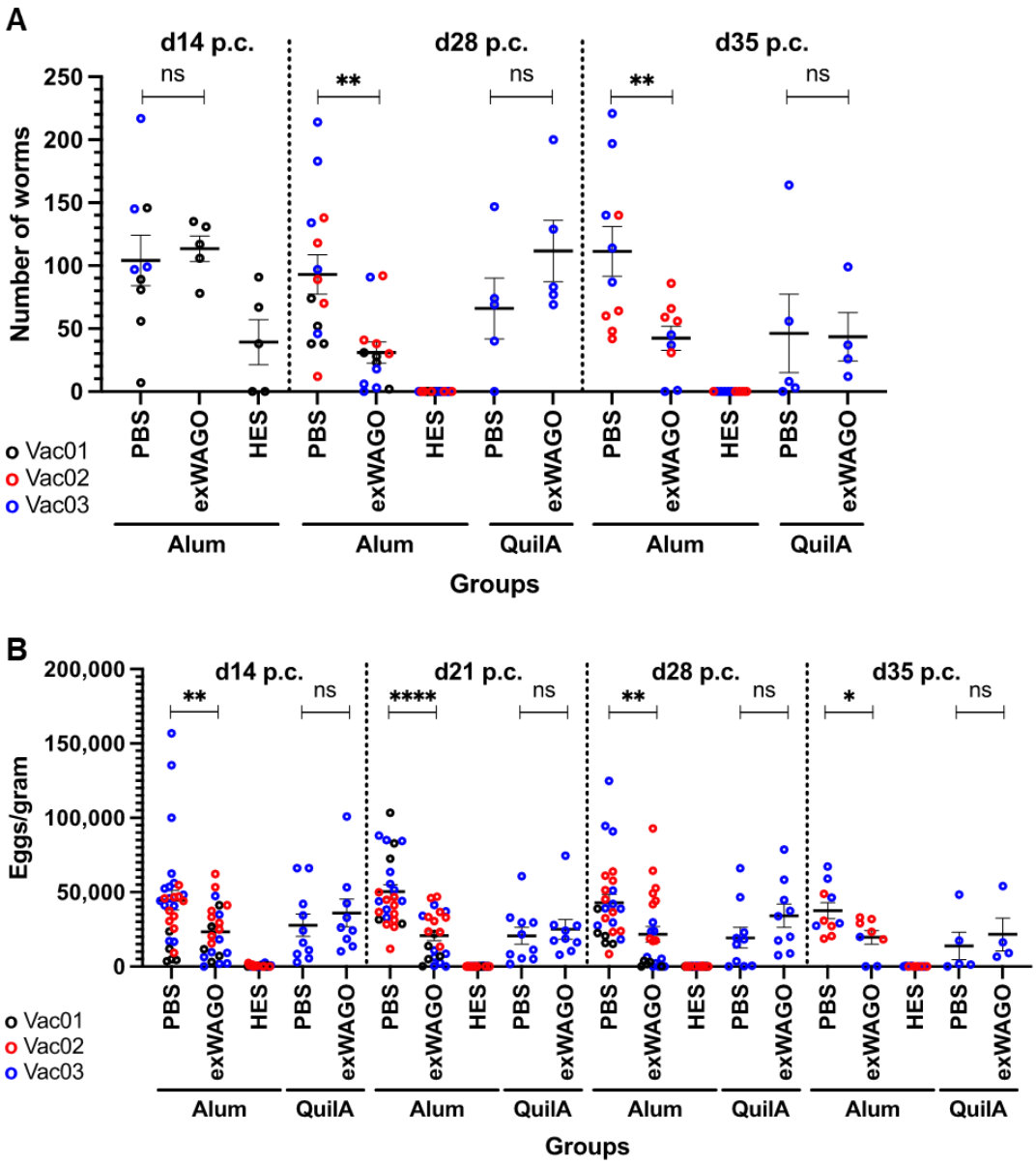


Figure 6.2 | exWAGO vaccination generates partial immunity against infection with *H. bakeri*. A) The number of adult worms counted in the small intestine 14, 28 and/or 35 days post-challenge (p.c.) with 200 L3 stage worms. B) The number of eggs per gram of faeces were counted 14, 21, 28 and/or 35 days post-challenge (p.c.) with 200 L3 stage worms. Data are merged from three independent experiments (Vac01, Vac02 and Vac03) and represent the mean ± S.E.M. (n = 4-29 mice per group). Normally distributed data were analysed using an unpaired t-test with Welch's correction. Data that were not normally distributed were analysed using an unpaired Mann-Whitney test. (ns = not significant, $p > 0.05$; ✱ = $p < 0.05$; ✱✱ = $p < 0.005$; ✱✱✱ = $p < 0.0005$; ✱✱✱✱ = $p < 0.0001$).

6.3 exWAGO vaccination elicits high IgG1 responses but not IgA

To assess the immune response evoked by the vaccination we first examined serum antibody responses by western blot and enzyme-linked immunosorbent assay (ELISA) analysis. Immunisation and challenge with *H. bakeri* evoked IgG responses as shown by western blot analysis using adult worm lysate, HES and recombinant exWAGO probed with pooled sera from exWAGO- and PBS-vaccinated (in alum) mice 28 days p.c. or from uninfected (naïve) mouse serum (Fig. 6.3). More specifically, challenge with *H. bakeri* larvae induces IgG antibodies against proteins found in the worm lysate and HES. This is shown by the strong signals detected in the western blots probed first with serum from exWAGO- and PBS-vaccinated (in alum) mice and then probed with an anti-mouse IgG secondary antibody, compared to uninfected naïve mouse serum (Fig. 6.3). Vaccination with exWAGO particularly induced high IgG levels that recognise exWAGO, as indicated by detection of the recombinant exWAGO protein (107 kDa) by serum from the exWAGO-vaccinated mice compared to the weaker signal detected by serum from the PBS-vaccinated mice and naïve mouse serum (Fig. 6.3). Detection of worm antigens by western blot analysis using naïve serum has previously been reported by McCoy *et al*, (2008).The findings showing that serum from naïve mice detects *H. bakeri* antigens in adult worm lysate and the recombinant exWAGO (faint band for 0.1 µg rexWAGO in Figure 6.3) suggest that naïve mice naturally express antibodies that recognise worm antigens. This perhaps reflects the co-evolution of helminths with their host. Whether the IgG antibodies in naïve serum bind worm antigens specifically or non-specifically remains to be determined.

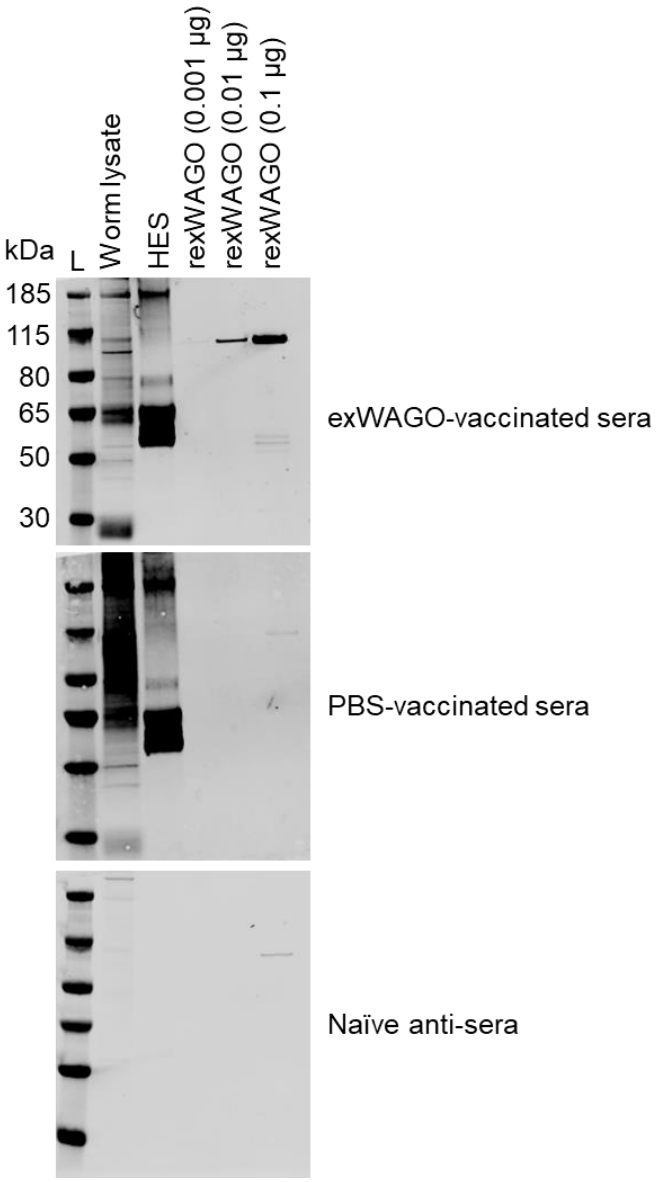


Figure 6.3 | Western blot analysis of total IgG responses for the exWAGO- and PBS-vaccinated groups. Adult *H. bakeri* worm lysate (1 µg), HES (1 µg), and recombinant exWAGO (rexWAGO) protein (0.001-0.1 µg) were run on a 4-12% Bis-Tris SDS gel and the proteins were transferred to a PVDF membrane. Pooled sera from mice from the Vac01 experiment (day 28 p.c., n = 5 mice) or from an unvaccinated and uninfected mouse (naïve serum, n = 1) were used as the primary antibody (1:1,000 dilution in 5% BSA/TBST). The blots were probed with goat anti-mouse IgG fluorescently labelled antibody (1:10,000 dilution in 5% BSA/TBST). L = ladder. These western blots were performed by Ms Yvonne Harcus under the guidance of the author.

Vaccination with recombinant exWAGO and *H. bakeri* challenge also elicits IgG1 antibodies specific to exWAGO as measured by ELISA (Fig. 6.4A). Mice have no detectable exWAGO-specific IgG1 antibodies pre-immunisation and pre-challenge (Fig. 6.4A). Immunisation with recombinant exWAGO (in alum or QuilA adjuvant) elicits high titres of exWAGO-specific IgG1 antibodies while immunisation with PBS (in alum or QuilA adjuvant) prior to challenge does not elicit exWAGO-specific IgG1 antibodies (Fig. 6.4A). Immunisation with HES (in alum) induces generation of exWAGO-specific IgG1 antibodies prior to challenge as expected, since exWAGO is present in HES (Fig. 6.4A). Interestingly, 28 days post-challenge we detect exWAGO-specific IgG1 antibodies in mice immunised with PBS (in alum or QuilA adjuvant) (Fig. 6.4A). The exWAGO-specific signal detected after worm challenge in PBS-vaccinated mice compared to pre-challenge (Fig. 6.4A), indicates that (at least some of) the exWAGO protein is accessible to the immune system and that mice naturally generate exWAGO-specific antibodies during infection. This is also seen by an increase in the exWAGO-specific IgG1 titres post-challenge in mice vaccinated with HES (in alum) (Fig. 6.4A). The exWAGO-specific IgG1 titres post-challenge in mice vaccinated with recombinant exWAGO (in alum or QuilA adjuvant) remain high (Fig. 6.4A).

Interestingly, immunisation of mice with recombinant exWAGO (in alum or QuilA) did not evoke exWAGO-specific IgA antibodies in serum (Fig. 6.4B). However, low titres of exWAGO-specific IgA antibodies are generated following larval challenge as shown by an increase in the signal detected in the PBS-vaccinated mice (in alum or QuilA) (Fig. 6.4B). The increase in the exWAGO-specific IgA signal observed in mice immunised with PBS is similar to the signal observed in the exWAGO-vaccinated mice (in alum or QuilA) and HES-vaccinated mice (in alum) (Fig. 6.4B). These results indicate that infection with *H. bakeri* generates exWAGO-specific IgA antibodies while exWAGO vaccination does not.

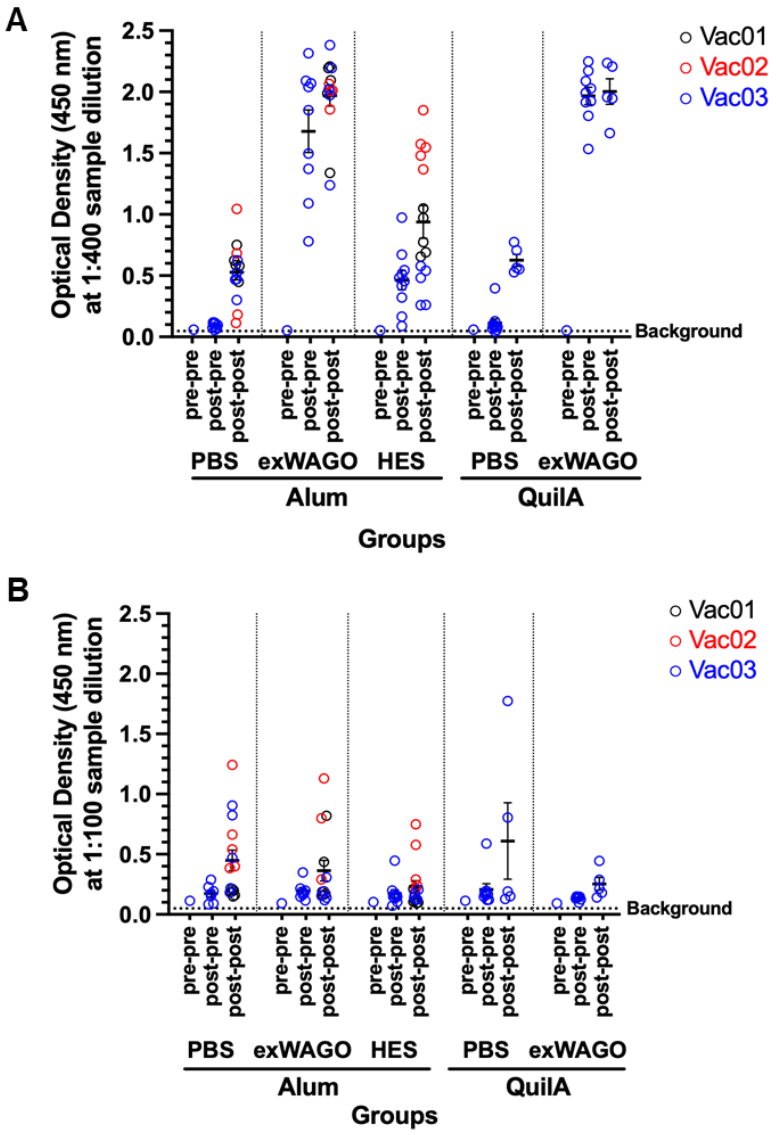


Figure 6.4 | Titres of exWAGO-specific A) IgG1 and B) IgA antibody responses as measured by ELISA from sera of PBS-, exWAGO- and HES-vaccinated mice. The pre-immunisation & pre-challenge data are measured from a pooled sample (n = 10-14 mice). The rest of the data are merged from three independent experiments (Vac01, Vac02 and Vac03) and represent the mean ± S.E.M. (n = 5-14 mice per group). The antibody levels post-infection were measured 28 days post-challenge. The horizontal dotted line represents the average background levels as detected from wells where no sample was added (blank control). ELISAs were performed by Ms Elaine Robertson, Ms Chanel Naar (under the guidance of the author) and the author. Pre-pre = pre-immunisation and pre-challenge; post-pre = post-immunisation and pre-challenge; post-post = post-immunisation and 28 days post-challenge.

6.4 exWAGO vaccination does not affect worm fitness

To investigate if worm fitness is affected by exWAGO vaccination we measured the ATP content of male and female adult worms (Gentile *et al*, 2019). Briefly, worms from mice harvested at day 28 and day 35 p.c. in the Vac03 experiment that had been vaccinated with PBS or recombinant exWAGO (in alum or QuilA) were collected, sexed, and processed to measure the amount of ATP present in each worm by a luminescence assay. Females have greater ATP content than male worms regardless of whether they were obtained from mice vaccinated with PBS or exWAGO and regardless of the adjuvant used in vaccination (Fig. 6.5). The difference in the amount of ATP in males and females could be attributed to the difference of size between the sexes and the presence of eggs in female worms. Moreover, we did not detect a significant change in the ATP content of worms obtained from PBS-vaccinated mice (in alum or QuilA) compared to worms obtained from exWAGO-vaccinated mice (in alum or QuilA) (Fig. 6.5). This indicates that fitness of worms as assessed by measuring the ATP content is not affected by any immune responses triggered by vaccination with recombinant exWAGO, regardless of the adjuvant used.

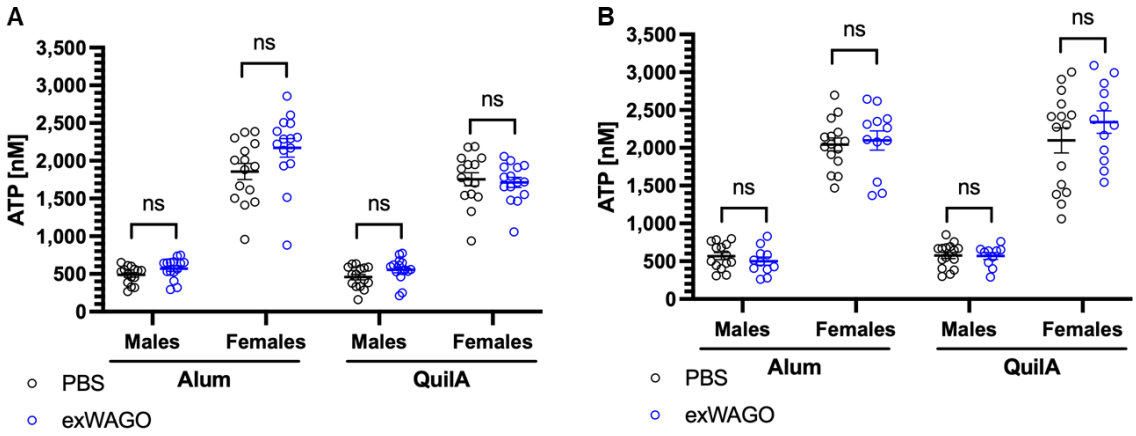


Figure 6.5 | ATP content of worms measured as a proxy of worm fitness A) 28 or B) 35 days post-challenge. Data are from Vac03 experiment and represent the mean ± S.E.M. (n = 3 worms per sex per mouse were obtained if possible, resulting in an n = 10-15 worms per sex per group). Normally distributed data were analysed using an unpaired t-test with Welch's correction. Data that were not normally distributed were analysed using an unpaired Mann-Whitney test. (ns = not significant, $p > 0.05$; ✱ = $p < 0.05$; ✱✱ = $p < 0.005$; ✱✱✱ = $p < 0.0005$; ✱✱✱✱ = $p < 0.0001$).

Discussion

6.5 Summary

In this chapter, we vaccinated mice with exWAGO and challenged them with *H. bakeri* larvae to test whether exWAGO is a potential new vaccine candidate and to determine the impact of blocking exWAGO on parasite survival *in vivo*. Our results indicate that exWAGO vaccination with alum adjuvant confers partial protection against infection when compared to PBS in alum (negative control) as illustrated by a significant reduction in worm and egg burdens (66.7% and 59.0% respectively, calculated as the average across the three experiments) (Fig. 6.2, Table 6.2). Interestingly, vaccination of exWAGO with QuilA adjuvant did not protect against infection (Fig. 6.2, Table 6.2) even though both QuilA and Alum induced high titres of serum exWAGO-specific IgG1 responses (Fig. 6.4A). Moreover, IgA can be actively transported through epithelial cells to mucosal surfaces (Pietrzak *et al*, 2020), hence this antibody is thought to be able to access the worm in the mouse gut lumen where exWAGO is secreted and hypothesised to act. However, exWAGO-specific IgA (measured in serum) was not generated by the exWAGO vaccine (Fig. 6.4B). Additionally, to understand how the vaccine works and whether the immune response generated affects worm fitness, we measured the ATP levels of the worms after vaccination. The exWAGO vaccine did not affect the ATP contents of the worms (Fig. 6.5), suggesting that the immune responses generated by immunisation do not directly attack the parasites but rather may block the function of the secreted exWAGO protein. Further experiments are required to understand the mechanism by which exWAGO vaccination in alum confers protection against *H. bakeri*. The data presented here suggest that exWAGO is important for parasite survival in the gut lumen of the host and that exWAGO could be a novel vaccine candidate against other parasitic nematodes of Clade V that excrete/secrete highly conserved orthologues of exWAGO.

6.6 exWAGO as a vaccine candidate

Analysis of the worm and egg burdens of each vaccine experiment individually (Vac01-03), indicates that vaccination with exWAGO in alum did not confer (significant) protection against larval challenge compared to the PBS-vaccinated mice in alum in the Vac02 experiment (Fig. 6.2, Table 6.2 and data not shown). This contrasts our results obtained for Vac01 and Vac03, hence in this thesis the average of the three experiments was presented. Several variables are different between all three experiments, including the batch of recombinant exWAGO used for

immunisation, the batch of L3 stage larvae used for the infection, and the housing conditions. It was particularly noted that during the Vac02 experiment the animal house hosting the mice was under refurbishment with construction work occurring. To this day it remains undetermined why the exWAGO immunisation did not confer protection in Vac02. Nevertheless, our (averaged) results showing that vaccination with a single protein can reduce parasite survival and engender partial protection against the gastrointestinal nematode *H. bakeri* are encouraging. To our knowledge, this is the first study reporting protection to *H. bakeri* challenge by vaccination using a single recombinant protein (other studies use a cocktail of antigens against *H. bakeri* challenge). At the time of writing this thesis, use of an Argonaute protein as a vaccine candidate has not been previously reported, although RNA-binding proteins have been shown to be vaccine targets in viral infections (Chukwudozie *et al*, 2020) and cancer research (Wang *et al*, 2003; Silva *et al*, 2020).

In this thesis, we assessed the possibility of exWAGO being a putative vaccine candidate against a bolus *H. bakeri* infection (infection with a single, unnaturally large dose of larvae). The bolus infection, however, does not simulate field conditions where humans, livestock and wild animals are chronically and repeatedly infected with gastrointestinal nematodes (Glover *et al*, 2019). Hence, to assess the efficacy of exWAGO as a vaccine candidate in a more natural infection, we could immunise mice with recombinant exWAGO and challenge them with a trickle *H. bakeri* infection. The low but frequent doses of larvae given during a trickle infection would simulate better the chronicity of the disease and the re-infections that occur naturally (Glover *et al*, 2019).

Additionally, it is of interest to test whether exWAGO vaccination could be used to generate immunity against other parasitic gastrointestinal nematodes of Clade V that express highly conserved orthologues of exWAGO (70-80% amino acid identity), such as the human-infective *N. americanus* and the ruminant-infective *T. circumcincta* and *H. contortus* (Chow *et al*, 2019) (Table 1.1). A similar concept to this idea has been tested with vaccination of recombinant versions of the *N. americanus* proteins Na-APR-1 and Na-GST-1 followed by infection with *N. brasiliensis* (Bouchery *et al*, 2018). *N. brasiliensis* expresses and secretes a highly conserved homologue of Na-APR-1 (83% amino acid identity) and two conserved homologues of Na-GST-1 (> 55% amino acid identity) (Bouchery *et al*, 2018). Immunisation of mice with either Na-APR-1 or Na-GST-1 conferred protection against infection with *N. brasiliensis*

(Bouchery *et al*, 2018). If exWAGO can successfully generate protection against other Clade V parasitic nematodes that infect humans and ruminants, it would be important to understand if the vaccine establishes long-term immunity or if frequent vaccine boosts would be required, as seen for the animal vaccine against *H. contortus* currently in use (Drurey *et al*, 2020). Moreover, exWAGO could also be trialled as part of a cocktail of antigens in a vaccine intervention, where immunisation would target co-infections in humans and animals. This is a concept similar to that of the pan-anthelmintic vaccine discussed earlier (Zhan *et al*, 2014), and the licensed DTaP/IPV/Hib/HepB vaccine for humans which is a 6-in-1 combination vaccine providing prophylaxis to various viral and bacterial infections (diphtheria, tetanus, polio, pertussis, *Haemophilus influenzae* type b and hepatitis B) (Public Health Agency, 2017).

Besides efficacy, safety is another crucial factor that needs to be addressed when developing any prophylactic or therapeutic drug. At the current stage of the project, we need to gather more information about the immune responses elicited by exWAGO vaccination in animals. This is discussed further below. However, any side effects induced by the vaccine should be monitored in the future, especially when testing in bigger animals and humans. One safety-related measurement that could be obtained with the current samples we have is examining the levels of serum IgE induced by the exWAGO vaccine. IgE is associated with allergy and has been linked to anaphylaxis (a potentially life-threatening allergic reaction) in children vaccinated against influenza (Nagao *et al*, 2016). Anaphylaxis is generally uncommon after vaccination but when it occurs it is rarely caused by the vaccine antigen itself (McNeil & DeStefano, 2018). It is usually caused by other substances in the vaccine, such as preservatives, egg proteins and adjuvants (McNeil & DeStefano, 2018). Even though allergic reactions to immunisation are rare, it has been observed that vaccination with helminth antigens can trigger allergic responses (Zhan *et al*, 2014). For example, vaccination of rabbits with extracts from the nematode *Ascaris lumbricoides* elicited high titres of IgE antibodies which were cross-reactive to house dust mite antigens, hence inducing allergy to dust mites (Nakazawa *et al*, 2013). Additionally, the phase I human clinical trial with recombinant Na-ASP-2, a candidate antigen against the human hookworm *N. americanus* which is secreted by infective larvae, revealed that people living in hookworm-endemic areas acquire IgE antibody responses against Na-ASP-2 pre-vaccination (Diemert *et al*, 2012). This led to urticarial reactions after vaccination with (a recombinant version of) the hookworm antigen in 3 out of 7 participants

(Diemert *et al*, 2012). These observations prompt us to question whether populations (human and animal) endemic for gastrointestinal nematodes also acquire IgE antibodies against exWAGO orthologues, as this would raise safety concerns for using exWAGO as a vaccine candidate.

6.7 A model of how exWAGO vaccine works

Immunisation of exWAGO in alum adjuvant generates exWAGO-specific IgG1 antibodies and confers partial protection against *H. bakeri* infection. However, since vaccination with exWAGO in QuilA adjuvant did not generate immunity to larval challenge but generated exWAGO-specific IgG1 titre, comparable to titres generated by vaccination with exWAGO in alum adjuvant (Fig. 6.2 and 6.4A), we infer that immune responses other than just IgG1 antibodies, for example cellular responses, are involved in the partial protection observed (Fig. 6.6). The contrasting results we obtained here with Imject alum and QuilA adjuvants, highlight the importance of selecting appropriate adjuvants for vaccine development, as adjuvants can affect vaccine efficacy (Sarkar *et al*, 2019). For example, a group investigating the use of the protein ASP1 (activation-associated secreted protein 1) as a vaccine in cattle against the gastrointestinal nematode *Ostertagia ostertagi* in QuilA or $Al(OH)_3$ (aluminum hydroxide) adjuvants found that cattle immunised with ASP1 in QuilA significantly exhibited reduced egg burdens after infection with *O. ostertagi*, while no reduction was observed after vaccination with ASP1 in $Al(OH)_3$ (González-Hernández *et al*, 2016).

How does the exWAGO vaccine confer partial protection against *H. bakeri* larval challenge? Since the ATP content of worms (which relates to their metabolic activity) remains unaltered during vaccination (Fig. 6.5), we hypothesise that the immune responses elicited by vaccination interfere with the function of the excreted/secreted forms of exWAGO rather than affecting the worms directly. This hypothesis is based on the fact that as an Argonaute protein, exWAGO is not expected to be on the surface or be membrane-associated, unlike the *S. mansoni* Sm-TSP-2 tetraspanin membrane protein and the *N. americanus* Na-APR-1 gut-anchored protease for example which induce antibodies that can bind the tegument and gut, respectively (Tran *et al*, 2006; Loukas *et al*, 2005*).* Other types of vaccines such as vaccination with HES or EVs is expected to affect *H. bakeri* worms directly and also block function of the ES products. This is because the immune responses generated after HES/EV vaccination would recognise antigens that are accessible on the worm surface, openings, and intestine,

since EVs and ES products can derive from the tegument, mouth, gut, and anal pore of worms (Drurey *et al*, 2020). Supporting this hypothesis are preliminary data from our lab which show that the ATP content of female worms is significantly reduced after vaccination of mice with *H. bakeri* EVs (at day 14 p.c.) (Ruby White, unpublished). This could imply that fecundity (i.e. egg production) of female worms is negatively affected by EV vaccination. In the future we could test whether antibodies induced by exWAGO vaccination pre-challenge can bind worms *in vitro*, compared to antibodies induced by PBS vaccination pre-challenge.

Furthermore, as exWAGO is found both encapsulated in EVs and in a vesicle-free form, we predict that mainly the non-vesicular form of exWAGO will be accessible to vaccine-elicited immune responses (such as neutralising antibodies) rather than the vesicular form of exWAGO which would be protected by the lipid bilayer of the EVs (Fig. 6.6). It is worth noting however, that EVs are expected to be taken up and processed by antigen-presenting cells (APCs) (Drurey *et al*, 2020), although this has not been proven. Processing of EV cargo could thus lead to peptides of the vesicular exWAGO presented on the surface of APCs, hence we cannot eliminate the idea that a response against the vesicular exWAGO would be initiated too after the host is challenged with *H. bakeri*. To continue with, whether the anti-exWAGO antibodies generated by vaccination pre-challenge attack the worm directly and/or whether these antibodies neutralise (mainly the non-vesicular) exWAGO is important to understand. If the latter stands then the anti-exWAGO antibodies could be used *in vitro* to neutralise/block the function of non-vesicular exWAGO, which is thought to suppress gene expression by associating with the 5’PPP secondary siRNAs (Chapter 3) that may direct it to host gene targets. This could be achieved by incubating EV-depleted HES with mouse cells or intestinal organoids in the presence or absence of the anti-exWAGO antibodies and investigating the effect on host gene expression changes.

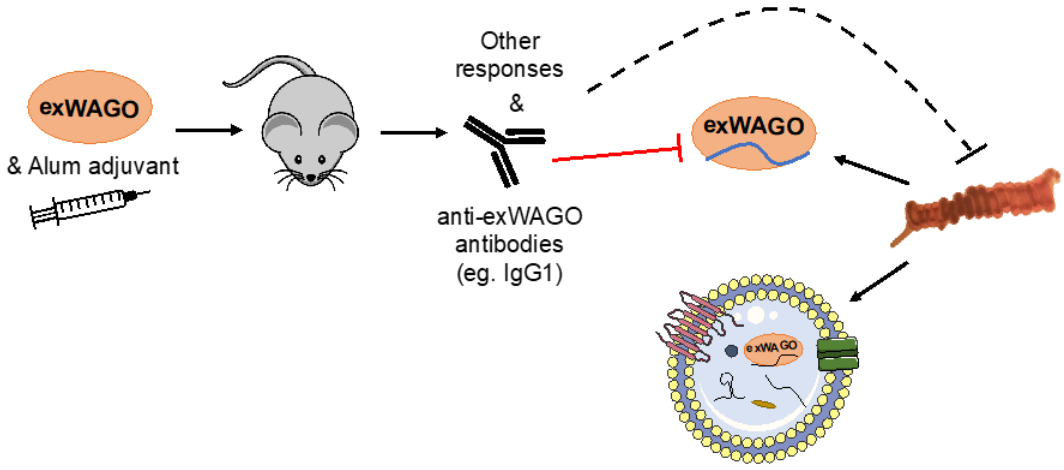


Figure 6.6 | A model of how the exWAGO vaccine is hypothesised to confer protection against *H. bakeri* infection. The exWAGO:alum vaccine induces exWAGO-specific antibodies in mice and presumably cellular responses which we have not examined yet. These immune responses lead to a reduction in the number of worms and eggs indirectly, presumably by blocking the function of the non-vesicular form of exWAGO mainly. The vesicular form of exWAGO is assumed to be protected by the lipid bilayer of the vesicle and hence be inaccessible to antibodies unless EVs are processed by APCs.

6.8 Future work

Understanding the vaccine-induced immune responses is highly valuable for several reasons. For example, this knowledge can inform us on how to generate successful vaccines for existing diseases where vaccine development has failed (e.g. HIV) and for successful generation of vaccines against novel/emerging diseases (Furman & Davis, 2015; Pollard & Bijker, 2020). Additionally, immunological data from vaccination trials can be utilised to predict vaccine efficacy prior to efficacy studies, allowing acceleration of vaccine development (Furman & Davis, 2015; Dudášová *et al*, 2021). Hence, future work is required to better characterise all of the immune responses elicited by the exWAGO vaccination and understand why exWAGO vaccination with QuilA adjuvant did not generate protection in contrast to alum adjuvant. QuilA is hypothesised to shift responses to type 1 immunity, while alum is thought to induce type 2 immune responses (Kensil, 2006). While type 2 immune responses are appreciated as important effectors against helminths, understanding the differences in the immune responses triggered by the two adjuvants could shed

light on which particular immune mechanism(s) must be evoked by the vaccine to control helminths, as this remains unclear (Perera & Ndao, 2021). Whether QuilA has skewed responses to type 1 immunity compared to alum could easily be tested with the current samples we have. This can be done by measuring serum IgG2a in the exWAGO-vaccinated mice with alum and QuilA. As IgG2a is a marker of type 1 immune responses, we would expect to detect high levels of IgG2a in mice vaccinated with QuilA adjuvant and negligible levels in the alum vaccinated mice if this hypothesis stands.

To test whether humoral (antibody-mediated) and/or cellular responses in exWAGO-vaccinated mice (in alum) are required for expulsion of *H. bakeri* adult worms, a serum transfer experiment could be performed. This requires transfer of antibodies purified from serum obtained from exWAGO-vaccinated mice (in alum) into unvaccinated mice, followed by infection with *H. bakeri* larvae and monitoring worm and egg burdens. Such an experiment has been performed by Hewitson *et al*, (2015) who showed that the sterile immunity elicited by vaccination with total HES in alum against *H. bakeri* is only partially attributed to IgG1 antibodies following serum transfer from HES-vaccinated mice into naïve animals. This suggests that HES protection is also mediated by cellular responses (Hewitson *et al*, 2015). The importance of cellular responses in conferring potent immunity against infection via vaccination of mice with *H. bakeri* EVs in alum was also shown by our lab in collaboration with the Maizels lab. Mice lacking ST2 ($ST2^{-/-}$), the IL-33 receptor, were vaccinated with EVs and their immune responses were compared to those of wild type mice vaccinated with EVs. The lack of ST2 impairs activation of ILC2s upon *H. bakeri* infection and results in a reduction in the number of ILC2s, macrophages and T helper cells (Coakley *et al*, 2017). Expression of proteins like Ym1 and RELMα by alternatively activated macrophages is also greatly decreased in ST2-deficient mice (Coakley *et al*, 2017). Although EV vaccination of $ST2^{-/-}$ mice generated comparable levels of antibodies (IgG1, IgA and IgM), only wild type mice had strong protection against the infection in contrast to $ST2^{-/-}$ mice (Coakley *et al*, 2017). These data suggest that EV protection is mainly mediated via cellular responses (Coakley *et al*, 2017). The aforementioned examples highlight the importance of cell-mediated immunity, hence it would be beneficial to study cellular responses in the future in mice vaccinated with exWAGO.

Chapter 7: Discussion

Gastrointestinal worms cause chronic infections and have the ability to manipulate the host immune system by releasing hundreds of immunomodulatory proteins and various other molecules such as sRNAs, some of which are encapsulated in EVs (Maizels *et al*, 2018; Britton *et al*, 2020; Drurey & Maizels, 2021). Our group identified the presence of an extracellular Argonaute (AGO) protein, termed exWAGO, in the EVs and the vesicle-free fraction of the excretory/secretory products of the model organism for gastrointestinal nematodes, *H. bakeri* (Buck *et al*, 2014). *H. bakeri* EVs carry miRNAs and secondary siRNAs, and are internalised by host cells *in vitro* (Buck *et al*, 2014; Coakley *et al*, 2017; Chow *et al*, 2019). Internalisation of EVs causes suppression of genes coding for proteins involved in immunity and inflammation (Buck *et al*, 2014). These data led to the hypothesis that exWAGO might be involved in manipulation of host gene expression via cross-species RNA interference, by associating with extracellular sRNAs. The hypothesis is based on the fact that AGO proteins usually regulate gene expression by binding sRNAs that guide AGO to its target gene to silence gene expression (Höck & Meister, 2008). Hence, this project set out to determine the function of exWAGO with a focus on its putative role inside the host. The work presented here encompasses characterisation of four main aspects of exWAGO biology: (1) identification of the sRNA guides it associates with, (2) developing approaches to detect the (host RNA) targets exWAGO is being directed to, (3) identification of the proteins it interacts with, and (4) testing whether exWAGO could be a vaccine candidate.

7.1 The sRNA guides that associate with exWAGO

exWAGO is classed as a worm-specific Argonaute protein (WAGO), a clade of AGO proteins unique to nematodes (Chow *et al*, 2019). The WAGO proteins are by far best studied in the free-living nematode species *C. elegans*, compared to parasitic nematode species. *C. elegans* expresses at least 13 WAGO proteins (Gu *et al*, 2009). WAGOs are known for their ability to bind sRNAs termed secondary siRNAs that are generated by RNA-dependent RNA polymerases (RdRPs) and have a 5'PPP moiety at their end, hence they are referred to as "Secondary AGOs". WAGOs are therefore involved in RNA interference pathways where there is amplification of the silencing trigger (Gu *et al*, 2009; Ketting & Cochella, 2020). In Chapter 3, I aimed to determine whether exWAGO directly binds 5'PPP secondary siRNAs inside and outside EVs. My data demonstrate that both the vesicular and non-vesicular exWAGO proteins

bind 5'PPP secondary siRNAs that are 22-23 nt in size and start with G. miRNAs, Y-RNAs, tRNAs and rRNAs are not bound by exWAGO even though they are also found in the ES products of *H. bakeri*. Identification of the 5'PPP sRNAs associated with exWAGO is consistent with the fact that exWAGO is a WAGO protein and an orthologue of the Secondary AGOs SAGO-1, SAGO-2 and PPW-1 which also bind 5'PPP secondary siRNAs (Yigit *et al*, 2006; Chow *et al*, 2019). These results prompted us to ask whether the vesicular and non-vesicular exWAGO bind the same 5'PPP secondary siRNAs, as this would imply that they target the same genes. Comparison of the sRNA guides associated with the two extracellular forms of exWAGO illustrates some overlap in the sequences bound but the relative abundance of these sequences varies. To our surprise, the vesicular and non-vesicular exWAGO also bind sequences unique to each exWAGO form. Overall, this suggests that the vesicular and non-vesicular exWAGO might regulate different genes but also some of the same genes but to a different extent. A similar concept to what we uncovered in this thesis with the vesicular and non-vesicular exWAGO-associated sRNAs being different was shown by Arroyo *et al*, (2011). Using qRT-PCR miRNA profiling arrays, Arroyo *et al*, (2011) showed that human plasma EVs and EV-depleted plasma carry different miRNA populations. For example, miR-16 and miR-92a are found highly enriched in the EV-depleted plasma whereas miR-940 is enriched in plasma EVs (Arroyo *et al*, 2011). Some miRNAs such as miR-142-3p and miR-150 are found both in EVs and the EV-depleted fraction, whereas some are found exclusively in each fraction (Arroyo *et al*, 2011). The authors tested some of the miRNAs that were enriched in EV-depleted plasma and showed that these are bound by human AGO2 (hAGO2), which was found to be unencapsulated in plasma (Arroyo *et al*, 2011). The authors suggest that the different miRNA populations observed in EVs or associated with AGO2 outside of EVs could originate from different cell types and represent a cell type-specific expression pattern and/or export mechanism (Arroyo *et al*, 2011). In addition to these hypotheses, our work also suggests different biogenesis/loading pathways for the exWAGO-siRNAs that are exported in EVs compared to those that are exported associated with the non-vesicular form of exWAGO. The finding that miRNAs remain preferentially unbound by exWAGO in the ES products suggests that these do not act via exWAGO. Instead, miRNAs may function by hijacking the host AGO protein, mAGO2 (Donnelly & Tran, 2021).

7.2 Putative host targets of exWAGO and its host protein interactors

In Chapter 4, I employed a biochemical technique to identify candidate genes that might be targeted by exWAGO:sRNA complexes *in vivo*. This technique (CLASH) involved immunoprecipitation of exWAGO:sRNA complexes from *H. bakeri*-infected mouse gut tissue, ligation of the guide sRNA to the target mRNA, and sequencing. The guide:target ligation step should generate chimeric reads, however, in this thesis we only analysed non-chimeric reads. By carrying out parallel studies on the host AGO protein, mAGO2, we show that the technique used is valid for identification of target genes from gut tissue outside of chimeric reads. Consistent with the literature, mAGO2 associates with 3' UTR and CDS regions of mRNA targets, where miRNA target sites are located (Bartel, 2009; Leung *et al*, 2011). On the other hand, the exWAGO data suggest that exWAGO associates with intronic regions of transcript genes. These preliminary data should be viewed with some caution. Identification of exWAGO targets is based on internalisation of exWAGO:guide complexes by mouse host cells in the gut. Detection of introns as targets of exWAGO might not be real based on the fact that introns are the most abundant type of gene location compared to other genomic loci (Mouse Genome Sequencing Consortium, 2002). What we are detecting might be background noise due to limited amount of exWAGO:guide complexes that are internalised by host cells compared to the amount of exWAGO:guide complexes not internalised by the host cells. This is based on the data suggesting that the major source of exWAGO:guide complexes in the gut tissue samples is derived from the parasite itself. Additionally, examination of the reads mapping to the intronic regions detected as exWAGO targets indicates the majority of these have persistent mismatches after mapping to the mouse genome (with up to 1 mismatch) and start with a G. These characteristics are reminiscents of the features the parasite 5'PPP secondary siRNAs possess. Hence, these sequences could be parasite guides which were not removed during data analysis (at the step of removing perfectly mapping *H. bakeri* reads) but that are highly complementary to the mouse genome.

Nevertheless, identification of introns as the site targeted by exWAGO could be consistent with data presented by Sarshad *et al*, (2018) who report that in the nucleus of stem cells, mammalian AGO2 proteins target intronic regions. The authors speculate that AGO2 proteins could be involved in co-transcriptional gene silencing in the nucleus (Sarshad *et al*, 2018). In agreement with the mass spectrometry data obtained in Chapter 5 and identification of the nuclear protein SFPQ in these data, we

cannot disregard the hypothesis that exWAGO might operate in the nucleus of host cells and that it might be involved in co-transcriptional gene silencing, hence targeting intronic regions. Overall, our proteomics data suggest that exWAGO interacts with host proteins involved in gene silencing and gene regulation, such as SFPQ, LSM14a and Ubap2l. Further work is required to validate the protein interactions identified and also to determine the subcellular localisation of exWAGO, as this could further shed light on how exWAGO functions.

7.3 Proposed model for the mode of action of exWAGO and sRNAs

Based on data previously published by our lab and the data gathered during this thesis our aim was to build a model for how exWAGO and the sRNAs secreted by *H. bakeri* may function in the gastrointestinal tract of mice to silence host gene expression (Fig. 7.1). The exWAGO protein bound to secondary siRNAs is found both unencapsulated and encapsulated in EVs. Published data shows that the *H. bakeri* EVs are internalised by host cells (Buck *et al*, 2014; Coakley *et al*, 2017) hence we assume that the vesicular exWAGO would enter host cells via EV uptake. EV uptake could occur via various pathways including receptor-ligand interactions, phagocytosis/endocytosis and direct membrane fusion (Drurey *et al*, 2020). Evidence on how non-vesicular AGO proteins traffic into recipient cells is lacking. However, a study showed that human AGO2 ('empty' or loaded with guide) can be internalised into recipient cells via the receptor neuropilin-1 (Prud'homme *et al*, 2016). Internalisation of the non-vesicular exWAGO might be facilitated by a receptor protein which we did not identify in this thesis by proteomic analysis. In our data we found the adaptor protein of the clathrin-mediated endocytosis pathway Dab2 (Finkielstein & Capelluto, 2016) and we hypothesise that it is involved in AGO trafficking. Dab2 is known to be involved in protein internalisation and my results suggest that it interacts with exWAGO and mAGO2. We speculate that Dab2 might interact with a receptor (not found here) that binds exWAGO (and mAGO2) to facilitate internalisation of non-vesicular exWAGO. This is discussed further in the next section (Section 7.4).

Moreover, the host protein interactors identified by mass spectrometry such as SFPQ and Ubap2l, suggest a model where exWAGO operates in complex with host proteins involved in gene silencing and gene regulation. SFPQ is a nuclear protein which was found to interact with hAGO2 and mAGO2 and is involved in post-transcriptional gene silencing (Bottini *et al*, 2017). Identification of SFPQ as a putative interactor of exWAGO suggests that exWAGO could act in the nucleus. exWAGO could also act

in specialised compartments found in the cytoplasm such as stress granules. This is based on identification of Ubap2l, a protein involved in the formation of stress granules which are thought to be involved in gene regulation (Leung *et al*, 2006; Buchan & Parker, 2009; Cirillo *et al*, 2020), as a putative interactor of exWAGO. Although we do not yet know how exactly exWAGO would modulate host gene silencing/regulation, in this thesis I showed that exWAGO is not able to slice targets.

Our preliminary data on target identification suggest that exWAGO is directed mainly to intronic regions of transcripts, implying that exWAGO could be involved in co-transcriptional gene silencing, but as discussed before these results might not be true. Based on my sRNA sequencing data, even though miRNAs are also secreted by *H. bakeri* inside and outside EVs, exWAGO does not associate with these *in vivo*. Along with the data that indicates that mAGO2 does not bind 5'PPP with high affinity in contrast to exWAGO, it is possible that *H. bakeri* miRNAs act independent of exWAGO and instead might function to hijack the host mAGO2 protein. This is currently under investigation. Taken together, I propose that *H. bakeri* has evolved two mechanisms for manipulating host gene expression using sRNAs via cross-species RNA interference: an exWAGO-dependent mechanism which involves 5'PPP secondary siRNAs and an exWAGO-independent mechanism which involves miRNAs and potentially other types of sRNAs.

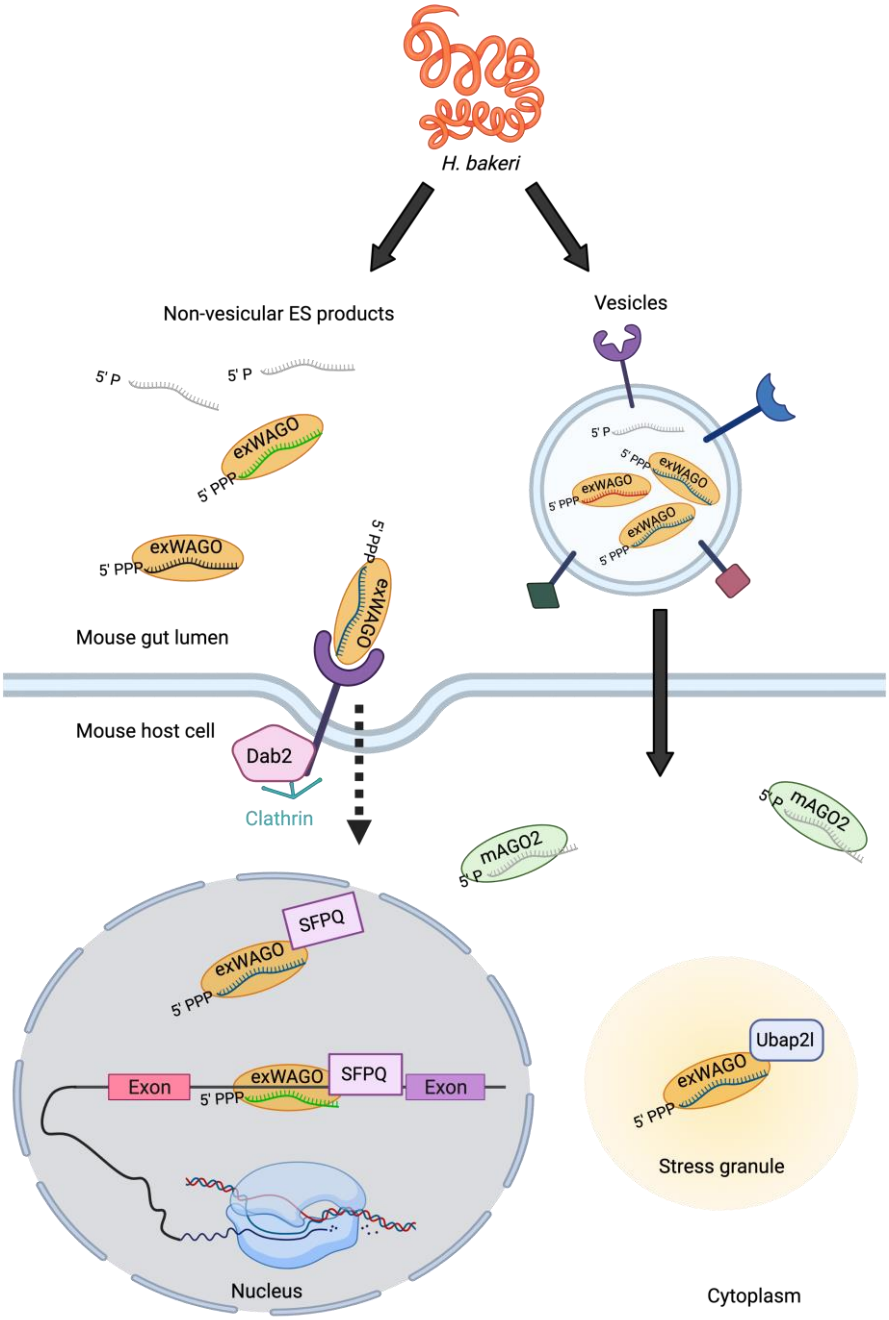


Figure 7.1 | A cartoon of hypothetical entry and functional pathways of exWAGO inside mouse host cells. exWAGO is secreted by *H. bakeri* inside EVs or outside EVs. EVs can be internalised by mouse host cells *in vitro* but it remains unknown if and how the non-vesicular exWAGO is internalised. The data gathered and presented in this thesis suggest that the non-vesicular exWAGO might enter host cells by interacting with the clathrin-associated adaptor protein Dab2. It is hypothesised that a membrane receptor facilitates the interaction of Dab2 and exWAGO. Once internalised, exWAGO functions in the cytoplasm and interacts with the stress granule-associated protein Ubap2l. exWAGO may also function in the nucleus based on identification of the nuclear protein interactor SFPQ. exWAGO might be involved in manipulation of host gene expression by targeting the intronic regions of targets. Figure created using BioRender.com.

7.4 RNA-mediated communication

In the last 15 years or so, there has been a great interest in extracellular RNA (exRNA) (Claycomb *et al*, 2017), with ground-breaking research reporting (1) the presence of sRNAs in several types of body fluids such as serum and urine (Chen *et al*, 2008), (2) the presence of sRNAs and mRNAs inside mammalian EVs, and (3) their transfer to recipient cells *in vitro* (Valadi *et al*, 2007). These findings suggest that RNA is a functional molecule involved in direct communication and dissemination of information between cells and between organisms of different species (Claycomb *et al*, 2017). Even though the first documentation showing that RNA is released outside of cells was made 50 years ago (Kolodny *et al*, 1972), the concept of RNA-mediated communication has been difficult to accept as the mechanistic basis of RNA export/import remain poorly understood to this day (Claycomb *et al*, 2017). Understanding the mechanism of RNA communication is important as it will not only help uncover new interactions between organisms, but it is also relevant to drug discovery and RNA delivery.

The mechanism of RNA-mediated communication is thought to have several regulatory steps. For example, how is RNA exported? It is generally well accepted that RNAs can be exported via EVs. Indeed, studies have shown that RNA-binding proteins such as AGO proteins, RNA helicases, and annexins are involved in loading sRNAs in EVs generated by *Arabidopsis* (He *et al*, 2021). exRNA is also found unencapsulated. The existence of unencapsulated RNAs however, raises the question of how these molecules are stable since the absence of the EV lipid bilayer makes them an easy target for degradation? Using a Proteinase K assay, Arroyo *et al*, (2011) showed that the non-vesicular hAGO2 found in serum plasma protects miRNAs from degradation. This means that RNA-binding proteins are (at least part of) the answer regarding the stability of extracellular sRNAs. In this way, we can infer that in communication mediated by sRNAs released by *H. bakeri*, at least one role of exWAGO could be to protect vesicle-free secondary siRNAs from degradation. Another question related to RNA export, is whether there is selectivity in the secretion process. Work in *Arabidopsis* by He *et al*, (2021) indicates that the AGO1 and RNA helicases RH11 and RH37 selectively bind to sRNAs that are enriched in EVs compared to non-EV-associated sequences, proposing that selectivity can be established by RNA-binding proteins. How exactly these proteins are involved in the sorting of sRNAs into EVs and how some proteins such as AGOs are also exported remains unknown (Pottash *et al*, 2019).

Following (selective) export of exRNA, we need to consider how the RNA gets delivered to the 'right' target cell(s) and how it is internalised. While several mechanisms of internalisation have been proposed and validated regarding EV uptake by recipient cells including receptor-mediated endocytosis and membrane fusion, how non-vesicular exRNA is internalised remains largely unknown (Kim *et al*, 2020). mRNA is reported to be transferred via membrane nanotube-like projections between mammalian cells that are in direct contact in *in vitro* co-culture experiments (Haimovich *et al*, 2017). Moreover, as mentioned above, a study by Prud'homme *et al*, (2016) reports that the cell-membrane receptor neuropilin-1 mediates internalisation of miRNAs and also extracellular miRNA-loaded hAGO2 complexes. Neuropilin-1 interacts with the endocytic adaptor protein synectin (GIPC1) which in turn interacts with the motor myosin Myo6 to mediate endocytosis of another cell-membrane protein which is involved in adhesion to the extracellular matrix, the α5β1 integrin in its active form (Valdembri *et al*, 2009). Interestingly, Myo6 also interacts with the adaptor protein Dab2 (Inoue *et al*, 2002), which was found to interact with exWAGO and mAGO2 in mouse intestinal tissue. Dab2 is involved in endocytosis of inactive integrin proteins (de Franceschi *et al*, 2015). It has been hypothesised that there could be a link between Dab2 and GIPC1 where they act in conjunction to facilitate internalisation of neuropilin-1 (Prud'homme *et al*, 2016). However, the published literature currently supporting this is limited. Nevertheless, the hypothesis that Dab2 interacts with a cell-membrane receptor that binds 'naked' sRNAs and/or AGO-bound sRNAs, and the idea that this interaction facilitates internalisation of the associated ligands inside host cells could be a plausible scenario for trafficking of exWAGO and mAGO2. One important future direction therefore is to identify the cell-surface receptor protein (such as neuropilin-1) that might explain how AGO proteins are internalised into recipient cells via Dab2 (Fig. 7.1).

With regards to how exRNA is transferred to the 'right' cells, it is assumed that specificity will be determined according to the means of internalisation. This is linked to how exRNA is exported (inside EVs, bound to an RNA-binding protein like AGO or carrier-free). Several studies have identified ligands or surface proteins on EVs carrying exRNA (such as tetraspanins) to interact with receptors on recipient cells (Kim *et al*, 2020). For example, in the helminth field, Kuipers *et al*, (2020) illustrated that *S. mansoni* EVs carrying glycans on their surface are internalised by dendritic cells by interacting with the dendritic cell receptor DC-SIGN (CD209). Hence, if exRNA is internalised via a receptor, it would be expected that only cells expressing

that receptor would be the candidate recipient cells. Currently it is not known whether AGO proteins have a mechanism for targeting specific cells or tissues (Pottash *et al*, 2019). Furthermore, in the context where an extracellular pathogen (such as a gastrointestinal worm) is the 'sender' of the exRNA messages, it might be safe to assume that secreted molecules including exRNA would target cells in the local environment. Nonetheless, potential distal effects of the secreted molecules should not be disregarded. This phenomenon has been observed by the McSorley lab who report abrogated detection of the IL-33 receptor (ST2) in the lung ILC2 cells in mice infected with *H. bakeri* (7 days post-infection) (Vacca *et al*, 2020). This is consistent with the abilities of the extracellular *H. bakeri* protein HpBARI to bind the ST2 receptor and block ligation of IL-33 (Vacca *et al*, 2020). These results illustrate the potent effects of gastrointestinal nematode ES products on sites distal of infection, even though *H. bakeri* is physically constrained to the small intestine.

For understanding RNA-based communication, it is of course crucial to understand whether and how the exRNA is internalised inside recipient cells is functional. This also links to whether the concentration of the RNA delivered into recipient cells is adequate to exert functional effects (Claycomb *et al*, 2017). Whether exRNA has a biological function is still debatable (Gruner & McManus, 2021). This is mainly because many studies use non-physiological concentrations of exRNA or because results could be explained by other confounding variables (Gruner & McManus, 2021). Several examples of studies examining the function of mammalian exRNA and their limitations are discussed by Gruner & McManus (2021). Quantitative analysis of miRNAs in mammalian plasma-derived EVs from patients with prostate cancer suggested that examination of 100 EVs is required to detect one miRNA molecule (Chevillet *et al*, 2014). This suggests that the number of miRNAs in EVs is very low to cause a biologically significant effect in the recipient cell, hence Chevillet *et al*, (2014) propose that it is unlikely for EVs to serve as mediators of miRNA-based communication on their own. A study by Albanese *et al*, (2021) suggests that delivery of EV cargo is also inefficient, supported by lack of detection of fusion of EVs with the recipient cell. The authors also failed to detect the functional effects of miRNAs in recipient cells based on luciferase reporter assays (Albanese *et al*, 2021). It is therefore important for us to understand whether the exRNAs that are secreted by *H. bakeri*, encapsulated in EVs or unencapsulated, are functional.

Although we have not quantified how many miRNAs and/or secondary siRNAs are present in *H. bakeri* EVs, we expect that there are at least 3 molecules of secondary siRNAs (bound to exWAGO) per EV. This is based on stoichiometric analysis which indicates that exWAGO is found at 3.4±1.1 copies per EV (Chow *et al*, 2019), and the assumption that most of (if not all) exWAGO molecules would be loaded with secondary siRNAs rather than be 'empty'. It is important to note however that the EVs purified are likely a heterogeneous population, thus even though on average there are 3.4±1.1 copies of exWAGO per EV, in reality some EVs might have no exWAGO whilst others have more than the average number reported. To investigate if *H. bakeri* vesicular sRNAs have a functional effect we need to consider how many EVs are secreted by each worm over a specified timeframe. Unpublished data from our lab estimate that approximately $1x10^7$ EVs are generated per worm during 14 days in culture (Coakley, unpublished). This is probably an underestimation of the EVs generated *in vivo* as the culture conditions used and the purification method (ultracentrifugation) employed can lead to reduction of EV release by the parasite or to loss of EVs respectively (Konoshenko *et al*, 2018). The total number of EVs secreted *in vivo* by *H. bakeri* will depend on the level of infection. Additionally, it is important to remember that a bolus infection in laboratory mice is not reflective of the low worm burdens and re-infections that occur in the wild. Therefore, we need to account for the fact that gastrointestinal worm infections are chronic, when examining the functional effects of exRNA. Moreover, we also need to consider the fact that sRNAs and exWAGO:secondary siRNA complexes are found outside of EVs too, and the functional effect of the non-vesicular sRNAs should be considered. Lastly, it is possible that the hundreds of proteins, thousands of sRNAs, and various other molecules secreted by *H. bakeri*, regardless of whether they are inside EVs or not, individually exert small effects. Some of these small effects might even be undetectable and/or non-quantifiable by the techniques currently available. However, the additive effect of all the molecules acting on the host simultaneously results in the observed immunomodulatory effects of helminths.

7.5 Prophylactic and therapeutic implications

The prevalence of parasitic worms and the devastating impact they have on human health and the economy requires development of prophylactic and therapeutic interventions. The immunogenic and immunomodulatory excretory/secretory products of helminths, which include sRNAs and exWAGO, are tested and (some) are employed for development of vaccines. The number of vaccines currently licensed

against parasitic worms in animals is limited, and no human vaccines are available (Claerebout & Geldhof, 2020; Perera & Ndao, 2021). In Chapter 6, we report that immunisation of mice with exWAGO in alum adjuvant elicits partial protection against challenge with *H. bakeri* larvae. Based on the high degree of conservation between exWAGO and its orthologues expressed by Clade V gastrointestinal nematodes, we propose that exWAGO and/or its orthologues could be a vaccine candidate against such helminths. As previously discussed (Section 6.6), to enhance the efficacy of the vaccine and also tackle co-infections, exWAGO could be trialled as part of a combinatorial vaccine where a cocktail of antigens is used to tackle animal- and human-infective helminths. This would require testing exWAGO vaccination in ruminants and humans prior to use in a pan-anthelmintic vaccine. Moreover, the use of protein epitopes for vaccination rather than the whole protein could be advantageous (eg. improved safety and low production cost) (Topuzoğullari *et al*, 2020). For this, protein epitopes that are conserved among exWAGO orthologues of the Clade V gastrointestinal nematodes and that are exposed on the surface of the protein hence accessible to immune responses could be identified.

Therapeutic interventions against helminths could also be developed based on the extracellular sRNAs secreted by the parasites. For example, RNA sponges could be engineered to sequester helminth exRNAs to inhibit the putative function of sRNA molecules from manipulating host gene expression. Hence, the parasite sRNAs secreted by *H. bakeri* (some of which are bound by exWAGO) outside of EVs detected in Chapter 3, could serve as a tool for testing such an anthelmintic strategy. This approach would also enable us to investigate the function of *H. bakeri* sRNAs inside host cells. A similar concept to this has been used to examine the effect of the sRNAs secreted by the *A. thaliana* pathogen *Hyaloperonospora arabidopsidis*, on the host (Dunker *et al*, 2020). By expressing a short-tandem-target-mimic in Arabidopsis to sequester *H. arabidopsidis* sRNAs, the authors showed that the pathogen's virulence can be reduced (Dunker *et al*, 2020). Additionally, the implications of sequestering host miRNAs from their normal functions should be investigated and considered when designing/developing such an intervention.

Identification of the RdRP involved in the biogenesis route of secondary siRNAs in *H. bakeri* could also help in identification of therapeutic agents. For example, blocking the function of the RdRP could disrupt generation of the secondary siRNAs and hence affect the parasite directly (assuming the secondary siRNAs play a role in intracellular

worm functions), as well as abrogate any exWAGO:sRNA-mediated effects that dysregulate host gene expression. One of the challenges of this intervention is that the drug would need to be delivered and taken up by the parasite without affecting the normal function of host proteins. However, as mammals do not express RdRPs, this should facilitate drug design (Pinzón *et al*, 2019).

The need to control and eliminate helminth infections is highlighted by the severe morbidities and the economic losses these parasites cause. While research has focused on identifying prophylactic and therapeutic interventions to tackle worm infections, more recently the scientific community also investigates the use of helminth products against allergies, based on the inverse correlation between allergy and prevalence of worms (Pearce *et al*, 2000). It is critical to consider a combinatorial approach where we control and eliminate helminth infections as they cause serious health problems and agricultural losses and also explore how helminth ES products dampen type 2 responses and use them to treat allergies and autoimmune diseases (Maruszewska-Cheruiyot *et al*, 2018).This is possible as allergies and autoimmune diseases trigger a similar type 2 immune response to that triggered by worms (Akdis *et al*, 2020). Some of these worm ES molecules have already been discovered and are tested in models of allergy. For example, the *H. bakeri* ES protein TGF-β mimic, TGM1, modulates T-cell responses by inducing expansion of regulatory T-cells and it is thought to have anti-inflammatory effects (Johnston *et al*, 2017; White *et al*, 2021). After expressing bioactive recombinant TGM1 in edible green algae, Smyth *et al*, (2021) orally administered the protein to a mouse model of chemically-induced colitis. Ulcerative colitis is one of the principal types of the autoimmune disease Inflammatory Bowel Disease (IBD), affecting approximately 10 to 20 per 100,000 people per year (Kaur *et al*, 2020). The results showed that oral delivery of the algal expressed TGM1 protects mice from loss of weight due to chemically-induced colitis and ameliorated disease symptoms, suggesting that TGM1 could be employed as an IBD treatment (Smyth *et al*, 2021). The role of exWAGO and its associated sRNAs in manipulation of host genes that control immune responses remains to be elucidated. However, we propose that identification of specific secondary siRNAs involved in suppression of type 2 immune responses could be exploited as an RNA therapeutic solution to allergies and autoimmune diseases. For example, strategies could include delivery of recombinant exWAGO loaded with the appropriate 5’PPP secondary siRNAs or delivery of 5’P siRNA mimics with the appropriate sequence to control expression of genes involved in immunity.

7.6 Future applications – exWAGO as a gene editing and RNA sponge tool

The discovery of prokaryotic AGOs (pAGOs) and their unusual features has led to the proposition that pAGOs could be novel genome-editing tools (Hegge *et al*, 2018). For example, the fact that pAGOs (1) bind small DNA guides rather than sRNAs, (2) can target DNA and (3) have cleavage activity, has illuminated their potential as a putative alternative platform to the CRISPR-Cas system (Hegge *et al*, 2018). Could exWAGO also be employed as a gene editing tool? Moreover, as exWAGO binds 5'PPP sRNAs with high affinity based on our data (Chapter 3), exWAGO could be exploited as a therapeutic and molecular tool at the same time against viruses that generate 5'PPP RNAs. For example, 'empty' exWAGO could be employed as a sponge to bind and sequester 5'PPP viral RNAs. In this way the function of 5'PPP viral sRNAs could be inhibited. Additionally, exWAGO could be engineered such that it possesses a degradation signal which is revealed following conformational change upon binding the viral 5'PPP RNA. The exWAGO:viral RNA complex could then be directed for degradation, limiting in this way immune evasion by viral sRNAs and hence inhibiting viral infection. It is hard, however, to know at the current time if the proposed uses of exWAGO are feasible as we do not yet understand how it functions. Nevertheless, future work uncovering how it operates inside mammalian cells will highlight potential applications of exWAGO.

7.7 Future work and outstanding questions

The field of cross-species RNAi and RNA-mediated communication in helminths is still in its infancy. exWAGO is the first AGO protein to be reported as extracellular in helminths and the second AGO known to be secreted from a parasite. To date, various questions remain regarding the mechanism of action of exWAGO. One of the main questions we wish to answer is whether exWAGO represses host gene expression and how. The work presented in this thesis provides a first mechanistic basis for how exWAGO might operate in cross-species RNAi inside host cells, but further work is required to pin-point the exact mechanism. Identification of the specific guide:target interactions would shed light on the targeting rules between the parasite guides and the host targets. This would inform how other nematodes might manipulate gene expression using secondary siRNAs.

Recent work by the Barr lab shows that *C. elegans* AGO proteins, including PPW-1, are secreted in EVs and these can be transferred from the reproductive tract of male

worms to the vulva of hermaphrodite worms during copulation (Nikonorova *et al*, 2021). These findings hence suggest that PPW-1, an exWAGO orthologue protein, could function in intraspecies communication. It would thus be of interest to understand the functional mechanism of exWAGO in RNAi inside the *H. bakeri* worms and whether it could also mediate communication with other *H. bakeri* worms. This also prompts the question of whether exWAGO facilitates communication between *H. bakeri* and the gut microbiota.

Furthermore, understanding why exWAGO exists in two extracellular forms (vesicular and non-vesicular) and identifying what dictates their route of export from the parasite could shed light on the mechanisms of export of AGO proteins in general. Work in this thesis has led to identification of Dab2 as a potential route of AGO trafficking which should be investigated further. All in all, future work based on exWAGO will add to our model of RNA-mediated communication, it will expand our understanding of cross-species RNAi, and potentially lead to development of anthelmintic interventions and molecular tools, such as gene editing and RNA sponge tools.

References

Adegnika AA, de Vries SG, Zinsou FJ, Honkepehedji YJ, Dejon Agobé JC, Vodonou KG, Bikangui R, Bouyoukou Hounkpatin A, Bache EB, Massinga Loembe M, van Leeuwen R, Molemans M, Kremsner PG, Yazdanbakhsh M, Hotez PJ, Bottazzi ME, Li G, Bethony JM, Diemert DJ, Grobusch MP, et al (2021) Safety and immunogenicity of co-administered hookworm vaccine candidates Na-GST-1 and Na-APR-1 in Gabonese adults: a randomised, controlled, double-blind, phase 1 dose-escalation trial. *Lancet Infect. Dis.* **21:** 275–285

Ait-Ammar A, Kula A, Darcis G, Verdikt R, De Wit S, Gautier V, Mallon PWG, Marcello A, Rohr O & Van Lint C (2020) Current Status of Latency Reversing Agents Facing the Heterogeneity of HIV-1 Cellular and Tissue Reservoirs. *Front. Microbiol.* **10:** 3060

Akdis CA, Arkwright PD, Brüggen MC, Busse W, Gadina M, Guttman-Yassky E, Kabashima K, Mitamura Y, Vian L, Wu J & Palomares O (2020) Type 2 immunity in the skin and lungs. *Allergy* **75:** 1582–1605

Albanese MI, Adam Chen Y-FI, Hü ls C, Gä rtner KI, TagawaID T, Mejias-PerezID E, KepplerID OT, Gö bel C, ZeidlerID R, Shein M, Schü tzID AK & HammerschmidtID W (2021) MicroRNAs are minor constituents of extracellular vesicles that are rarely delivered to target cells. *PLOS Genet.* **17:** e1009951

Alemu A, Bitew ZW & Worku T (2020) Intestinal parasites co-infection among tuberculosis patients in Ethiopia: a systematic review and meta-analysis. *BMC Infect. Dis.* **20:** 1–10

Allaire JM, Crowley SM, Law HT, Chang SY, Ko HJ & Vallance BA (2018) The Intestinal Epithelium: Central Coordinator of Mucosal Immunity. *Trends Immunol.* **39:** 677–696

Allen JE & Sutherland TE (2014) Host protective roles of type 2 immunity: Parasite killing and tissue repair, flip sides of the same coin. *Semin. Immunol.* **26:** 329–340

Alló M, Agirre E, Bessonov S, Bertucci P, Acuña LG, Buggiano V, Bellora N, Singh B, Petrillo E, Blaustein M, Miñana B, Dujardin G, Pozzi B, Pelisch F, Bechara E, Agafonov DE, Srebrow A, Lührmann R, Valcárcel J, Eyras E, et al (2014) Argonaute-1 binds transcriptional enhancers and controls constitutive and alternative splicing in human cells. *PNAS* **111:** 15622–15629

Alló M, Buggiano V, Fededa JP, Petrillo E, Schor I, De La Mata M, Agirre E, Plass M, Eyras E, Abou Elela S, Klinck R, Chabot B & Kornblihtt AR (2009) Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat. Struct. Mol. Biol.* **16:** 717–725

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402

Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schäffer AA & Yu YK (2005) Protein database searches using compositionally adjusted substitution matrices. *FEBS J.* **272:** 5101–5109

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. *Babraham Bioinforma.* Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Angeles-Albores D, Lee RYN, Chan J & Sternberg PW (2018) Two new functions in the WormBase Enrichment Suite. *MicroPublication Biol.* **2018:** 10.17912/W25Q2N

Angeles-Albores D, N. Lee RY, Chan J & Sternberg PW (2016) Tissue enrichment analysis for C. elegans genomics. *BMC Bioinformatics* **17:** 1–10

Aoki K, Moriguchi H, Yoshioka T, Okawa K & Tabara H (2007) In vitro analyses of the production and activity of secondary small interfering RNAs in C. elegans. *EMBO J.* **26:** 5007–5019

Arnold A, Rahman MM, Lee MC, Muehlhaeusser S, Katic I, Gaidatzis D, Hess D, Scheckel C, Wright JE, Stetak A, Boag PR & Ciosk R (2014) Functional characterization of C. elegans Y-box-binding proteins reveals tissue-specific functions and a critical role in the formation of polysomes. *Nucleic Acids Res.* **42:** 13353–13369

Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, Mitchell PS, Bennett CF, Pogosova-Agadjanyan EL, Stirewalt DL, Tait JF & Tewari M (2011) Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc. Natl. Acad. Sci.*

Artis D, Wang ML, Keilbaugh SA, He W, Brenes M, Swain GP, Knight PA, Donaldson DD, Lazar MA, Miller HRP, Schad GA, Scott P & Wu GD (2004) RELMβ/FIZZ2 is a goblet cell-specific immune-effector molecule in the gastrointestinal tract. *Proc. Natl. Acad. Sci. U. S. A.* **101:** 13596

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM & Sherlock G (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25:** 25–29

Ayache J, Bénard M, Ernoult-Lange M, Minshall N, Standart N, Kress M & Weil D (2015) P-body assembly requires DDX6 repression complexes rather than decay or Ataxin2/2L complexes. *Mol. Biol. Cell* **26:** 2579

Babu S & Nutman TB (2019) Immune Responses to Helminth Infection. *Clin. Immunol.***:** 437-447.e1

Ball G, Selkirk ME & Knox DP (2007) The effect of vaccination with a recombinant Nippostrongylus brasiliensis acetylcholinesterase on infection outcome in the rat. *Vaccine* **25:** 3365–3372

Bansemir AD & Sukhdeo MVK (1994) The food resource of adult Heligmosomoides polygyrus in the small intestine. *J. Parasitol.* **80:** 24–28

Bantscheff M, Schirle M, Sweetman G, Rick J & Kuster B (2007) Quantitative mass spectrometry in proteomics: A critical review. *Anal. Bioanal. Chem.* **389:** 1017–1031

Bartel DP (2009) MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136:** 215–233

Bartel DP (2018) Metazoan MicroRNAs. *Cell* **173:** 20–51

Behm-Ansmant I, Rehwinkel J, Doerks T, Stark A, Bork P & Izaurralde E (2006) mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev.* **20:** 1885

Behnke JM, Menge DM & Noyes H (2009) Heligmosomoides bakeri: A model for exploring the biology and genetics of resistance to chronic gastrointestinal nematode infections. *Parasitology* **136:** 1565–1580

Berkachy R, Smyth DJ, Schnoeller C, Harcus Y, Maizels RM, Selkirk ME & Gounaris K (2021) Characterisation of the secreted apyrase family of Heligmosomoides polygyrus. *Int. J. Parasitol.* **51:** 39–48

Bermúdez-Barrientos JR, Ramírez-S O, Chow FW-N, Buck AH & Abreu-Goodger C (2019) Disentangling sRNA-Seq data to study RNA communication between species. *Nucleic Acids Res.* **48:** 21

Bethony J, Loukas A, Smout M, Brooker S, Mendez S, Plieskatt J, Goud G, Bottazzi ME, Zhan B, Wang Y, Williamson A, Lustigman S, Correa-Oliveira R, Xiao S & Hotez PJ (2005) Antibodies against a secreted protein from hookworm larvae reduce the intensity of hookworm infection in humans and vaccinated laboratory animals. *FASEB J.* **19:** 1743–1745

Bhattacharjee S, Roche B & Martienssen RA (2019) RNA-induced initiation of transcriptional silencing (RITS) complex structure and function. *RNA Biol.* **16:** 1133–1146

Billi AC, Fischer SEJ & Kim JK (2014) Endogenous RNAi pathways in C. elegans. In *WormBook : the online review of C. elegans biology*, Ahringer J (ed) pp 1–49. The C. elegans Research Community

Blasius AL & Beutler B (2010) Intracellular Toll-like Receptors. *Immunity* **32:** 305–315

Boersema PJ, Raijmakers R, Lemeer S, Mohammed S & Heck AJR (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* **4:** 484–494

Boes J & Helwigh A (2000) Animal models of intestinal nematode infections of humans. *Parasitology* **121:** S97-111

Boland A, Tritschler F, Heimstädt S, Izaurralde E & Weichenrieder O (2010) Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. *EMBO Rep.* **11:** 522–527

Bottazzi M-E & Diemert DJ (2019) Safety and Immunogenicity of the Na-APR-1 Hookworm Vaccine in Healthy Adults. *ClinicalTrials.gov* Available at: https://clinicaltrials.gov/ct2/show/NCT01717950 [Accessed November 30, 2021]

Bottini S, Hamouda-Tekaya N, Mategot R, Zaragosi L-E, Audebert S, Pisano S, Grandjean V, Mauduit C, Benahmed M, Barbry P, Repetto E & Trabucchi M (2017) Post-transcriptional gene silencing mediated by microRNAs is controlled by nucleoplasmic Sfpq. *Nat. Commun.* **8:** 1189

Bouchery T, Filbey K, Shepherd A, Chandler J, Patel D, Schmidt A, Camberis M, Peignier A, Smith AAT, Johnston K, Painter G, Pearson M, Giacomin P, Loukas A, Bottazzi M-E, Hotez P & LeGros G (2018) A novel blood-feeding detoxification pathway in Nippostrongylus brasiliensis L3 reveals a potential checkpoint for arresting hookworm development. *PLOS Pathog.* **14:** e1006931

Bouchery T, Le Gros G & Harris N (2019) ILC2s—Trailblazers in the Host Response Against Intestinal Helminths. *Front. Immunol.* **10:** 623

Britton C, Laing R & Devaney E (2020) Small RNAs in parasitic nematodes-forms and functions. *Parasitology* **147:** 855–864

Britton C, Winter AD, Gillan V & Devaney E (2014) microRNAs of parasitic helminths-Identification, characterization and potential as drug targets. *Int. J. Parasitol. Drugs Drug Resist.* **4:** 85–94

Broughton JP, Lovci MT, Huang JL, Yeo GW & Pasquinelli AE (2016) Pairing beyond the Seed Supports MicroRNA Targeting Specificity. *Mol. Cell* **64:** 320–333

Brueggemann AB, Pai R, Crook DW & Beall B (2007) Vaccine Escape Recombinants Emerge after Pneumococcal Vaccination in the United States. *PLOS Pathog.* **3:** e168

Buchan JR & Parker R (2009) Eukaryotic Stress Granules: The Ins and Outs of Translation. *Mol. Cell* **36:** 932–941

Buck AH & Blaxter M (2013) Functional diversification of Argonautes in nematodes: An expanding universe. *Biochem. Soc. Trans.* **41:** 881–886

Buck AH, Coakley G, Simbari F, McSorley HJ, Quintana JF, Le Bihan T, Kumar S, Abreu-Goodger C, Lear M, Harcus Y, Ceroni A, Babayan SA, Blaxter M, Ivens A & Maizels RM (2014) Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity. *Nat. Commun.* **5:** 5488

Cai P, Gobert GN & McManus DP (2016) MicroRNAs in Parasitic Helminthiases: Current Status and Future Perspectives. *Trends Parasitol.* **32:** 71–86

Cai Q, He B, Weiberg A, Buck AH & Jin H (2019) Small RNAs and extracellular vesicles: New mechanisms of cross-species communication and innovative tools for disease control. *PLOS Pathog.* **15:** e1008090

Cai Q, Qiao L, Wang M, He B, Lin F-M, Palmquist J, Huang S-D & Jin H (2018) Plants send small RNAs in extracellular vesicles to fungal pathogen to silence virulence genes. *Science* **360:** 1126–1129

Cai X, Hagedorn CH & Cullen BR (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10:** 1957–1966

Camberis M, Gros G Le & Urban J (2003) Animal Model of Nippostrongylus brasiliensis and Heligmosomoides polygyrus. *Curr. Protoc. Immunol.* **55:** 19.12.1-19.12.27

Campos-Melo D, Hawley ZCE, Droppelmann CA & Strong MJ (2021) The Integral Role of RNA in Stress Granule Formation and Function. *Front. Cell Dev. Biol.* **9:** 621779

De Candia P, De Rosa V, Casiraghi M & Matarese G (2016) Extracellular RNAs: A secret arm of immune system regulation. *J. Biol. Chem.* **291:** 7221–7228

Cardin SE & Borchert GM (2017) Viral MicroRNAs, Host MicroRNAs Regulating Viruses, and Bacterial MicroRNA-Like RNAs. *Methods Mol. Biol.* **1617:** 39–56

Carthew RW & Sontheimer EJ (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136:** 642–655

Casulli A (2021) New global targets for NTDs in the WHO roadmap 2021–2030. *PLoS Negl. Trop. Dis.* **15:** e0009373

Charlesworth AG, Seroussi U, Lehrbach NJ, Renaud MS, Sundby AE, Molnar RI, Lao RX, Willis AR, Woock JR, Aber MJ, Diao AJ, Reinke AW, Ruvkun G & Claycomb JM (2021) Two isoforms of the essential C. elegans Argonaute CSR-1 differentially regulate sperm and oocyte fertility. *Nucleic Acids Res.* **49:** 8836–8865

Charlier J, Rinaldi L, Musella V, Ploeger HW, Chartier C, Vineer HR, Hinney B, von Samson-Himmelstjerna G, Băcescu B, Mickiewicz M, Mateus TL, Martinez-Valladares M, Quealy S, Azaizeh H, Sekovska B, Akkari H, Petkevicius S, Hektoen L, Höglund J, Morgan ER, et al (2020) Initial assessment of the economic burden of major parasitic helminth infections to the ruminant livestock industry in Europe. *Prev. Vet. Med.* **182:** 105103

Chaudhuri AA, So AY-L, Sinha N, Gibson WSJ, Taganov KD, O'Connell RM & Baltimore D (2011) MicroRNA-125b Potentiates Macrophage Activation. *J. Immunol.* **187:** 5062–5068

Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, Guo J, Zhang Y, Chen J, Guo X, Li Q, Li X, Wang W, Zhang Y, Wang J, Jiang X, Xiang Y, Xu C, Zheng P, Zhang J, et al (2008) Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* **18:** 997–1006

Cheng G, Luo R, Hu C, Cao J & Jin Y (2013) Deep sequencing-based identification of pathogen-specific microRNAs in the plasma of rabbits infected with Schistosoma japonicum. *Parasitology* **140:** 1751–1761

Chevillet JR, Kang Q, Ruf IK, Briggs HA, Vojtech LN, Hughes SM, Cheng HH, Arroyo JD, Meredith EK, Gallichotte EN, Pogosova-Agadjanyan EL, Morrissey C, Stirewalt DL, Hladik F, Yu EY, Higano CS & Tewari M (2014) Quantitative and stoichiometric analysis of the microRNA content of exosomes. *Proc. Natl. Acad. Sci.* **111:** 14888–14893

Chow FWN, Koutsovoulos G, Ovando-Vázquez C, Neophytou K, Bermúdez-Barrientos JR, Laetsch DR, Robertson E, Kumar S, Claycomb JM, Blaxter M, Abreu-Goodger C & Buck AH (2019) Secretion of an Argonaute protein by a parasitic nematode and the evolution of its siRNA guides. *Nucleic Acids Res.* **47:** 3594–3606

Chu Y, Yokota S, Liu J, Kilikevicius A, Johnson KC & Corey DR (2021) Argonaute binding within human nuclear RNA and its impact on alternative splicing. *RNA* **27:** 991–1003

Chukwudozie OS, Chukwuanukwu RC, Iroanya OO, Eze DM, Duru VC, Dele-Alimi TO, Kehinde BD, Bankole TT, Obi PC & Okinedo EU (2020) Attenuated subcomponent vaccine design targeting the SARS-CoV-2 nucleocapsid phosphoprotein RNA binding domain: In silico analysis. *J. Immunol. Res.* **2020:** 2837670

Cirillo L, Cieren A, Barbieri S, Khong A, Schwager F, Parker R & Gotta M (2020) UBAP2L Forms Distinct Cores that Act in Nucleating Stress Granules Upstream of G3BP1. *Curr. Biol.* **30:** 698-707.e6

Claerebout E & Geldhof P (2020) Helminth Vaccines in Ruminants: From Development to Application. *Vet. Clin. North Am. Food Anim. Pract.* **36:** 159–171

Claycomb J, Abreu-Goodger C & Buck AH (2017) RNA-mediated communication between helminths and their hosts: The missing links. *RNA Biol.* **14:** 436–441

Clerc M, Devevey G, Fenton A & Pedersen AB (2018) Antibodies and coinfection drive variation in nematode burdens in wild mice. *Int. J. Parasitol.* **48:** 785–792

Coakley G, Maizels RM & Buck AH (2015) Exosomes and Other Extracellular Vesicles: The New Communicators in Parasite Infections. *Trends Parasitol.* **31:** 477–489

Coakley G, McCaskill JL, Borger JG, Simbari F, Robertson E, Millar M, Harcus Y, McSorley HJ, Maizels RM & Buck AH (2017) Extracellular Vesicles from a Helminth Parasite Suppress Macrophage Activation and Constitute an Effective Vaccine for Protective Immunity. *Cell Rep.* **19:** 1545–1557

Coghlan A, Tyagi R, Cotton JA, Holroyd N, Rosa BA, Tsai IJ, Laetsch DR, Beech RN, Day TA, Hallsworth-Pepin K, Ke HM, Kuo TH, Lee TJ, Martin J, Maizels RM, Mutowo P, Ozersky P, Parkinson J, Reid AJ, Rawlings ND, et al (2018) Comparative genomics of the major parasitic worms. *Nat. Genet.* **51:** 163–174

Cohen JD & Sundaram M V. (2020) C. elegans Apical Extracellular Matrices Shape Epithelia. *J. Dev. Biol.* **8:** 1–26

Colombo SAP & Grencis RK (2020) Immunity to Soil-Transmitted Helminths: Evidence From the Field and Laboratory Models. *Front. Immunol.* **11:** 1286

COMBAR (2020) Helminth infection control in farmed ruminants: key priority research needs Available at: https://www.combar-ca.eu/sites/default/files/Research_priorities_for_helminth_control_in_ruminants.pdf

Connolly KJ & Kvalsvig JD (1993) Infection, nutrition and cognitive performance in children. *Parasitology* **107:** S187–S200

Coronado S, Barrios L, Zakzuk J, Regino R, Ahumada V, Franco L, Ocampo Y & Caraballo L (2017) A recombinant cystatin from Ascaris lumbricoides attenuates inflammation of DSS-induced colitis. *Parasite Immunol.* **39:**

Cortés A, Muñoz-Antoli C, Esteban JG & Toledo R (2017) Th2 and Th1 Responses: Clear and Hidden Sides of Immunity Against Intestinal Helminths. *Trends Parasitol.* **33:** 678–693

Crooks GE, Hon G, Chandonia J-M & Brenner SE (2004) WebLogo: A Sequence Logo Generator. *Genome Res.***:** 1188–1190

Cross JH (1996) Enteric Nematodes of Humans. In *Medical Microbiology*, Baron S (ed) Galveston, Texas: University of Texas Medical Branch at Galveston Available at: https://www.ncbi.nlm.nih.gov/books/NBK8261/ [Accessed November 2, 2021]

Czech B & Hannon GJ (2010) Small RNA sorting: matchmaking for Argonautes. *Nat. Rev. Genet.* **12:** 19–31

D'Elia R, Behnke JM, Bradley JE & Else KJ (2009) Regulatory T Cells: A Role in the Control of Helminth-Driven Intestinal Pathology and Worm Survival. *J. Immunol.* **182:** 2340–2348

Dainichi T, Maekawa Y, Ishii K, Zhang T, Fawzy Nashed B, Sakai T, Takashima M & Himeno K (2001) Nippocystatin, a Cysteine Protease Inhibitor from Nippostrongylus brasiliensis, Inhibits Antigen Processing and Modulates Antigen-Specific Immune Response. *Infect. Immun.* **69:** 7380

Davis MPA, van Dongen S, Abreu-Goodger C, Bartonicek N & Enright AJ (2013) Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods* **63:** 41–49

Day K, Howard R, Prowse S, Chapman C & Mitchell G (1979) Studies on chronic versus transient intestinal nematode infections in mice. I. A. comparison of responses to excretory/secretory (ES) products of Nippostrongylus brasiliensis and Nematospiroides dubius worms. *Parasite Immunol.* **1:** 217–239

De N, Young L, Lau PW, Meisner NC, Morrissey D V. & MacRae IJ (2013) Highly complementary target RNAs promote release of guide RNAs from human argonaute2. *Mol. Cell* **50:** 344–355

Deng P, Muhammad S, Cao M & Wu L (2018) Biogenesis and regulatory hierarchy of phased small interfering RNAs in plants. *Plant Biotechnol. J.* **16:** 965–975

Dexheimer PJ & Cochella L (2020) MicroRNAs: From Mechanism to Organism. *Front. cell Dev. Biol.* **8:** 409

Diemert DJ, Freire J, Valente V, Fraga CG, Talles F, Grahek S, Campbell D, Jariwala A, Periago MV, Enk M, Gazzinelli MF, Bottazzi ME, Hamilton R, Brelsford J, Yakovleva A, Li G, Peng J, Correa-Oliveira R, Hotez P & Bethony J (2017) Safety and immunogenicity of the Na-GST-1 hookworm vaccine in Brazilian and American adults. *PLoS Negl. Trop. Dis.* **11:** e0005574

Diemert DJ, Pinto AG, Freire J, Jariwala A, Santiago H, Hamilton RG, Periago MV, Loukas A, Tribolet L, Mulvenna J, Correa-Oliveira R, Hotez PJ & Bethony JM (2012) Generalized urticaria induced by the Na-ASP-2 hookworm vaccine: implications for the development of vaccines against helminths. *J. Allergy Clin. Immunol.* **130:** 169–176

Ditgen D, Anandarajah EM, Meissner KA, Brattig N, Wrenger C & Liebau E (2014) Harnessing the Helminth Secretome for Therapeutic Immunomodulators. *Biomed Res. Int.* **2014:** 964350

Donnelly S & Tran N (2021) Commandeering the mammalian Ago2 miRNA network: a newly discovered mechanism of helminth immunomodulation. *Trends Parasitol.* **37:** 1031–1033

Drurey C, Coakley G & Maizels RM (2020) Extracellular vesicles: new targets for vaccines against helminth parasites. *Int. J. Parasitol.* **50:** 623–633

Drurey C, Lindholm HT, Coakley G, Poveda MC, Löser S, Doolan R, Gerbe F, Jay P, Harris N, Oudhoff MJ & Maizels RM (2022) Intestinal epithelial tuft cell induction is negated by a murine helminth and its secreted products. *J. Exp. Med.* **219:** e20211140

Drurey C & Maizels RM (2021) Helminth extracellular vesicles: Interactions with the host immune system. *Mol. Immunol.* **137:** 124–133

Dudášová J, Laube R, Valiathan C, Wiener MC, Gheyas F, Fišer P, Ivanauskaite J, Liu F & Sachs JR (2021) A method to estimate probability of disease and vaccine efficacy from clinical trial immunogenicity data. *NPJ Vaccines* **6:** 133

Dunker F, Trutzenberg A, Rothenpieler JS, Kuhn S, Pröls R, Schreiber T, Tissier A, Kemen A, Kemen E, Hückelhoven R & Weiberg A (2020) Oomycete small RNAs bind to the plant RNA-induced silencing complex for virulence. *Elife* **9:** e56096

Ebert MS & Sharp PA (2012) Roles for MicroRNAs in Conferring Robustness to Biological Processes. *Cell* **149:** 515–524

Eichenberger RM, Ryan S, Jones L, Buitrago G, Polster R, de Oca MM, Zuvelek J, Giacomin PR, Dent LA, Engwerda CR, Field MA, Sotillo J & Loukas A (2018a) Hookworm secreted extracellular vesicles interact with host cells and prevent inducible colitis in mice. *Front. Immunol.* **9:** 850

Eichenberger RM, Talukder MH, Field MA, Wangchuk P, Giacomin P, Loukas A & Sotillo J (2018b) Characterization of Trichuris muris secreted proteins and extracellular vesicles provides new insights into host–parasite communication. *J. Extracell. Vesicles* **7:** 1428004

Van Eijl RAPM, Van Den Brand T, Nguyen LN & Mulder KW (2017) Reactivity of human AGO2 monoclonal antibody 11A9 with the SWI/SNF complex: A case study for rigorously defining antibody selectivity. *Sci. Rep.* **7:** 1–11

Elbashir S, Martinez J, Patkaniowska A, Lendeckel W & Tuschl T (2001a) Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *EMBO J.* **20:** 6877–6888

Elbashir SM, Lendeckel W & Tuschl T (2001b) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15:** 188–200

Ender C, Krek A, Friedländer MR, Beitzinger M, Weinmann L, Chen W, Pfeffer S, Rajewsky N & Meister G (2008) A Human snoRNA with MicroRNA-Like Functions. *Mol. Cell* **32:** 519–528

Erdmann RM & Picard CL (2020) RNA-directed DNA Methylation. *PLOS Genet.* **16:** e1009034

Evans C, Hardin J & Stoebel DM (2018) Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Brief. Bioinform.* **19:** 776–792

Everts B, Hussaarts L, Driessen NN, Meevissen MHJ, Schramm G, Ham AJ van der, Hoeven B van der, Scholzen T, Burgdorf S, Mohrs M, Pearce EJ, Hokke CH, Haas H, Smits HH & Yazdanbakhsh M (2012) Schistosome-derived omega-1 drives Th2 polarization by suppressing protein synthesis following internalization by the mannose receptor. *J. Exp. Med.* **209:** 1753–1767

Fabbri M, Garzon R, Cimmino A, Liu Z, Zanesi N, Callegari E, Liu S, Alder H, Costinean S, Fernandez-Cymering C, Volinia S, Guler G, Morrison CD, Chan KK, Marcucci G, Calin GA, Huebner K & Croce CM (2007) MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proc. Natl. Acad. Sci. U. S. A.* **104:** 15805–15810

Fabián Flores-Jasso C, Salomon WE & Zamore PD (2013) Rapid and specific purification of Argonaute-small RNA complexes from crude cell lysates. *RNA* **19:** 271–279

Faehnle CR & Joshua-Tor L (2007) Argonautes confront new small RNAs. *Curr. Opin. Chem. Biol.* **11:** 569–577

Fan M, Zhang Y, Huang Z, Liu J, Guo X, Zhang H & Luo H (2014) Optimizations of SiRNA Design for the Activation of Gene Transcription by Targeting the TATA-Box Motif. *PLoS One* **9:** e108253

Fang Z & Rajewsky N (2011) The Impact of miRNA Target Sites in Coding Sequences and in 3′UTRs. *PLoS One* **6:** e18067

Fei Q, Xia R & Meyers BC (2013) Phased, Secondary, Small Interfering RNAs in Posttranscriptional Regulatory Networks. *Plant Cell* **25:** 2400–2415

Fields BD & Kennedy S (2019) Chromatin Compaction by Small RNAs and the Nuclear RNAi Machinery in C. elegans. *Sci. Rep.* **9:** 9030

Figliuolo da Paz V, Ghishan FK & Kiela PR (2020) Emerging Roles of Disabled Homolog 2 (DAB2) in Immune Regulation. *Front. Immunol.* **11:** 580302

Finkielstein C V. & Capelluto DGS (2016) Disabled-2: A modular scaffold protein with multifaceted functions in signaling. *BioEssays* **38:** S45–S55

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE & Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391:** 806–811

Fowler EK, Mohorianu I, Smith DT, Dalmay T & Chapman T (2018) Small RNA populations revealed by blocking rRNA fragments in Drosophila melanogaster reproductive tissues. *PLoS One* **13:** e0191966

de Franceschi N, Hamidi H, Alanko J, Sahgal P & Ivaska J (2015) Integrin traffic-the update. *J. Cell Sci.* **128:** 839–852

Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, Barnes I, Berry A, Bignell A, Carbonell Sala S, Chrast J, Cunningham F, Di Domenico T, Donaldson S, Fiddes IT, García Girón C, et al (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47:** D766–D773

Friend K, Campbell ZT, Cooke A, Kroll-Conner P, Wickens MP & Kimble J (2012) A conserved PUF-Ago-eEF1A complex attenuates translation elongation. *Nat. Struct. Mol. Biol.* **19:** 176–184

Frohn A, Christian Eberl H, Stö hr J, Glasmacher E, Rü del S, Heissmeyer V, Mann M & Meister G (2012) Dicer-dependent and-independent Argonaute2 Protein Interaction Networks in Mammalian Cells. *Mol. Cell. Proteomics* **11:** 1442–1456

Fukaya T, Iwakawa H oki & Tomari Y (2014) MicroRNAs Block Assembly of eIF4F Translation Initiation Complex in Drosophila. *Mol. Cell* **56:** 67–78

Furman D & Davis MM (2015) New approaches to understanding the immune response to vaccination and infection. *Vaccine* **33:** 5271–5281

Garcia-Silva MR, Cabrera-Cabrera F, Ferreira Cura Das Neves R, Souto-Padrón T, De Souza W & Cayota A (2014) Gene Expression Changes Induced by Trypanosoma cruzi Shed Microvesicles in Mammalian Host Cells: Relevance of tRNA-Derived Halves. *Biomed Res. Int.* **2014:** 1–11

Garcia Silva MR, Tosar JP, Frugier M, Pantano S, Bonilla B, Esteban L, Serra E, Rovira C, Robello C & Cayota A (2010) Cloning, characterization and subcellular localization of a Trypanosoma cruzi argonaute protein defining a new subfamily distinctive of trypanosomatids. *Gene* **466:** 26–35

Garçon N, Hem S & Friede M (2013) Evolution of adjuvants across the centuries. In *Vaccines*, Plotkin S Orenstein W & Offit P (eds) pp 58–70. W.B. Saunders

Gebert L & Macrae IJ (2019) Regulation of microRNA function in animals. *Nat. Rev. Mol. Cell Biol.* **20:** 21–37

Geldhof P, De Maere V, Vercruysse J & Claerebout E (2007) Recombinant expression systems: the obstacle to helminth vaccines? *Trends Parasitol.* **23:** 527–532

Gentile ME, Li Y, Robertson A, Shah K, Fontes G, Kaufmann E, Polese B, Khan N, Parisien M, Munter HM, Mandl JN, Diatchenko L, Divangahi M & King IL (2019) NK cell recruitment limits tissue damage during an enteric helminth infection. *Mucosal Immunol.* **13:** 357–370

Gerbe F, Legraverend C & Jay P (2012) The intestinal epithelium tuft cells: specification and function. *Cell. Mol. Life Sci.* **69:** 2907–2917

Gill HS, Gray GD, Watson DL & Husband AJ (1993) Isotype-specific antibody responses to Haemonchus contortus in genetically resistant sheep. *Parasite Immunol.* **15:** 61–67

Girard A, Sachidanandam R, Hannon GJ & Carmell MA (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442:** 199–202

Girgis NM, Gundra UM & Loke P (2013) Immune Regulation during Helminth Infections The Co-Evolution of Helminths and the Mammalian Immune System. *PLOS Pathog.* **9:** e1003250

Glover M, Colombo SAP, Thornton DJ & Grencis RK (2019) Trickle infection and immunity to Trichuris muris. *PLoS Pathog.* **15:** e1007926

González-Hernández A, Van Coppernolle S, Borloo J, Van Meulder F, Paerewijck O, Peelaers I, Leclercq G, Claerebout E & Geldhof P (2016) Host protective ASP-based vaccine against the parasitic nematode Ostertagia ostertagi triggers NK cell activation and mixed IgG1-IgG2 response. *Sci. Rep.* **6:** 29496

Griffiths JA, Richard AC, Bach K, Lun ATL & Marioni JC (2018) Detection and removal of barcode swapping in single-cell RNA-seq data. *Nat. Commun.* **9:** 2667

Grishok A, Sinskey JL & Sharp PA (2005) Transcriptional silencing of a transgene by RNAi in the soma of C. elegans. *Genes Dev.* **19:** 683–696

Gruner HN & McManus MT (2021) Examining the evidence for extracellular RNA function in mammals. *Nat. Rev. Genet.* **22:** 448–458

Gu W, Shirayama M, Conte D, Vasale J, Batista PJ, Claycomb JM, Moresco JJ, Youngman EM, Keys J, Stoltz MJ, Chen C-CG, Chaves DA, Duan S, Kasschau KD, Fahlgren N, Yates JR, Mitani S, Carrington JC & Mello CC (2009) Distinct Argonaute-Mediated 22G-RNA Pathways Direct Genome Surveillance in the C. elegans Germline. *Mol. Cell* **36:** 231–244

Guang S, Bochner AF, Burkhart KB, Burton N, Pavelec DM & Kennedy S (2010) Small regulatory RNAs inhibit RNA Polymerase II during the elongation phase of transcription. *Nature* **465:** 1097–1101

Guang S, Bochner AF, Pavelec DM, Burkhart KB, Harding S, Lachowiec J & Kennedy S (2008) An argonaute transports siRNAs from the cytoplasm to the nucleus. *Science* **321:** 537–541

Guérin TM, Palladino F & Robert VJ (2014) Transgenerational functions of small RNA pathways in controlling gene expression in C. elegans. *Epigenetics* **9:** 37–44

Gulland F (1992) The role of nematode parasites in Soay sheep (Ovis aries L.) mortality during a population crash. *Parasitology* **105:** 493–503

Haimovich G, Ecker CM, Dunagin MC, Eggan E, Raj A, Gerst JE & Singer RH (2017) Intercellular mRNA trafficking via membrane nanotube-like extensions in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **114:** E9873–E9882

Halliez MCM & Buret AG (2015) Gastrointestinal Parasites and the Neural Control of Gut Functions. *Front. Cell. Neurosci.* **9:** 452

Hamilton A, Voinnet O, Chappell L & Baulcombe D (2002) Two classes of short interfering RNA in RNA silencing. *EMBO J.* **21:** 4671–4679

Hamilton AJ & Baulcombe DC (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286:** 950–952

Hammad H & Lambrecht BN (2015) Barrier Epithelial Cells and the Control of Type 2 Immunity. *Immunity* **43:** 29–40

Hammer M, Mages J, Dietrich H, Servatius A, Howells N, Cato ACB & Lang R (2006) Dual specificity phosphatase 1 (DUSP1) regulates a subset of LPS-induced genes and protects mice from lethal endotoxin shock. *J. Exp. Med.* **203:** 15–20

Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT & Kim VN (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125:** 887–901

Harris N & Gause WC (2011) B cell function in the immune response to helminths. *Trends Immunol.* **32:** 80–88

Hart BL & Hart LA (2018) How mammals stay healthy in nature: the evolution of behaviours to avoid parasites and pathogens. *Philos. Trans. R. Soc. B Biol. Sci.* **373:** 20170205

Hausser J, Syed AP, Bilen B & Zavolan M (2013) Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. *Genome Res.* **23:** 604–615

He B, Cai Q, Qiao L, Huang CY, Wang S, Miao W, Ha T, Wang Y & Jin H (2021) RNA-binding proteins contribute to small RNA loading in plant extracellular vesicles. *Nat. Plants* **7:** 342–352

He Z, Jing S, Yang T, Chen J, Huang F, Zhang W, Peng Z, Liu B, Ma X, Wu L, Pan T, Zhang X, Li L, Cai W, Tang X, Zhang J & Zhang H (2020) PIWIL4 Maintains HIV-1 Latency by Enforcing Epigenetically Suppressive Modifications on the 5′ Long Terminal Repeat. *J. Virol.* **94:** e01923-19

Hegge JW, Swarts DC & Van Der Oost J (2018) Prokaryotic argonaute proteins: Novel genome-editing tools? *Nat. Rev. Microbiol.* **16:** 5–11

Heil F, Hemmi H, Hochrein H, Ampenberger F, Kirschning C, Akira S, Lipford G, Wagner H & Bauer S (2004) Species-Specific Recognition of Single-Stranded RNA via Till-like Receptor 7 and 8. *Science* **303:** 1526–1529

Helwak A, Kudla G, Dudnakova T & Tollervey D (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153:** 654–665

Helwak A & Tollervey D (2014) Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). *Nat. Protoc.* **9:** 711–728

Hentzschel F, Obrova K & Marti M (2020) No evidence for Ago2 translocation from the host erythrocyte into the Plasmodium parasite [version 2; peer review: 2 approved]. *Wellcome Open Res.* **5:** 92

Herbert DR, Yang J-Q, Hogan SP, Groschwitz K, Khodoun M, Munitz A, Orekov T, Perkins C, Wang Q, Brombacher F, Urban JF, Jr., Rothenberg ME & Finkelman FD (2009) Intestinal epithelial cell secretion of RELM-β protects against gastrointestinal worm infection. *J. Exp. Med.* **206:** 2947–2957

Hewitson JP, Filbey KJ, Esser-von Bieren J, Camberis M, Schwartz C, Murray J, Reynolds LA, Blair N, Robertson E, Harcus Y, Boon L, Huang SC-C, Yang L, Tu Y, Miller MJ, Voehringer D, Le Gros G, Harris N & Maizels RM (2015) Concerted Activity of IgG1 Antibodies and IL-4/IL-25-Dependent Effector Cells Trap Helminth Larvae in the Tissues following Vaccination with Defined Secreted Antigens, Providing Sterile Immunity to Challenge Infection. *PLOS Pathog.* **11:** e1004676

Hewitson JP, Grainger JR & Maizels RM (2009) Helminth immunoregulation: The role of parasite secreted proteins in modulating host immunity. *Mol. Biochem. Parasitol.* **167:** 1–11

Hewitson JP, Ivens AC, Harcus Y, Filbey KJ, McSorley HJ, Murray J, Bridgett S, Ashford D, Dowle AA & Maizels RM (2013) Secretion of Protective Antigens by Tissue-Stage Nematode Larvae Revealed by Proteomic Analysis and Vaccination-Induced Sterile Immunity. *PLoS Pathog.* **9:** e1003492

Hicks JA, Li L, Matsui M, Chu Y, Volkov O, Johnson KC & Corey DR (2017) Human GW182 Paralogs Are the Central Organizers for RNA-Mediated Control of Transcription. *Cell Rep.* **20:** 1543–1552

Höck J & Meister G (2008) The Argonaute protein family. *Genome Biol.*

Hokke CH & van Diepen A (2017) Helminth glycomics – glycan repertoires and host-parasite interactions. *Mol. Biochem. Parasitol.* **215:** 47–57

Holoch D & Moazed D (2015) RNA-mediated epigenetic regulation of gene expression. *Nat. Publ. Gr.* **16:** 71–84

Hoogstrate SW, Volkers RJ, Sterken MG, Kammenga JE & Snoek LB (2014) Nematode endogenous small RNA pathways. *Worm* **3:** e28234

Hooper L V. (2015) Epithelial Cell Contributions to Intestinal Immunity. *Adv. Immunol.* **126:** 129–172

Hotez PJ, Alvarado M, Basá M-G, Bolliger I, Bourne R, Boussinesq M, Brooker SJ, Shah Brown A, Buckle G, Budke CM, Halasa YA, Jasrasaria R, Johns NE, Keiser J, King CH, Lozano R, Murdoch ME, O'hanlon S, Bastien S, Pion DS, et al (2014) The Global Burden of Disease Study 2010: Interpretation and Implications for the Neglected Tropical Diseases. *PLoS Negl. Trop. Dis.* **8:** e2865

Hotez PJ, Bundy DAP, Beegle K, Brooker S, Drake L, Silva N de, Montresor A, Engels D, Jukes M, Chitsulo L, Chow J, Laxminarayan R, Michaud C, Bethony J, Correa-Oliveira R, Shuhua X, Fenwick A & Savioli L (2006) Helminth Infections: Soil-transmitted Helminth Infections and Schistosomiasis. In *Disease Control Priorities in Developing Countries*, Jamison D Breman J AR M Alleyne G Claeson M Evans D Jha

P Mills A & Musgrove P (eds) Washington, DC: The International Bank for Reconstruction and Development / The World Bank

Hotez PJ, Molyneux DH, Fenwick A, Kumaresan J, Sachs SE, Sachs JD & Savioli L (2007) Control of Neglected Tropical Diseases. *N. Engl. J. Med.* **357:** 1018–1027

Hoy AM, Lundie RJ, Ivens A, Quintana JF, Nausch N, Forster T, Jones F, Kabatereine NB, Dunne DW, Mutapi F, MacDonald AS & Buck AH (2014) Parasite-derived microRNAs in host serum as novel biomarkers of helminth infection. *PLoS Negl. Trop. Dis.* **8:** e2701

Hsu CC, Li HP, Hung YH, Leu YW, Wu WH, Wang FS, Lee K Der, Chang PJ, Wu CS, Lu YJ, Huang THM, Chang YS & Hsiao SH (2010) Targeted methylation of CMV and E1A viral promoters. *Biochem. Biophys. Res. Commun.* **402:** 228–234

Huang V & Li LC (2014) Demystifying the nuclear function of Argonaute proteins. *RNA Biol.* **11:** 18–24

Hudson PJ, Dobson AP & Newborn D (1998) Prevention of Population Cycles by Parasite Removal. *Science* **282:** 2256–2258

Hunt VL, Hino A, Yoshida A & Kikuchi T (2018) Comparative transcriptomics gives insights into the evolution of parasitism in Strongyloides nematodes at the genus, subclade and species level. *Sci. Rep.* **8:** 1–9

Huntzinger E, Kuzuoğlu-Öztürk D, Braun JE, Eulalio A, Wohlbold L & Izaurralde E (2013) The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. *Nucleic Acids Res.* **41:** 978–94

Inoue A, Sato O, Homma K & Ikebe M (2002) DOC-2/DAB2 is the binding partner of myosin VI. *Biochem. Biophys. Res. Commun.* **292:** 300–307

Ismail HAHA, Jeon H-K, Yu Y-M, Do C & Lee Y-H (2010) Intestinal Parasite Infections in Pigs and Beef Cattle in Rural Areas of Chungcheongnam-do, Korea. *Korean J. Parasitol.* **48:** 347–349

Iwasaki S, Kobayashi M, Yoda M, Sakaguchi Y, Katsuma S, Suzuki T & Tomari Y (2010) Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol. Cell* **39:** 292–299

Iyer SS & Cheng G (2012) Role of Interleukin 10 Transcriptional Regulation in Inflammation and Autoimmune Disease. *Crit. Rev. Immunol.* **32:** 23–63

Jakymiw A, Lian S, Eystathioy T, Li S, Satoh M, Hamel JC, Fritzler MJ & Chan EKL (2005) Disruption of GW bodies impairs mammalian RNA interference. *Nat. Cell Biol.* **7:** 1167–1174

Jazurek M, Ciesiolka A, Starega-Roslan J, Bilinska K & Krzyzosiak WJ (2016) Identifying proteins that bind to specific RNAs - focus on simple repeat expansion diseases. *Nucleic Acids Res.* **44:** 9050–9070

Jenkins TP, Rathnayaka Y, Perera PK, Peachey LE, Nolan MJ, Krause L, Rajakaruna RS & Cantacessi C (2017) Infections by human gastrointestinal helminths are associated with changes in faecal microbiota diversity and composition. *PLoS One* **12:** e0184719

Jia T-W, Melville S, Utzinger J, King CH & Zhou X-N (2012) Soil-Transmitted Helminth Reinfection after Drug Treatment: A Systematic Review and Meta-Analysis. *PLoS Negl. Trop. Dis.* **6:** e1621

Jinek M, Fabian MR, Coyle SM, Sonenberg N & Doudna JA (2010) Structural insights into the human GW182-PABC interaction in microRNA-mediated deadenylation. *Nat. Struct. Mol. Biol.* **17:** 238–240

Johnson NR, Yeoh JM, Coruh C & Axtell MJ (2016) Improved placement of multi-mapping small RNAs. *G3* **6:** 2103–2111

Johnston CJC, Robertson E, Harcus Y, Grainger JR, Coakley G, Smyth DJ, McSorley HJ & Maizels R (2015) Cultivation of Heligmosomoides Polygyrus: An Immunomodulatory Nematode Parasite and its Secreted Products. *J. Vis. Exp.* **98:** e52412

Johnston CJC, Smyth DJ, Kodali RB, White MPJ, Harcus Y, Filbey KJ, Hewitson JP, Hinck CS, Ivens A, Kemter AM, Kildemoes AO, Le Bihan T, Soares DC, Anderton SM, Brenn T, Wigmore SJ, Woodcock H V., Chambers RC, Hinck AP, McSorley HJ, et al (2017) A structurally distinct TGF-β mimic from an intestinal helminth parasite potently induces regulatory T cells. *Nat. Commun.* **8:** 1741

Juillerat A, Lewit-Bentley A, Guillotte M, Gangnard S, Hessel A, Baron B, Vigan-Womas I, England P, Mercereau-Puijalon O & Bentley GA (2011) Structure of a Plasmodium falciparum PfEMP1 rosetting domain reveals a role for the N-terminal segment in heparin-mediated rosette inhibition. *Proc. Natl. Acad. Sci. U. S. A.* **108:** 5243–5248

Kaur L, Gordon M, Baines PA, Iheozor-Ejiofor Z, Sinopoulou V & Akobeng AK (2020) Probiotics for induction of remission in ulcerative colitis. *Cochrane Database Syst. Rev.* **3:** CD005573

Kaya E, Doxzen KW, Knoll KR, Wilson RC, Strutt SC, Kranzusch PJ, Doudna JA, Designed JAD & Performed SCS (2016) A bacterial Argonaute with noncanonical guide RNA specificity. *PNAS* **113:** 4057–4062

Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, Furlotte NA, Eskin E, Nellåker C, Whitley H, Cleak J, Janowitz D, Hernandez-Pliego P, Edwards A, Belgard TG, Oliver PL, et al (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477:** 289–294

Kebeta MM, Hine BC, Walkden-Brown SW, Kahn LP & Doyle EK (2021) Protective efficacy of Barbervax® in Merino weaner sheep trickle infected with five doses of Haemonchus contortus infective larvae. *Vet. Parasitol.* **292:** 109386

Kennedy DA & Read AF (2017) Why does drug resistance readily evolve but vaccine resistance does not? *Proc. R. Soc. B Biol. Sci.* **284:** 20162562

Kensil C (2006) Immunomodulatory adjuvants from Quillaja saponaria. In *Immunopotentiators in Modern Vaccines*, Schijns V & O'Hagan D (eds) pp 109–122. Academic Press

Ketting FR & Cochella L (2020) Concepts and functions of small RNA pathways in C. elegans. *Curr. Top. Dev. Biol.* **144:** 45–89

Ketting RF, Fischer SEJ, Bernstein E, Sijen T, Hannon GJ & Plasterk RHA (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes Dev.* **15:** 2654–2659

Kim H, Ding Y-H, Zhang G, Yan Y-H, Conte Jr D, Dong M-Q & Mello CC (2021) HDAC1 SUMOylation promotes Argonaute-directed transcriptional silencing in C. elegans. *Elife***:** e63299

Kim S, Jeon OH & Jeon YJ (2020) Extracellular RNA: Emerging roles in cancer cell communication and biomarkers. *Cancer Lett.* **495:** 33–40

Kim YS & Ho SB (2010) Intestinal Goblet Cells and Mucins in Health and Disease: Recent Insights and Progress. *Curr. Gastroenterol. Rep.* **12:** 319–330

King IL & Li Y (2018) Host-Parasite Interactions Promote Disease Tolerance to Intestinal Helminth Infection. *Front. Immunol.* **9:** 2128

Knutie SA, Wilkinson CL, Wu QC, Ortega CN & Rohr JR (2017) Host resistance and tolerance of parasitic gut worms depend on resource availability. *Oecologia* **183:** 1031–1040

Kolodny GM, Culp LA & Rosenthal LJ (1972) Secretion of RNA by normal and transformed cells. *Exp. Cell Res.* **73:** 65–72

Konoshenko MY, Lekchnov EA, Vlassov A V. & Laktionov PP (2018) Isolation of Extracellular Vesicles: General Methodologies and Latest Trends. *Biomed Res. Int.* **2018:** 8545347

Kowalski MP & Krude T (2015) Functional roles of non-coding Y RNAs. *Int. J. Biochem. Cell Biol.* **66:** 20–29

Kuipers ME, Nolte-'t Hoen ENM, van der Ham AJ, Ozir-Fazalalikhan A, Nguyen DL, de Korne CM, Koning RI, Tomes JJ, Hoffmann KF, Smits HH & Hokke CH (2020) DC-SIGN mediated internalisation of glycosylated extracellular vesicles from Schistosoma mansoni increases activation of monocyte-derived dendritic cells. *J. Extracell. Vesicles* **9:** 1753420

Kumar M, Gouw M, Michael S, Sámano-Sánchez H, Pancsa R, Glavina J, Diakogianni A, Valverde JA, Bukirova D, Čalyševa J, Palopoli N, Davey NE, Chemes LB & Gibson TJ (2020) ELM—the eukaryotic linear motif resource in 2020. *Nucleic Acids Res.* **48:** D296

Kuscu C, Kumar P, Kiran M, Su Z, Malik A & Dutta A (2018) tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *RNA* **24:** 1093–1105

Kuzuoğlu-Öztürk D, Huntzinger E, Schmidt S & Izaurralde E (2012) The Caenorhabditis elegans GW182 protein AIN-1 interacts with PAB-1 and subunits of the PAN2-PAN3 and CCR4-NOT deadenylase complexes. *Nucleic Acids Res.* **40:** 5651–65

De La Torre-Escudero E, Gerlach JQ, Bennett APS, Cwiklinski K, Jewhurst HL, Huson KM, Joshi L, Kilcoyne M, O'neill S, Dalton JP & Robinson MW (2019) Surface molecules of extracellular vesicles secreted by the helminth pathogen Fasciola hepatica direct their internalisation by host cells. *PLoS Negl. Trop. Dis.* **13:** e0007087

Lai EC (2002) Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.* **30:** 363–364

Lalani AI, Zhu S, Gokhale S, Jin J & Xie P (2017) TRAF Molecules in Inflammation and Inflammatory Diseases. *Curr. Pharmacol. Reports* **4:** 64–90

Langer C, Rücker FG, Buske C, Döhner H & Kuchenbauer F (2012) Targeted therapies through microRNAs: pulp or fiction? *Ther. Adv. Hematol.* **3:** 97–104

Langmead B, Trapnell C, Pop M & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10:** R25

Lao RX, Seroussi U, Wadi L, Willis AR, Maity T, Reinke AW, Blaxter M, Abreu-Goodger C, Buck AH & Claycomb JM (2020) Understanding Argonaute/Small RNA-based Intercellular Communication. In 25th Annual Meeting of the RNA Society, RNA2020 Available at: https://app.oxfordabstracts.com/events/1385/program-app/submission/189924 [Accessed September 13, 2021]

Lee RC, Feinbaum RL & Ambros V (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75:** 843–854

Lee Y, Jeon K, Lee JT, Kim S & Kim VN (2002) MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.* **21:** 4663–4670

Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH & Kim VN (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23:** 4051–4060

Lello J, Boag B, Fenton A, Stevenson IR & Hudson PJ (2004) Competition and mutualism among the gut helminths of a mammalian host. *Nature* **428:** 840–844

Lendner M, Doligalska M, Lucius R & Hartmann S (2008) Attempts to establish RNA interference in the parasitic nematode Heligmosomoides polygyrus. *Mol. Biochem. Parasitol.* **161:** 21–31

Lenz S, Sinn LR, O'Reilly FJ, Fischer L, Wegner F & Rappsilber J (2021) Reliable identification of protein-protein interactions by crosslinking mass spectrometry. *Nat. Commun.* **12:** 1–11

Leung AKL, Calabrese JM & Sharp PA (2006) Quantitative analysis of Argonaute protein reveals microRNA-dependent localization to stress granules. *Proc. Natl. Acad. Sci. USA* **103:** 18125–18130

Leung AKL, Young AG, Bhutkar A, Zheng GX, Bosson AD, Nielsen CB & Sharp PA (2011) Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat. Struct. Mol. Biol.* **18:** 237–244

Li K, Rodosthenous RS, Kashanchi F, Gingeras T, Gould SJ, Kuo LS, Kurre P, Lee H, Leonard JN, Liu H, Lombo TB, Momma S, Nolan JP, Ochocinska MJ, Michiel D, Sadovsky Y, Sánchez-Madrid F, Valdes KM, Vickers KC, Weaver AM, et al (2018) Advances, challenges, and opportunities in extracellular RNA biology: insights from the NIH exRNA Strategic Workshop. *JCI Insight* **3:** e98942

Li LC, Okino ST, Zhao H, Pookot D, Place RF, Urakami S, Enokida H & Dahiya R (2006) Small dsRNAs induce transcriptional activation in human cells. *Proc. Natl. Acad. Sci. U. S. A.* **103:** 17337–17342

Lian SL, Li S, Abadal GX, Pauley BA, Fritzler MJ & Chan EKL (2009) The C-terminal half of human Ago2 binds to multiple GW-rich regions of GW182 and requires GW182 to mediate silencing. *RNA* **15:** 804–13

Lingel A, Simon B, Izaurralde E & Sattler M (2004) Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat. Struct. Mol. Biol.* **11:** 576–577

Liu J, Carmell MA, Rivas F V, Marsden CG, Thomson JM, Song J-J, Hammond SM, Joshua-Tor L & Hannon GJ (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305:** 1437–41

Liu J, Rivas F V., Wohlschlegel J, Yates JR, Parker R & Hannon GJ (2005a) A role for the P-body component GW182 in microRNA function. *Nat. Cell Biol.* **7:** 1161–1166

Liu J, Valencia-Sanchez MA, Hannon GJ & Parker R (2005b) MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat. Cell Biol.* **7:** 719–723

Liu J, Zhu L, Wang J, Qiu L, Chen Y, Davis RE & Cheng G (2019a) Schistosoma japonicum extracellular vesicle miRNA cargo regulates host macrophage functions facilitating parasitism. *PLoS Pathog.* **15:** e1007817

Liu M, Roth A, Yu M, Morris R, Bersani F, Rivera MN, Lu J, Shioda T, Vasudevan S, Ramaswamy S, Maheswaran S, Diederichs S & Haber DA (2013) The IGF2 intronic miR-483 selectively enhances transcription from IGF2 fetal promoters and enhances tumorigenesis. *Genes Dev.* **27:** 2543–2548

Liu Y, Zhang Y, Dong P, An R, Xue C, Ge Y, Wei L & Liang X (2015) Digestion of Nucleic Acids Starts in the Stomach. *Sci. Rep.* **5:** 1–11

Liu Z, Johnson ST, Zhang Z & Corey DR (2019b) Expression of TNRC6 (GW182) Proteins Is Not Necessary for Gene Silencing by Fully Complementary RNA Duplexes. *Nucleic Acid Ther.* **29:** 323–334

Logan J, Pearson MS, Manda SS, Choi YJ, Field M, Eichenberger RM, Mulvenna J, Nagaraj SH, Fujiwara RT, Gazzinelli-Guimaraes P, Bueno L, Mati V, Bethony JM, Mitreva M, Sotillo J & Loukas A (2020) Comprehensive analysis of the secreted proteome of adult Necator americanus hookworms. *PLoS Negl. Trop. Dis.* **14:** 1–30

Loukas A, Bethony JM, Mendez S, Fujiwara RT, Goud GN, Ranjit N, Zhan B, Jones K, Bottazzi ME & Hotez PJ (2005) Vaccination with Recombinant Aspartic Hemoglobinase Reduces Parasite Load and Blood Loss after Hookworm Infection in Dogs. *PLoS Med.* **2:** 1008–1017

Loukas A, Hotez PJ, Diemert D, Yazdanbakhsh M, McCarthy JS, Correa-Oliveira R, Croese J & Bethony JM (2016) Hookworm infection. *Nat. Rev. Dis. Prim.* **2:** 1–18

Lukasik A, Wójcikowski M & Zielenkiewicz P (2016) Tools4miRs – one place to gather all the tools for miRNA analysis. *Bioinformatics* **32:** 2722–2724

Luo Y, Na Z & Slavoff SA (2018) P-Bodies: Composition, Properties, and Functions. *Biochemistry* **57:** 2424–2431

Lytle JR, Yario TA & Steitz JA (2007) Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5 UTR as in the 3 UTR. *Proc. Natl. Acad. Sci. USA* **104:** 9667–9672

Macpherson AJ, Geuking MB & McCoy KD (2012) Homeland Security: IgA immunity at the frontiers of the body. *Trends Immunol.* **33:** 160–167

MacRae I, Ma E, Zhou M, Robinson C & Doudna J (2008) In vitro reconstitution of the human RISC-loading complex. *PNAS* **105:** 512–517

Maduro MF (2009) Structure and evolution of the C. elegans embryonic endomesoderm network. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1789:** 250–260

Maizels RM (2020) Regulation of immunity and allergy by helminth parasites. *Allergy* **75:** 524–534

Maizels RM, Smits HH & McSorley HJ (2018) Modulation of Host Immunity by Helminths: The Expanding Repertoire of Parasite Effector Molecules. *Immunity* **49:** 801–818

Maniar JM & Fire AZ (2011) EGO-1, a C. elegans RdRP, Modulates Gene Expression via Production of mRNA-Templated Short Antisense RNAs. *Curr. Biol.* **21:** 449–459

Marcilla A, Trelis M, Cortés A, Sotillo J, Cantalapiedra F, Minguez MT, Valero ML, Sánchez del Pino MM, Muñoz-Antoli C, Toledo R & Bernal D (2012) Extracellular Vesicles from Parasitic Helminths Contain Specific Excretory/Secretory Proteins and Are Internalized in Intestinal Host Cells. *PLoS One* **7:** e45974

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17:** 10–12

Maruszewska-Cheruiyot M, Donskow-Lysoniewska K & Doligalska M (2018) Helminth Therapy: Advances in the use of Parasitic Worms Against Inflammatory Bowel Diseases and its Challenges. *Helminthologia* **55:** 1–11

Matsumoto M, Sasaki Y, Yasuda K, Takai T, Muramatsu M, Yoshimoto T & Nakanishi K (2013) IgG and IgE Collaboratively Accelerate Expulsion of Strongyloides venezuelensis in a Primary Infection. *Infect. Immun.* **81:** 2518–2527

Mayya VK & Duchaine TF (2019) Ciphers and Executioners: How 3′-Untranslated Regions Determine the Fate of Messenger RNAs. *Front. Genet.* **10:** 6

Mazumder B, Seshadri V & Fox PL (2003) Translational control by the 3′-UTR: the ends specify the means. *Trends Biochem. Sci.* **28:** 91–98

McCoy KD, Stoel M, Stettler R, Merky P, Fink K, Senn BM, Schaer C, Massacand J, Odermatt B, Oettgen HC, Zinkernagel RM, Bos NA, Hengartner H, Macpherson AJ & Harris NL (2008) Polyclonal and Specific Antibodies Mediate Protective Immunity against Enteric Helminth Infection. *Cell Host Microbe* **4:** 362–373

McCue AD, Panda K, Nuthikattu S, Choudury SG, Thomas EN & Slotkin RK (2015) ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. *EMBO J.* **34:** 20–35

Mckay DM, Shute A & Lopes F (2017) Helminths and intestinal barrier function. *Tissue Barriers* **5:** e1283385

McKellar QA & Jackson F (2004) Veterinary anthelmintics: old and new. *Trends Parasitol.* **20:** 456–461

McNeil MM & DeStefano F (2018) Vaccine-associated hypersensitivity. *J. Allergy Clin. Immunol.* **141:** 463–472

McSorley HJ & Maizels RM (2012) Helminth Infections and Host Immune Regulation. **25:** 585–608

Meister G (2013) Argonaute proteins: Functional insights and emerging roles. *Nat. Rev. Genet.* **14:** 447–59

Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G & Tuschl T (2004) Human Argonaute 2 Mediates RNA Cleavage Targeted by miRNAs and siRNAs. *Mol. Cell* **15:** 185–197

Meister G, Landthaler M, Peters L, Chen PY, Urlaub H, Lührmann R & Tuschl T (2005) Identification of Novel Argonaute-Associated Proteins. *Curr. Biol.* **15:** 2149–2155

Mellacheruvu D, Wright Z, Couzens AL, Lambert J-P, St-Denis NA, Li T, Miteva Y V, Hauri S, Sardiu ME, Low TY, Halim VA, Bagshaw RD, Hubner NC, Al-Hakim A, Bouchard A, Faubert D, Fermin D, Dunham WH, Goudreault M, Lin Z-Y, et al (2013) The CRAPome: a contaminant repository for affinity purification–mass spectrometry data. *Nat. Methods* **10:** 730–736

Meng Y, Shao C, Ma X & Wang H (2013) Introns targeted by plant microRNAs: a possible novel mechanism of gene regulation. *Rice* **6:** 8

Mette MF, Aufsatz W, Van der Winden J, Matzke MA & Matzke AJM (2000) Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* **19:** 5194–5201

Miyoshi T, Takeuchi A, Siomi H & Siomi MC (2010) A direct role for Hsp90 in pre-RISC formation in Drosophila. *Nat. Struct. Mol. Biol.* **17:** 1024–1026

Mlecnik B, Galon J & Bindea G (2018) Comprehensive functional analysis of large lists of genes and proteins. *J. Proteomics* **171:** 2–10

von Moltke J, Ji M, Liang H-E & Locksley RM (2015) Tuft-cell-derived IL-25 regulates an intestinal ILC2–epithelial response circuit. *Nature* **529:** 221–225

Montaño KJ, Cuéllar C & Sotillo J (2021a) Rodent Models for the Study of Soil-Transmitted Helminths: A Proteomics Approach. *Front. Cell. Infect. Microbiol.* **11:** 639573

Montaño KJ, Loukas A & Sotillo J (2021b) Proteomic approaches to drive advances in helminth extracellular vesicle research. *Mol. Immunol.* **131:** 1–5

Moore MJ, Scheel TKH, Luna JM, Park CY, Fak JJ, Nishiuchi E, Rice CM & Darnell RB (2015) miRNA–target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity. *Nat. Commun.* **6:** 1–17

Motamedi MR, Verdel A, Colmenares SU, Gerber SA, Gygi SP & Moazed D (2004) Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119:** 789–802

Motta J-P, Martin L & Vergnolle N (2011) Proteases/Antiproteases in Inflammatory Bowel Diseases. In *Proteases and Their Receptors in Inflammation*, Vergnolle N & Chignard M (eds) pp 173–215. Springer, Basel

Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* **420:** 520–562

Moyer TJ, Kato Y, Abraham W, Chang JYH, Kulp DW, Watson N, Turner HL, Menis S, Abbott RK, Bhiman JN, Melo MB, Simon HA, Herrera-De la Mata S, Liang S, Seumois G, Agarwal Y, Li N, Burton DR, Ward AB, Schief WR, et al (2020) Engineered immunogen binding to alum adjuvant enhances humoral immunity. *Nat. Med.* **26:** 430–440

Mulcahy LA, Pink RC & Carter DRF (2014) Routes and mechanisms of extracellular vesicle uptake. *J. Extracell. Vesicles***:** 3

Nagao M, Fujisawa T, Ihara T & Kino Y (2016) Highly increased levels of IgE antibodies to vaccine components in children with influenza vaccine–associated anaphylaxis. *J. Allergy Clin. Immunol.* **137:** 861–867

Nakazawa T, Fazal Khan A, Yasueda H, Saito A, Fukutomi Y, Takai T, Zaman K, Yunus M & Takeuchi H (2013) Immunization of Rabbits with Nematode Ascaris lumbricoides Antigens Induces Antibodies Cross-Reactive to House Dust Mite Dermatophagoides farinae Antigens. *Biosci. Biotechnol. Biochem.* **77:** 145–150

Nikonorova IA, Wang J, Cope AL, Tilton P, Power KM, Walsh JD, Akella JS, Krauchunas AR, Shah P & Barr MM (2021) Tracking extracellular vesicle (EV) cargo as a platform for studying EVomics, signaling, and targeting in vivo. *bioRxiv***:** 2021.09.23.461577

Nisbet AJ, McNeilly TN, Wildblood LA, Morrison AA, Bartley DJ, Bartley Y, Longhi C, McKendrick IJ, Palarea-Albaladejo J & Matthews JB (2013) Successful immunization against a parasitic nematode by vaccination with recombinant proteins. *Vaccine* **31:** 4017–4023

Nisbet AJ, Meeusen EN, González JF & Piedrafita DM (2016) Immunity to Haemonchus contortus and Vaccine Development. *Adv. Parasitol.* **93:** 353–396

Nottrott S, Simard MJ & Richter JD (2006) Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat. Struct. Mol. Biol.* **13:** 1108–1114

Nyström EEL, Birchenough GMH, Post S van der, Arike L, Gruber AD, Hansson GC & Johansson MEV (2018) Calcium-activated Chloride Channel Regulator 1 (CLCA1) Controls Mucus Expansion in Colon by Proteolytic Activity. *EBioMedicine* **33:** 134–143

Oikonomou G & Shaham S (2011) The Glia of Caenorhabditis elegans. *Glia* **59:** 1253–1263

Okoye IS, Coomes SM, Pelly VS, Czieso S, Papayannopoulos V, Tolmachova T, Seabra MC & Wilson MS (2014) MicroRNA-Containing T-Regulatory-Cell-Derived Exosomes Suppress Pathogenic T Helper 1 Cells. *Immunity* **41:** 89–103

Olsen PH & Ambros V (1999) The lin-4 Regulatory RNA Controls Developmental Timing in Caenorhabditis elegans by Blocking LIN-14 Protein Synthesis after the Initiation of Translation. *Dev. Biol.* **216:** 671–680

Ørom UA, Nielsen FC & Lund AH (2008) MicroRNA-10a Binds the 5′UTR of Ribosomal Protein mRNAs and Enhances Their Translation. *Mol. Cell* **30:** 460–471

Orr A, Quagraine J, Suwondo P, George S, Harrison L, Dornas F, Evans B, Caccone A, Humphries D, Wilson M & Cappello M (2019) Genetic Markers of Benzimidazole

Resistance among Human Hookworms (Necator americanus) in Kintampo North Municipality, Ghana. *Am. J. Trop. Med. Hyg.* **100:** 351–356

Osbourn M, Soares DC, Vacca F, Ivens AC, Maizels RM, McSorley Correspondence HJ, Suzanne Cohen E, Scott IC, Gregory WF, Smyth DJ, Toivakka M, Kemter AM, le Bihan T, Wear M, Hoving D, Filbey KJ, Hewitson JP, Henderson H, Gonzà lez-Cı A, Errington C, et al (2017) HpARI Protein Secreted by a Helminth Parasite Suppresses Interleukin-33 Article HpARI Protein Secreted by a Helminth Parasite Suppresses Interleukin-33. *Immunity* **47:** 739–751

Pak J & Fire A (2007) Distinct populations of primary and secondary effectors during RNAi in C. elegans. *Science* **315:** 241–244

Pal-Bhadra M, Leibovitch BA, Gandhi SG, Rao M, Bhadra U, Birchler JA & Elgin SCR (2004) Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. *Science* **303:** 669–672

Pare JM, Tahbaz N, López-Orozco J, LaPointe P, Lasko P & Hobman TC (2009) Hsp90 Regulates the Function of Argonaute 2 and Its Recruitment to Stress Granules and P-Bodies. *Mol. Biol. Cell* **20:** 3273–3284

Park JH & Shin C (2014) MicroRNA-directed cleavage of targets: mechanism and experimental approaches. *BMB Rep* **47:** 417–423

Park MS, Phan HD, Busch F, Hinckley SH, Brackbill JA, Wysocki VH & Nakanishi K (2017) Human Argonaute3 has slicer activity. *Nucleic Acids Res.* **45:** 11867–11877

Parker JS & Barford D (2006) Argonaute: a scaffold for the function of short regulatory RNAs. *Trends Biochem. Sci.* **31:** 622–630

Parker JS, Roe SM & Barford D (2004) Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J.* **23:** 4727–4737

Parker JS, Roe SM & Barford D (2005) Structural insights into mRNA recognition from a PIWI domain – siRNA guide complex. *Nature* **434:** 663–666

Pearce N, Sunyer J, Cheng S, Chinn S, Björkstén B, Burr M, Keil U, Anderson HR & Burney P (2000) Comparison of asthma prevalence in the ISAAC and the ECRHS. ISAAC Steering Committee and the European Community Respiratory Health Survey. International Study of Asthma and Allergies in Childhood. *Eur. Respir. J.* **16:** 420–426

Pearson MS, Pickering DA, McSorley HJ, Bethony JM, Tribolet L, Dougall AM, Hotez PJ & Loukas A (2012) Enhanced Protective Efficacy of a Chimeric Form of the Schistosomiasis Vaccine Antigen Sm-TSP-2. *PLoS Negl. Trop. Dis.* **6:** e1564

Pearson MS, Pickering DA, Tribolet L, Cooper L, Mulvenna J, Oliveira LM, Bethony JM, Hotez PJ & Loukas A (2010) Neutralizing Antibodies to the Hookworm Hemoglobinase Na-APR-1: Implications for a Multivalent Vaccine against Hookworm Infection and Schistosomiasis. *J. Infect. Dis.* **201:** 1561–1569

Pedersen AB & Greives TJ (2008) The interaction of parasites and resources cause crashes in a wild mouse population. *J. Anim. Ecol.* **77:** 370–377

Perera DJ & Ndao M (2021) Promising Technologies in the Field of Helminth Vaccines. *Front. Immunol.* **12:** 711650

Pesce J, Ramalingam T, Mentink-Kane M, Wilson M, El Kasmi K, Smith A, Thompson R, Cheever A, Murray P & Wynn T (2009) Arginase-1-expressing macrophages suppress Th2 cytokine-driven inflammation and fibrosis. *PLoS Pathog.* **5:** e1000371

Peters D, Radine C, Reese A, Budach W, Sohn D, Jänicke RU & Wek RC (2017) The DEAD-box RNA helicase DDX41 is a novel repressor of p21 WAF1/CIP1 mRNA translation. *J. Biol. Chem.* **292:** 8331–41

Peters L & Meister G (2007) Argonaute Proteins: Mediators of RNA Silencing. *Mol. Cell* **26:** 611–23

Petersen CP, Bordeleau ME, Pelletier J & Sharp PA (2006) Short RNAs repress translation after initiation in mammalian cells. *Mol. Cell* **21:** 533–542

Peterson SM, Thompson JA, Ufkin ML, Sathyanarayana P, Liaw L & Congdon CB (2014) Common features of microRNA target prediction tools. *Front. Genet.* **5:** 23

Pfaff AW, Schulz-Key H, Soboslay PT, Taylor DW, MacLennan K & Hoffmann WH (2002) Litomosoides sigmodontis cystatin acts as an immunomodulator during experimental filariasis. *Int. J. Parasitol.* **32:** 171–178

Pfaff J & Meister G (2013) Argonaute and GW182 proteins: an effective alliance in gene silencing. *Biochem. Soc. Trans.* **41:** 855–860

Pietrzak B, Tomela K, Olejnik-Schmidt A, Mackiewicz A & Schmidt M (2020) Secretory IgA in Intestinal Mucosal Secretions as an Adaptive Barrier against Microbial Cells. *Int. J. Mol. Sci.* **21:** 1–15

Pinzón N, Bertrand S, Subirana L, Busseau I, Escrivá H & Seitz H (2019) Functional lability of RNA-dependent RNA polymerases in animals. *PLoS Genet.* **15:** e1007915

Place RF, Li LC, Pookot D, Noonan EJ & Dahiya R (2008) MicroRNA-373 induces expression of genes with complementary promoter sequences. *PNAS* **105:** 1608–1613

Pollard AJ & Bijker EM (2020) A guide to vaccinology: from basic principles to new developments. *Nat. Rev. Immunol.* **21:** 83–100

Portnoy V, Lin SHS, Li KH, Burlingame A, Hu ZH, Li H & Li LC (2016) saRNA-guided Ago2 targets the RITA complex to promoters to stimulate transcription. *Cell Res.* **26:** 320–335

Pottash AE, Kuffner C, Noonan-Shueh M & Jay SM (2019) Protein-based vehicles for biomimetic RNAi delivery. *J. Biol. Eng.* **13:** 1–13

Powell DR (2019) Degust: interactive RNA-seq analysis. Available at: https://degust.erc.monash.edu/

Pratt AJ & MacRae IJ (2009) The RNA-induced Silencing Complex: A Versatile Gene-silencing Machine. *J. Biol. Chem.* **284:** 17897

Proudfoot L (2004) Parasitic helminths tip the balance: potential anti-inflammatory therapies. *Immunology* **113:** 438

Prud'homme GJ, Glinka Y, Lichner Z, Yousef GM, Prud'homme GJ, Glinka Y, Lichner Z & Yousef GM (2016) Neuropilin-1 is a receptor for extracellular miRNA and AGO2/miRNA complexes and mediates the internalization of miRNAs that modulate cell function. *Oncotarget* **7:** 68057–68071

Public Health Agency (2017) The hexavalent DTaP/IPV/Hib/HepB (6 in 1) combination vaccine Belfast Available at: https://www.publichealth.hscni.net/sites/default/files/The hexavalent_DTaP_IPV_Hib_HepB_combo_factsheet_0.pdf [Accessed December 2, 2021]

Pullan RL, Smith JL, Jasrasaria R & Brooker SJ (2014) Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. *Parasit. Vectors* **7:** 37

Quinlan AR & Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841–842

Quintana JF, Kumar S, Ivens A, Chow FWN, Hoy AM, Fulton A, Dickinson P, Martin C, Taylor M, Babayan SA & Buck AH (2019) Comparative analysis of small RNAs released by the filarial nematode Litomosoides sigmodontis in vitro and in vivo. *PLoS Negl. Trop. Dis.* **13:** e0007811

Quintana JF, Makepeace BL, Babayan SA, Ivens A, Pfarr KM, Blaxter M, Debrah A, Wanji S, Ngangyung HF, Bah GS, Tanya VN, Taylor DW, Hoerauf A & Buck AH (2015) Extracellular Onchocerca-derived small RNAs in host nodules and blood. *Parasites and Vectors* **8:** 1–11

R Core Team (2021) R: A language and environment for statistical computing. Available at: https://www.r-project.org/

Raina M & Ibba M (2014) tRNAs as regulators of biological processes. *Front. Genet.* **5:** 171

Rapin A, Chuat A, Lebon L, Zaiss MM, Marsland BJ & Harris NL (2020) Infection with a small intestinal helminth, Heligmosomoides polygyrus bakeri, consistently alters microbial communities throughout the murine small and large intestine. *Int. J. Parasitol.* **50:** 35–46

Rausch S, Midha A, Kuhring M, Affinass N, Radonic A, Kühl AA, Bleich A, Renard BY & Hartmann S (2018) Parasitic Nematodes Exert Antimicrobial Activity and Benefit From Microbiota-Driven Support for Host Immune Regulation. *Front. Immunol.* **9:** 2282

Reed KJ, Svendsen JM, Brown KC, Montgomery BE, Marks TN, Vijayasarathy T, Parker DM, Nishimura EO, Updike DL & Montgomery TA (2020) Widespread roles for piRNAs and WAGO-class siRNAs in shaping the germline transcriptome of Caenorhabditis elegans. *Nucleic Acids Res.* **48:** 1811–1827

Reinhart BJ & Bartel DP (2002) Small RNAs correspond to centromere heterochromatic repeats. *Science* **297:** 1831

Reitz M, Hartmann W, Rüdiger N, Orinska Z, Brunn M-L & Breloer M (2018) Interleukin-9 promotes early mast cell-mediated expulsion of Strongyloides ratti but is dispensable for generation of protective memory. *Sci. Rep.* **8:** 8636

Reynolds LA, Filbey KJ & Maizels RM (2012) Immunity to the model intestinal helminth parasite Heligmosomoides polygyrus. *Semin. Immunopathol.* **34:** 829–846

Ricafrente A, Nguyen H, Tran N & Donnelly S (2021) An Evaluation of the Fasciola hepatica miRnome Predicts a Targeted Regulation of Mammalian Innate Immune Responses. *Front. Immunol.* **11:** 608686

Ricci EP, Limousin T, Soto-Rifo R, Rubilar PS, Decimo D & Ohlmann T (2013) miRNA repression of translation in vitro takes place during 43S ribosomal scanning. *Nucleic Acids Res.* **41:** 586–598

Rivas F V, Tolia NH, Song J-J, Aragon JP, Liu J, Hannon GJ & Joshua-Tor L (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat. Struct. Mol. Biol.* **12:** 340–349

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G & Mesirov JP (2011) Integrative genomics viewer. *Nat. Biotechnol.* **29:** 24–26

Robinson M, Alvarado R, To J, Hutchinson A, Dowdell S, Lund M, Turnbull L, Whitchurch C, O'Brien B, Dalton J & Donnelly S (2012) A helminth cathelicidin-like protein suppresses antigen processing and presentation in macrophages via inhibition of lysosomal vATPase. *FASEB J.* **26:** 4614–4627

Robinson M, Wahid F, Behnke JM & Gilbert F (1989) Immunological relationships during primary infection with Heligmosomoides polygyrus (Nematospiroides dubius): dose-dependent expulsion of adult worms. *Parasitology* **98:** 115–124

Robinson MD, McCarthy DJ & Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:** 139

Rocca G La, Olejniczak SH, González AJ, Briskin D, Vidigal JA, Spraggon L, DeMatteo RG, Radler MR, Lindsten T, Ventura A, Tuschl T, Leslie CS & Thompson CB (2015) In vivo, Argonaute-bound microRNAs exist predominantly in a reservoir of low molecular weight complexes not associated with mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **112:** 767–772

Rolot M & Dewals BG (2018) Macrophage Activation and Functions during Helminth Infection: Recent Advances from the Laboratory Mouse. *J. Immunol. Res.* **2018:** 2790627

Rosenzvit MC, Camicia F, Kamenetzky L, Muzulin PM & Gutierrez AM (2006) Identification and intra-specific variability analysis of secreted and membrane-bound proteins from Echinococcus granulosus. *Parasitol. Int.* **55:** S63–S67

Ruby JG, Jan CH & Bartel DP (2007) Intronic microRNA precursors that bypass Drosha processing. *Nature* **448:** 83–86

Ryazansky S, Kulbachinskiy A & Aravin AA (2018) The Expanded Universe of Prokaryotic Argonaute Proteins. *MBio* **9:** e01935-18

Sadik N, Cruz L, Gurtner A, Rodosthenous RS, Dusoswa SA, Ziegler O, Van Solinge TS, Wei Z, Salvador-Garicano AM, Gyorgy B, Broekman M & Balaj L (2018) Extracellular RNAs: A New Awareness of Old Perspectives. *Methods Mol. Biol.* **1740:** 1–15

Saksouk N, Simboeck E & Déjardin J (2015) Constitutive heterochromatin formation and transcription in mammals. *Epigenetics and Chromatin* **8:** 1–17

Sanan-Mishra N, Abdul Kader Jailani A, Mandal B & Mukherjee SK (2021) Secondary siRNAs in Plants: Biosynthesis, Various Functions, and Applications in Virology. *Front. Plant Sci.* **12:** 610283

Sandberg K, Samson WK & Ji H (2013) Decoding noncoding RNA: the long and short of it. *Circ. Res.* **113:** 240–241

Sarkar I, Garg R & van Drunen Littel-van den Hurk S (2019) Selection of adjuvants for vaccines targeting specific pathogens. *Expert Rev. Vaccines* **18:** 505–521

Sarshad AA, Juan AH, Muler AIC, Anastasakis DG, Wang X, Genzor P, Feng X, Tsai PF, Sun HW, Haase AD, Sartorelli V & Hafner M (2018) Argonaute-miRNA Complexes Silence Target mRNAs in the Nucleus of Mammalian Stem Cells. *Mol. Cell* **71:** 1040-1050.e8

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P & Cardona A (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9:** 676–682

Schneider CA, Rasband WS & Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9:** 671–675

Schönemeyer A, Lucius R, Sonnenburg B, Brattig N, Sabat R, Schilling K, Bradley J & Hartmann S (2001) Modulation of human T cell responses and macrophage functions by onchocystatin, a secreted protein of the filarial nematode Onchocerca volvulus. *J. Immunol.* **167:** 3207–3215

Schwarz DS, Hutvágner G, Du T, Xu Z, Aronin N & Zamore PD (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115:** 199–208

Seggerson K, Tang L & Moss EG (2002) Two genetic circuits repress the Caenorhabditis elegans heterochronic gene lin-28 after translation initiation. *Dev. Biol.* **243:** 215–225

Seitz H, Tushir JS & Zamore PD (2011) A 5′-uridine amplifies miRNA/miRNA* asymmetry in Drosophila by promoting RNA-induced silencing complex formation. *Silence* **2:** 4

Sen GL & Blau HM (2005) Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nat. Cell Biol.* **7:** 633–636

Seth M, Shirayama M, Gu W, Ishidate T, Conte D & Mello CC (2013) The C. elegans CSR-1 argonaute pathway counteracts epigenetic silencing to promote germline gene expression. *Dev. Cell* **27:** 656–663

Sharpe C, Thornton DJ & Grencis RK (2018) A sticky end for gastrointestinal helminths; the role of the mucus barrier. *Parasite Immunol.* **40:** 1–10

Sheldon J & Soriano V (2008) Hepatitis B virus escape mutants induced by antiviral therapy. *J. Antimicrob. Chemother.* **61:** 766–768

Sheu-Gruttadauria J & MacRae IJ (2017) Structural Foundations of RNA Silencing by Argonaute. *J. Mol. Biol.* **429:** 2619–2639

Shore S, Henderson JM, Lebedev A, Salcedo MP, Zon G, Mccaffrey AP, Paul N & Hogrefe RI (2016) Small RNA Library Preparation Method for Next-Generation Sequencing Using Chemical Modifications to Prevent Adapter Dimer Formation. *PLoS One* **11:** e0167009

Shurtleff MJ, Temoche-Diaz MM, Karfilis K V., Ri S & Schekman R (2016) Y-box protein 1 is required to sort microRNAs into exosomes in cells and in a cell-free reaction. *Elife* **5:** e19276

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD & Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7:** 539

Sijen T, Fleenor J, Simmer F, Thijssen KL, Parrish S, Timmons L, Plasterk RHA & Fire A (2001) On the Role of RNA Amplification in dsRNA-Triggered Gene Silencing. *Cell* **107:** 465–476

Silva L, Egea J, Villanueva L, Ruiz M, Llopiz D, Repáraz D, Aparicio B, Lasarte-cia A, Lasarte JJ, de Galarreta MR, Lujambio A, Sangro B & Sarobe P (2020) Cold-Inducible RNA Binding Protein as a Vaccination Platform to Enhance Immunotherapeutic Responses Against Hepatocellular Carcinoma. *Cancers (Basel).* **12:** 1–18

Silva NR de, Brooker S, Hotez PJ, Montresor A, Engels D & Savioli L (2003) Soil-transmitted helminth infections: updating the global picture. *Trends Parasitol.* **19:** 547–551

Singh M, Cornes E, Li B, Quarato P, Bourdon L, Dingli F, Loew D, Proccacia S & Cecere G (2021) Translation and codon usage regulate Argonaute slicer activity to trigger small RNA biogenesis. *Nat. Commun.* **12:** 3492

Siomi MC, Sato K, Pezic D & Aravin AA (2011) PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* **12:** 246–258

Smith WD, Van Wyk JA & Van Strijp MF (2001) Preliminary observations on the potential of gut membrane proteins of Haemonchus contortus as candidate vaccine antigens in sheep on naturally infected pasture. *Vet. Parasitol.* **98:** 285–297

Smyth DJ, Ren B, White MPJ, McManus C, Webster H, Shek V, Evans C, Pandhal J, Fields F, Maizels RM & Mayfield S (2021) Oral delivery of a functional algal-expressed TGF-β mimic halts colitis in a murine DSS model. *J. Biotechnol.* **340:** 1–12

Sorobetea D, Svensson-Frej M & Grencis R (2018) Immunity to gastrointestinal nematode infections. *Mucosal Immunol.* **11:** 304–315

Sotillo J, Pearson M, Potriquet J, Becker L, Pickering D, Mulvenna J & Loukas A (2016) Extracellular vesicles secreted by Schistosoma mansoni contain protein vaccine candidates. *Int. J. Parasitol.* **46:** 1–5

Stein M, Keshav S, Harris N & Gordon S (1992) Interleukin 4 potently enhances murine macrophage mannose receptor activity: a marker of alternative immunologic macrophage activation. *J. Exp. Med.* **176:** 287–292

Steinmann P, Utzinger J, Du ZW & Zhou XN (2010) Multiparasitism: A Neglected Reality on Global, Regional and Local Scale. *Adv. Parasitol.* **73:** 21–50

Stout RFJ, Verkhratsky A & Parpura V (2014) Caenorhabditis elegans glia modulate neuronal activity and behavior. *Front. Cell. Neurosci.* **8:** 67

Strachan DP (1989) Hay fever, hygiene, and household size. *BMJ* **299:** 1259–1260

Strunz EC, Addiss DG, Stocks ME, Ogden S, Utzinger J & Freeman MC (2014) Water, Sanitation, Hygiene, and Soil-Transmitted Helminth Infection: A Systematic Review and Meta-Analysis. *PLOS Med.* **11:** e1001620

Sugiyama T, Cam H, Verdel A, Moazed D & Grewal SIS (2005) RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc. Natl. Acad. Sci.* **102:** 152–157

Sun Y, Liu G, Li Z, Chen Y, Liu Y, Liu B & Su Z (2013) Modulation of dendritic cell function and immune response by cysteine protease inhibitor from murine nematode parasite Heligmosomoides polygyrus. *Immunology* **138:** 370–381

Sundby AE, Molnar RI & Claycomb JM (2021) Connecting the Dots: Linking Caenorhabditis elegans Small RNA Pathways and Germ Granules. *Trends Cell Biol.* **31:** 387–401

Sutherland TE, Rückerl D, Logan N, Duncan S, Wynn TA & Allen JE (2018) Ym1 induces RELMα and rescues IL-4Rα deficiency in lung repair during nematode infection. *PLOS Pathog.* **14:** e1007423

Svendsen JM & Montgomery TA (2018) piRNA Rules of Engagement. *Dev. Cell* **44:** 657–658

Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin E V., Patel DJ & Van Der Oost J (2014) The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.* **21:** 743–753

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ & Mering C von (2019) STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47:** D607–D613

Tan GS, Garchow BG, Liu X, Yeung J, Morris JP, Cuellar TL, McManus MT & Kiriakidou M (2009) Expanded RNA-binding activities of mammalian Argonaute 2. *Nucleic Acids Res.* **37:** 7533–7545

Tarver JE, Donoghue PCJ & Peterson KJ (2012) Do miRNAs have a deep evolutionary history? *BioEssays* **34:** 857–866

Tay Y, Zhang J, Thomson AM, Lim B & Rigoutsos I (2008) MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **455:** 1124–1128

Taylor PJ, Hagen J, Faruqu FN, Al-Jamal KT, Quigley B, Beeby M, Selkirk ME & Sarkies P (2020) Trichinella spiralis secretes abundant unencapsulated small RNAs with potential effects on host gene expression. *Int. J. Parasitol.* **50:** 697–705

Teotia S, Singh D, Tang G, Teotia S, Singh D & Tang G (2017) DNA Methylation in Plants by microRNAs. In *Plant Epigenetics. RNA Technologies*, Rajewsky N Jurga S & Barciszewski J (eds) pp 247–262. Springer, Cham Available at: https://link.springer.com/chapter/10.1007/978-3-319-55520-1_13 [Accessed November 11, 2021]

Teschendorf C, Warrington K, Siemann D & Muzyczka N (2002) Comparison of the EF-1 alpha and the CMV promoter for engineering stable tumor cell lines using recombinant adeno-associated virus. *Anticancer Res.* **22:** 3325–3330

Threadgold LT (1963) The Tegument and Associated Structures of Fasciola Hepatica. *J. Cell Sci.* **s3**-**104:** 505–512

Tijsterman M, Okihara KL, Thijssen K & Plasterk RH. (2002) PPW-1, a PAZ/PIWI Protein Required for Efficient Germline RNAi, Is Defective in a Natural Isolate of C. elegans. *Curr. Biol.* **12:** 1535–1540

Tolia NH & Joshua-Tor L (2007) Slicer and the Argonautes. *Nat. Chem. Biol.* **3:** 36–43

Topuzoğullari M, Acar T, Pelit Arayici P, Uçar B, Uğurel E, Abamor EŞ, Arasoğlu T, Turgut-Balik D & Derman S (2020) An insight into the epitope-based peptide vaccine design strategy and studies against COVID-19. *Turkish J. Biol.* **44:** 215–227

Tran MH, Pearson MS, Bethony JM, Smyth DJ, Jones MK, Duke M, Don TA, McManus DP, Correa-Oliveira R & Loukas A (2006) Tetraspanins on the surface of Schistosoma mansoni are protective antigens against schistosomiasis. *Nat. Med.* **12:** 835–840

Tran N, Ricafrente A, To J, Lund M, Marques TM, Gama-Carvalho M, Cwiklinski K, Dalton JP & Donnelly S (2021) Fasciola hepatica hijacks host macrophage miRNA machinery to modulate early innate immune responses. *Sci. Rep.* **11:** 6712

Travers J & Rothenberg ME (2015) Eosinophils in mucosal immune responses. *Mucosal Immunol.* **8:** 464–475

Tribolet L, Cantacessi C, Pickering DA, Navarro S, Doolan DL, Trieu A, Fei H, Chao Y, Hofmann A, Gasser RB, Giacomin PR & Loukas A (2015) Probing of a human proteome microarray with a recombinant pathogen protein reveals a novel mechanism by which hookworms suppress B-cell receptor signaling. *J. Infect. Dis.* **211:** 416–425

Tritten L, Burkman E, Moorhead A, Satti M, Geary J, Mackenzie C & Geary T (2014a) Detection of Circulating Parasite-Derived MicroRNAs in Filarial Infections. *PLoS Negl. Trop. Dis.* **8:** e2971

Tritten L, O'Neill M, Nutting C, Wanji S, Njouendoui A, Fombad F, Kengne-Ouaffo J, Mackenzie C & Geary T (2014b) Loa loa and Onchocerca ochengi miRNAs detected in host circulation. *Mol. Biochem. Parasitol.* **198:** 14–17

Tsuboyama K, Tadakuma H & Tomari Y (2018) Conformational Activation of Argonaute by Distinct yet Coordinated Actions of the Hsp70 and Hsp90 Chaperone Systems. *Mol. Cell* **70:** 722-729.e4

Vacca F, Chauché C, Jamwal A, Hinchy EC, Heieis G, Webster H, Ogunkanbi A, Sekne Z, Gregory WF, Wear M, Perona-Wright G, Higgins MK, Nys JA, Cohen S & McSorley HJ (2020) A helminth-derived suppressor of ST2 blocks allergic responses. *Elife* **9:** e54017

Vagin V V., Sigova A, Li C, Seitz H, Gvozdev V & Zamore PD (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313:** 320–324

Valadi H, Ekström K, Bossios A, Sjöstrand M, Lee JJ & Lötvall JO (2007) Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.* **9:** 654–659

Valdembri D, Caswell PT, Anderson KI, Schwarz JP, König I, Astanina E, Caccavari F, Norman JC, Humphries MJ, Bussolino F & Serini G (2009) Neuropilin-1/GIPC1

Signaling Regulates α5β1 Integrin Traffic and Function in Endothelial Cells. *PLoS Biol.* **7:** e1000025

Valkov N & Das S (2020) Y RNAs: Biogenesis, Function and Implications for the Cardiovascular System. *Adv. Exp. Med. Biol.* **1229:** 327–342

Vastenhouw NL, Fischer SEJ, Robert VJP, Thijssen KL, Fraser AG, Kamath RS, Ahringer J & Plasterk RHA (2003) A Genome-Wide Screen Identifies 27 Genes Involved in Transposon Silencing in C. elegans. *Curr. Biol.* **13:** 1311–1316

Vercruysse J, Charlier J, Dijk J Van, Morgan ER, Geary T, Samson-Himmelstjerna G von & Claerebout E (2018) Control of helminth ruminant infections by 2030. *Parasitology* **145:** 1655–1664

Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SIS & Moazed D (2004) RNAi-Mediated Targeting of Heterochromatin by the RITS Complex. *Science* **303:** 672–676

Vidal K, Grosjean I, evillard JP, Gespach C & Kaiserlian D (1993) Immortalization of mouse intestinal epithelial cells by the SV40-large T gene. Phenotypic and immune characterization of the MODE-K cell line. *J. Immunol. Methods* **166:** 63–73

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS & Martienssen RA (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297:** 1833–1837

Wakelin D (1996) Helminths: Pathogenesis and Defenses. In *Medical Microbiology*, Baron S (ed) Galveston, Texas: University of Texas Medical Branch at Galveston Available at: https://www.ncbi.nlm.nih.gov/books/NBK8191/ [Accessed November 1, 2021]

Walk ST, Blum AM, Ewing SAS, Weinstock J V. & Young VB (2010) Alteration of the murine gut microbiota during infection with the parasitic helminth Heligmosomoides polygyrus. *Inflamm. Bowel Dis.* **16:** 1841–1849

Walker JA & McKenzie ANJ (2017) TH2 cell development and function. *Nat. Rev. Immunol.* **18:** 121–133

Wang HL V., Dinwiddie BL, Lee H & Chekanova JA (2015) Stress-induced endogenous siRNAs targeting regulatory intron sequences in Brachypodium. *RNA* **21:** 145–163

Wang S, Wu W & Claret FX (2017a) Mutual regulation of microRNAs and DNA methylation in human cancers. *Epigenetics* **12:** 187–197

Wang T, Fan L, Watanabe Y, McNeill PD, Moulton GG, Bangur C, Fanger GR, Okada M, Inoue Y, Persing DH & Reed SG (2003) L523S, an RNA-binding protein as a potential therapeutic target for lung cancer. *Br. J. Cancer* **88:** 887–894

Wang Z, Xi J, Hao X, Deng W, Liu J, Wei C, Gao Y, Zhang L & Wang H (2017b) Red blood cells release microparticles containing human argonaute 2 and miRNAs to target genes of Plasmodium falciparum. *Emerg. Microbes Infect.* **6:** e75

Wang ZB & Xu J (2020) Better Adjuvants for Better Vaccines: Progress in Adjuvant Delivery Systems, Modifications, and Adjuvant-Antigen Codelivery. *Vaccines* **8:** 128

Wassenegger M, Heimes S, Riedel L & Sänger HL (1994) RNA-directed de novo methylation of genomic sequences in plants. *Cell* **76:** 567–576

Wedeles CJ (2018) A CSR-1 guided H3K36me3 Signature Maintains Germline Gene Expression Programs In Caenorhabditis elegans. Available at: http://hdl.handle.net/1807/95709 [Accessed November 15, 2021]

Wedeles CJ, Wu MZ & Claycomb JM (2013) Protection of germline gene expression by the C. elegans Argonaute CSR-1. *Dev. Cell* **27:** 664–671

Wedeles CJ, Wu MZ & Claycomb JM (2014) Silent no more: Endogenous small RNA pathways promote gene expression. *Worm* **3:** e28641

Weiberg A, Wang M, Lin F-M, Zhao H, Zhang Z, Kaloshia I, Huang H-D & Jin H (2013) Fungal Small RNAs Suppress Plant Immunity by Hijacking Host RNA Interference Pathways. *Science* **342:** 118–123

Weick EM & Miska EA (2014) piRNAs: from biogenesis to function. *Development* **141:** 3458–3471

White MPJ, McManus CM & Maizels RM (2020) Regulatory T-cells in helminth infection: induction, function and therapeutic potential. *Immunology* **160:** 248–260

White MPJ, Smyth DJ, Cook L, Ziegler SF, Levings MK & Maizels RM (2021) The parasite cytokine mimic Hp-TGM potently replicates the regulatory effects of TGF-β on murine CD4+ T cells. *Immunol. Cell Biol.* **99:** 848–864

Whitehead B, Boysen AT, Mardahl M & Nejsum P (2020) Unique glycan and lipid composition of helminth-derived extracellular vesicles may reveal novel roles in host-parasite interactions. *Int. J. Parasitol.* **5:** 647–654

Whitman JD, Sakanari JA & Mitreva M (2021) Areas of Metabolomic Exploration for Helminth Infections. *ACS Infect. Dis.* **7:** 206

WHO (2020) Soil-transmitted helminth infections. Available at: https://www.who.int/news-room/fact-sheets/detail/soil-transmitted-helminth-infections [Accessed November 1, 2021]

Wightman B, Ha I & Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75:** 855–862

Williamson AL, Brindley PJ, Abbenante G, Prociv P, Berry C, Girdwood K, Pritchard DI, Fairlie DP, Hotez PJ, Dalton JP & Loukas A (2002) Cleavage of hemoglobin by hookworm cathepsin D aspartic proteases and its potential contribution to host specificity. *FASEB J.* **16:** 1458–1460

Willkomm S, Zander A, Gust A, Grohmann D, Klenk H-P, Adams MWW & Garrett RA (2015) A Prokaryotic Twist on Argonaute Function. *Life* **5:** 538–553

Win SY, Win M, Thwin EP, Htun LL, Hmoon MM, Chel HM, Thaw YN, Soe NC, Phyo TT, Thein SS, Khaing Y, Than AA & Bawm S (2020) Occurrence of gastrointestinal parasites in small ruminants in the central part of Myanmar. *J. Parasitol. Res.* **2020:** 8826327

Wit J, Dilks CM & Andersen EC (2021) Complementary Approaches with Free-living and Parasitic Nematodes to Understanding Anthelmintic Resistance. *Trends Parasitol.* **37:** 240–250

Wolday D, Gebrecherkos T, Arefaine ZG, Kiros YK, Gebreegzabher A, Tasew G, Abdulkader M, Abraha HE, Desta AA, Hailu A, Tollera G, Abdella S, Tesema M, Abate

E, Endarge KL, Hundie TG, Miteku FK, Urban BC, Schallig HHDF, Harris VC, et al (2021) Effect of co-infection with intestinal parasites on COVID-19 severity: A prospective observational cohort study. *EClinicalMedicine* **39:**

Wu L, Zhou H, Zhang Q, Zhang J, Ni F, Liu C & Qi Y (2010) DNA Methylation Mediated by a MicroRNA Pathway. *Mol. Cell* **38:** 465–475

Wu X, Pan Y, Fang Y, Zhang J, Xie M, Yang F, Yu T, Ma P, Li W & Shu Y (2020) The Biogenesis and Functions of piRNAs in Human Diseases. *Mol. Ther. Nucleic Acids* **21:** 108–120

Wu Z, Wang L, Tang Y & Sun X (2017) Parasite-Derived Proteins for the Treatment of Allergies and Autoimmune Diseases. *Front. Microbiol.* **8:** 2164

Xu F, Feng X, Chen X, Weng C, Yan Q, Xu T, Hong M & Guang S (2018) A Cytoplasmic Argonaute Protein Promotes the Inheritance of RNAi. *Cell Rep.* **23:** 2482–2494

Yang J, Zhao Y, Kalita M, Li X, Jamaluddin M, Tian B, Edeh CB, Wiktorowicz JE, Kudlicki A & Brasier AR (2015) Systematic Determination of Human Cyclin Dependent Kinase (CDK)-9 Interactome Identifies Novel Functions in RNA Splicing Mediated by the DEAD Box (DDX)-5/17 RNA Helicases. *Mol. Cell. Proteomics* **14:** 2701–2721

Yarosh CA, Iacona JR, Lutz CS & Lynch KW (2015) PSF: nuclear busy-body or nuclear facilitator? *WIREs RNA* **6:** 351–367

Yi R, Qin Y, Macara IG & Cullen BR (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17:** 3011–3016

Yigit E, Batista PJ, Bei Y, Pang KM, Chen CCG, Tolia NH, Joshua-Tor L, Mitani S, Simard MJ & Mello CC (2006) Analysis of the C. elegans Argonaute Family Reveals that Distinct Argonautes Act Sequentially during RNAi. *Cell* **127:** 747–757

Youn J-Y, Dunham WH, Hong SJ, Knight JDR, Bashkurov M, Chen GI, Bagci H, Rathod B, MacLeod G, Eng SWM, Angers S, Morris Q, Fabian M, Côté J-F & Gingras A-C (2018) High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies. *Mol. Cell* **69:** 517-532.e11

Youngman EM & Claycomb JM (2014) From early lessons to new frontiers: The worm as a treasure trove of small RNA biology. *Front. Genet.* **5:** 416

Zagoskin M, Wang J, Neff AT, Veronezi GMB & Davis RE (2021) Nematode Small RNA Pathways in the Absence of piRNAs. *bioRxiv***:** 2021.07.23.453445

Zaiss MM, Rapin A, Lebon L, Dubey LK, Mosconi I, Sarter K, Piersigilli A, Menin L, Walker AW, Rougemont J, Paerewijck O, Geldhof P, McCoy KD, Macpherson AJ, Croese J, Giacomin PR, Loukas A, Junt T, Marsland BJ & Harris NL (2015) The Intestinal Microbiota Contributes to the Ability of Helminths to Modulate Allergic Inflammation. *Immunity* **43:** 998–1010

Zakeri A, Hansen EP, Andersen SD, Williams AR & Nejsum P (2018) Immunomodulation by Helminths: Intracellular Pathways and Extracellular Vesicles. *Front. Immunol.* **9:** 2349

Zawawi A & Else KJ (2020) Soil-Transmitted Helminth Vaccines: Are We Getting Closer? *Front. Immunol.* **11:** 576748

Zhan B, Beaumier CM, Briggs N, Jones KM, Keegan BP, Bottazzi ME & Hotez PJ (2014) Advancing a multivalent 'Pan-anthelmintic' vaccine against soil-transmitted nematode infections. *Expert Rev. Vaccines* **13:** 321–331

Zhang C (2009) Novel functions for small RNA molecules. *Curr. Opin. Mol. Ther.* **11:** 641–651

Zhang C, Montgomery TA, Fischer SEJ, Garcia SMDA, Riedel CG, Fahlgren N, Sullivan CM, Carrington JC & Ruvkun G (2012) The Caenorhabditis elegans RDE-10/RDE-11 Complex Regulates RNAi by Promoting Secondary siRNA Amplification. *Curr. Biol.* **22:** 881–890

Zhang H, Sheng C, Yin Y, Wen S, Yang G, Cheng Z & Zhu Q (2015) PABPC1 interacts with AGO2 and is responsible for the microRNA mediated gene silencing in high grade hepatocellular carcinoma. *Cancer Lett.* **367:** 49–57

Zhang Y, Fan M, Geng G, Liu B, Huang Z, Luo H, Zhou J, Guo X, Cai W & Zhang H (2014) A novel HIV-1-encoded microRNA enhances its viral replication by targeting the TATA box region. *Retrovirology* **11:** 23

Zhou H & Rigoutsos I (2014) MiR-103a-3p targets the 5' UTR of GPRC5A in pancreatic cells. *RNA* **20:** 1431–1439

Zhu L, Liu J, Dao J, Lu K, Li H, Gu H, Liu J, Feng X & Cheng G (2016) Molecular characterization of S. japonicum exosome-like vesicles reveals their regulatory roles in parasite-host interactions. *Sci. Rep.* **6:** 1–14

Supplementary Tables

All the Supplementary Tables are found at:

https://drive.google.com/drive/folders/1RkAksgNF_RWW09Vg8DPejy9YjZxfm1ON?usp=sharing

Supplementary Table 3.1 | The sRNA guides deemed exclusive to the non-vesicular exWAGO when compared to the vesicular exWAGO. The table shows the read sequence, annotation, fold change of sequences found in SUP IPs relative to EV IPs, average expression, the FDR, p-value, and the number of reads as counts and as CPM.

Supplementary Table 3.2 | The sRNA guides deemed exclusive to the vesicular exWAGO when compared to the non-vesicular exWAGO. The table shows the read sequence, annotation, fold change of sequences found in SUP IPs relative to EV IPs, average expression, the FDR, p-value, and the number of reads as counts and as CPM.

Supplementary Table 4.1 | The 5,396 clusters identified as putative mAGO2 targets. The table details the clusters identified and includes information on canonical and non-canonical mapping, the number of reads as counts and as CPM. The file also includes a summary of the mapping information, a list of the annotated genes used for gene ontology analysis and the cluster sequences that were used for generation of the weblogo.

Supplementary Table 4.2 | The 911 clusters identified as putative exWAGO targets. The table details the clusters identified and includes information on canonical and non-canonical mapping, the number of reads as counts and as CPM. The file also includes a summary of the mapping information, a list of the annotated genes used for gene ontology analysis and the cluster sequences that were used for generation of the weblogo.

Supplementary Table 4.3 | Gene Ontology biological process term analysis of the annotated clusters identified in the 911 putative exWAGO targets. The table shows the biological processes identified (FDR < 0.5), the number of gene expected and identified for each term, the raw p-value and the FDR.

Supplementary Table 4.4 | Gene Ontology biological process term analysis of the annotated clusters identified in the 5,396 putative mAGO2 targets. The table shows the biological processes identified (FDR < 0.5), the number of gene expected and identified for each term, the raw p-value and the FDR.

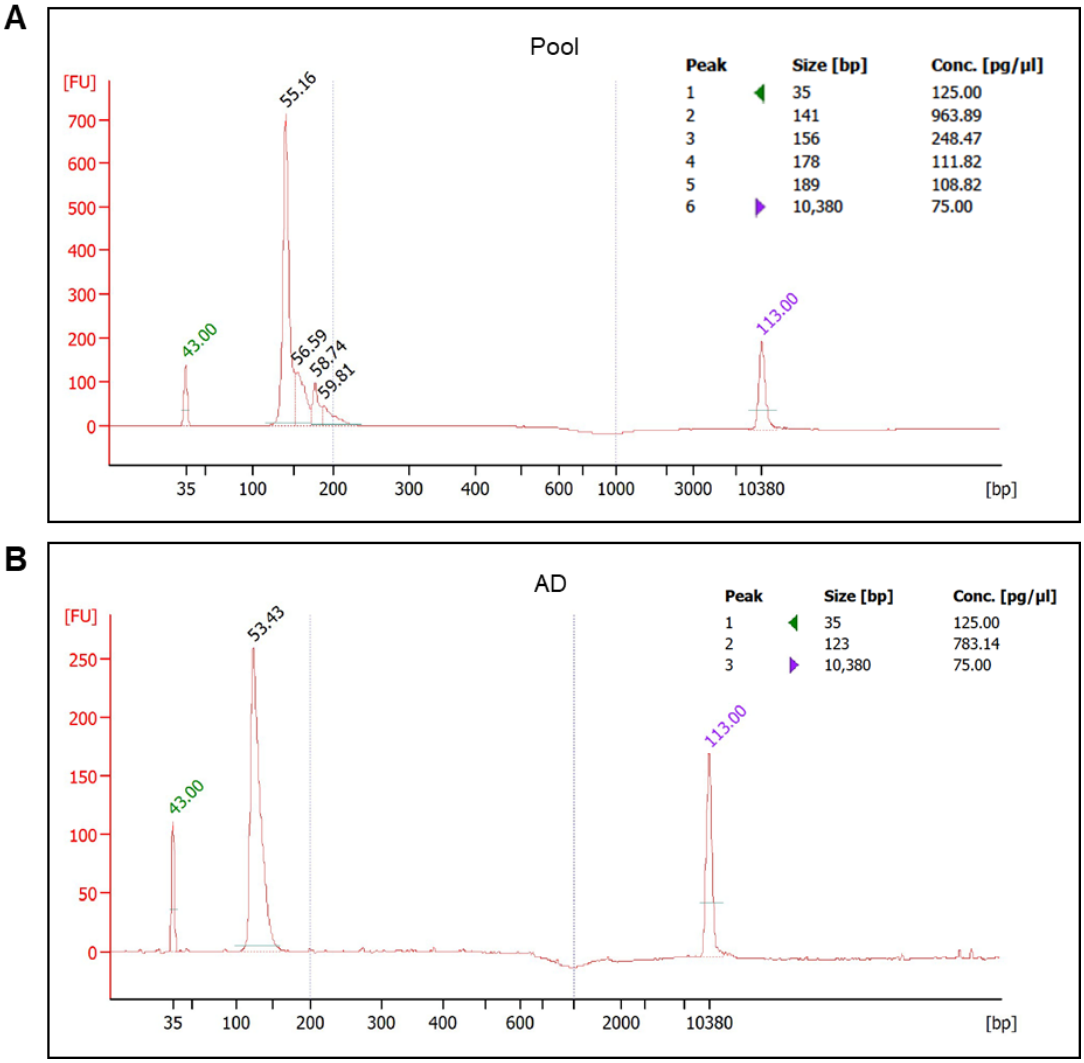
Supplementary Table 5.1 | Proteins identified from quantitative proteomic analysis following exWAGO IP from adult *H. bakeri*. The table shows the name of the proteins and the number of peptides detected, the fold change between the experimental (anti3) and negative control (naïve) samples. The spectral counts, abundance, and the *C. elegans* homologue genes are also detailed.

Supplementary Table 5.2 | Proteins identified from qualitative proteomic analysis following exWAGO IP from exWAGO-expressing MODE-K cells. The IPs were performed used different conditions (A – FLAG IP using high salt buffer; B – Protein G IP using high salt buffer; C – FLAG IP using van Nues protocol). The table shows the details of the proteins detected in the experimental (exWAGO-expressing MODE-K) and negative control (parental MODE-K) samples and the number of spectral counts detected per sample. The 'Count' columns show the total number of samples where each protein was identified.

Supplementary Table 5.3 | Proteins identified from qualitative proteomic analysis following exWAGO IP from exWAGO-expressing MODE-K cells. The IPs were performed using FLAG beads and van Nues protocol. The table shows the details of the proteins detected in the experimental (exWAGO-expressing MODE-K) and negative control (parental MODE-K) samples and the number of significant peptides detected per sample. Significance was determined as $p \leq 0.05$. The 'Count' columns show the total number of samples where each protein was identified.

Supplementary Table 5.4 | Proteins identified from qualitative proteomic analysis following exWAGO IP from *H. bakeri*-infected and uninfected gut tissue. The table shows the details of the proteins detected in the experimental (*H. bakeri*-infected tissue) and negative control (uninfected tissue) samples and the number of significant peptides detected per sample. Significance was determined as $p \leq 0.05$. The 'Count' columns show the total number of samples where each protein was identified.

Supplementary Figures



Supplementary Figure 4.1 | Quality control analyses of the libraries generated by the CLASH modified protocol using the Bioanalyzer.

(A) The electropherogram of the CLASH pool that was sequenced. (B) The electropherogram of the adapter dimer control. AD = Adapter Dimer control, bp = base pairs.

Appendix

Appendix A | The protein sequences of exWAGO and its homologues from the nematodes clade V and clade III.

>H.bakeri.HBAK_0000298601_mRNA_1_exWAGO

MDQLKTGMGQLSVGAVALPEKRSPGGIGNKVDFVTNLTELSLKPNVPYYKYDIRM
YIVYKGNDALEHLKELTKQTKDDFPEQERKSAAVAVYKHLCKTYKDVFLPDGALLY
DRAAVLFSAQRQLKLDGEEKQFMLPASVVSSAGPDATGIRVVIKKVKDQFQVTSN
DLSKAVNVRDMERDKGILEVLNLAVSQKGYMETSQFVTYGSGVHYLFDHRALGFR
DNELPELMDGKYMGIGLTKSVKVLEGDSGKGNSAFVVTDVTKGAFHVDEQNLMEK
ISQMSIFFDQRTGQSSFNAKNAMQPFNQKAILQQIKGLYVRTTYGKKKTFPIGNLAA
AANALKFQTADGAQCTVEQYFKKHYNIQLKYPGMFTVSERHNPHTYYPVELLTVA
PSQRVTLQQQTPDQVASMIKASATLPQTRLHQTKIMKDALDITPRNHNLATAGISVA
NGFTAVSGRVLPSPRIAYGGNQILRPVDNCKWNGDRSVFLEPAKLTNWAVCVTLT
QQDARRLQIKEYISRVEMRCRNRGMQVDPVAEVFTLKHQTFDGLKEWYASQKQK
NRRYLMFITSDGIKQHDSIKLLEVEYQIVSQEIKGSKVDAVVTKNQNQTLDNVVAKIN
MKLGGVNYNVMLGVKNDDKAFSWLNDKDRMFVGFEISNPPALSKVEIERGASYK
MPSVLGWGANCAGNHQQYIGDYVYIQPRQSDMMGAKLSELIVDILKRFRAATTIAP
RHIVLYFSGISEGQFSLVTDTYMRAVNTGIASLSPNYKPSVTAVAVSKDHNERIYKT
NISGNRATEQNIPPGTVIDTKIVSPVINEFYLNSHSAFQGTAKTPKYSLLADNSKIPLD
VIEGMTHGLCYLHEIVTSTVSVPVPLIVADRCAKRGHNVYIANSNQGEHSVNTIDEA
NAKLVNDGDLKKVRYNA

>C.elegans_SAGO1

MSNITQVTSSMASASLSNKAPLPVGHQPLAEKKPKEVNQEGTPVQIVTNMRKINLE
KNHSIFKYSVQVLFVYQKSDGTELVLEKSKSVGSGCDHERSKSHCLRVYRKAAKQ
CQELKSGGPFCYDSQGCLYSFSKLKNDEFSTNITGSDISNNPKFLRVEFKLAKVQE
SFQTTTNDVAKSVNCRPALQEKTILEAMNQIVSTAPINHPNVLTIGNCVHYLYDDTNI
DIRSITGEGGKSSAVGASKSVRTLEGTGKTPCLYMATELKTTLFHPDNCSLLKVFM
DYRGFNGSLKANSPFVLKNKNAFIGLWCYTTHGKCSDWKDDRPMIKIKDFGLSAK
ETTFERDNKKISVFNYFQVKYNMTLKYPDLFTVVARGKDGKNQHIPVECLDLCNSQ
TVRTEQMVGTEQADLIKLAAAKPHDRKKITDTVVNSIGLASEPKGIISVGAPESVTGL
VLPKPDIYFSGGKKVFWNDPKKRGPATDFMPAGTFIKPTKLTNWEVVFDNGVQLV
DCIQHLTSTMRQLGMEVSNPTVSLINRGYLRSIFENAKAANRQLIMFITKSMNNYHT
EIKCLEQEFDLLTQDIRFETAVKLAQQQNTRKNIIYKTNMKLGGLNYELRSGVFSNS
KRLIIGFETSQRGGLGDAPIAIGFAANMMSHSQQFAGGYMFVKKSADNYGPVIPEIL
LTILKQAKANRPNDRPDELLIYFSGVSEGQHALVNEYYANQVKAACGLFNESFRPHI
TLILASKVHNTRVYKSENGGGVCNVEPGTVIDHTIVSPVLSEWYHAGSLARQGTSK
LVKYSLIFNTKKNEKLSVYERLTNELCYEMQIVFHPTSLPIPLHIAGTYSERGSQMLA
LKKPIYTNGEFNQVATNEQLGYASKKLFGTRFN

>C.elegans_SAGO2

MEKQLKAMSVSDKPAAPAAQKLGTAPLAAKKTRNEEWGTKVNIDTNIRKLTIKPNQ
PIYKYAVQVNYVFRKPDGTEATIEMSKSAKKGTEHDNDKTRCQNVYNEAIKRYDEL
KTGGPFFYDRQASLYTLTKLKNESISFVVTDKICKRQNFKEAQFVLKKVDQSFQST
SNDVIKTTNSCPANADKTLLEAMNIIVSGPAFENKNVITVGACVHYLIDPTGVDVAYK
EYPEGQLYSGVGVSKSVKTLEGTDKKVPSLFMTTEMKTTLFHPDYAPLVELLQTFR
GFSTTLKANSPAAQRIEKAFVGLDVVLNYGVHKGLGEDGVVMKIRRFHTSAKETCF
EVEKSTREFTNVFDYFKKKYGITLKYPDLFTIEAKGKQGKIHFPAEVLLLCPNQTVTN
DQMINNEQADMIKMSAAQPHIRKTTTDTIVRNVGLASNNIYGFIKVEDPVNLEGMVL

PKPKIAFAGNRLADLANPKSRFPTDFNRAGQYYDAKELTKWELVFVQNEEVQGLA
KQLADEMVNNGMKCSNPTMSFIIRGDLEPIFKKAKAAGTQLLFFVVKSRYNYHQQI
KALEQKYDVLTQEIRAETAEKVFRQPQTRLNIINKTNMKLGGLNYAIGSEAFNKPNR
LIVGFVTSQRVGGNPDYPISVGFAANMLKHHQKFAGGYVYVHRDRDVFGSIIKDTL
LTIFKTCTEQRGRPDDILLYFNGVSEGQFSMINEEFSARVKEACMAFQKEGTPPFR
PHITIIASSKAHNERLYKSDKGRIVNLEPGTVVDHTIVSNVYTEWYHASAVARQGTA
KATKFTLIFTTKAGPQAEPLWHLEQLTNDLCYDHQIVFHPVGLPVPLYIADRYSQRG
AMVLAANQGPIYNEGQIDLAATNSAYGYGEKKLFTTRFNA

>C.elegans_PPW-1

MEKQLEAMFVSDRPAAPAAQKLGTAPLAAKKTRNVERGTKVNIDTNIRKLTIKPNQ
PIYKYAVQVNYVFRKPDGTEATIEMSKSAKKGTEHDNDKTRCQKVYNEAIKRYDEL
KTGGPFFYDRQASLYTLTKLKNESISFVVTDKICKRQNFKEAQFVLKKVDQSFQST
SNDVIRTTNSCPANADKTLLEAMNIIVSGPAFENKNVITVGACVHYLIDPTVVDVAYK
EYAEGQLYSGVGASKSVKTLEGTDKKVPSLFMTTEMKTTLFHPDYAPLVELLQTFR
GFSTTLKANSPAAQRIEKAFVGLDVVLNYGVHKGLGEDGVVMKIRRFHTSAKETCF
EVEKSTREFTNVFDYFKKKYGITLKYPDLFTIEAKGKQGKIHFPAEVLLLCPNQTVTN
DQMINNEQADMIKMSAAQPHIRKTTTDTIVRNVGLASNNIYGFIKVEDPVNLEGMVL
PKPKIAFAGNQLADLANPKSRFPTDFNRAGQYYDAKELTKWELVFVQNEEVQGLA
KQLADEMVNNGMKCSNPTMSFIIRGDLEPIFKKAKAAGTQLLFFVVKSRYNYHQQI
KALEQKYDVLTQEIRAETAEKVFRQPQTRLNIINKTNMKLGGLNYAIGSEAFNKPNR
LIVGFVTSQRVGGNPDYPISVGFAANMLKHHQKFAGGYVYVHRDRDVFGSIIKDTL
LAIFKTCTEQRGRPDDILLYFNGVSEGQFSMINEEFSARVKEACMAFQKEGTPPFR
PHITIIASSKAHNERLYKSDKGRIVNLEPGTVVDHTIVSNVYTEWYHASAVARQGTA
KATKFTLIFTTKAGPQAEPLWHLEQLTNDLCYDHQIVFHPVGLPVPLYIADRYSQRG
AMVLAANQGPIYNEGQIDLAATNSAYGYGEKKLFTTRFNA

>ACAN_exon_1

MADQLKKAMGDLTVSTIALPEKRPSGTTGANTEFVANLTSLKLKPNVPFYKYDMR
MYIVYKGNDGKEHLKELTKQTKDDFPEQERKTGTVLVYKHLLKSYPNIFPQDGALL
YDRAAVLFSAQKQIKLDGDEKVFTLPASLVPSAGEDAVGVRVVIKKVTDGFQVTSN
DLQKVVNVRDMEKDKGILEVLSLAMSQKGYMETSQFVTYGSGVHYLFDHRALGFK
DHEVPELMDGKYMGVGLSKSVKVLEGEKGQPCGAYVVTDVTKGAFHMDDQNLLE
KISQMSMFIDPRSGQSHFSVQAVTQPFNQKNILQLIKGLYVRTTYGKKKTFPIGNLA
QPANQLKFQTTDGTQCTVEQYFKKHYNIQLKYPGMFTVSERHNPHTYYPVELLRV
APSQRVTLQQQTPDQVAAMIKACATLPQNRLHQTKLLKDALAIKEGNPHLSAAGIS
VVNGFTSVPGRVLPSPSIVYGGNQLVKPVDNCKWNGDRSRFLEPARLHNWAVCA
TLTQNDSRRLNVKEYVAKIEGRCRQRGMDVEPCAEIFNLQRQNFESLKEWYASQK
AKNRRYLMFLTSDGIKQHDLIKLLEIEYQIVSQEIKGSKVDAVVSRNQNQTLDNVVA
KINEKLGGVNYNIMLGARPTDDVNKWISDKDRMFVGFEISNPPALSKVEIERGATY
RMPSVLGWGANCAKNPQQYLGDYVYIEPRQSDMMGAKLSELIVTILKRFRAATDV
APRHIVLYFSGISEGQWSMVADTYMRAIHTGIKSLSASYKPSLTALTVSKDHNERIY
KANITGNRATEQNIPPGTVVDTKIVSPVINEFYLNAHSAFQGTAKTPKYALVYDDSNI
PMNVVEGMTHGLCYLHEIVTATVSVPVPLIVADRCAKRGHNVYIANSNQRDAVGSI
KEANERLVNQGELQKVRYNA

>ACEY_exon_1

MADQLKKAMGDLTVSTIALPEKRPPGTTGANTEFVANLTSLKLKPNVPFYKYDIRM
YIVYKGKDGKEHLKELTKQTKDDFPEQERKTGTVLVYKHLLKSHPNIFPQDGALLY
DRAAVLFSAQKQIKLDGDEKVFTLPANLVPSAGEDAVGVRVVVKKVTDGFQVTSN
DLQKAVNVRDIEKDKGILEVLSLAMSQKGYMETSQFVTYGSGVHYLFDHRALGFK

DHEVPELMDGKYMGIGLSKSVKVLEGEKGQPCGAYVVTDVTKGAFHMDDQNLLE
KISQMSMFIDPRSGQSHFSVQAATQPFNQKNILQLIKGLYVRTTYGKKKTFPIGNLA
QPANQLKFQTTDGTQCTVEQYFKKHYNIQLKYPGMFTVSERHNPHTYYPVELLRV
APSQRVTLQQQTPDQVATMIKACATLPQNRLHQTKLLKDALAIKEGNPHLSAAGIS
VVNGFTSVPGRVLPSPSIVYGGNQLVKPVDNCKWNGDRSRFLEPARLHNWAVCA
TLTQNDSRRLNVKEYVARIEGRCRQRGVDVEPCAEIFNLQRQNFESLKEWYASQK
AKNRRYLMFLTSDGIKQHDLIKLLEIEYQIVSQEIKGSKVDAVLSRNQNQTLDNVVA
KINEKLGGVNYNIMLGTRPTDDVNKWISDKDRMFVGFEISNPPALSKVEIERGATYR
MPSVLGWGANCAKNPQQYLGDYVYIEPRQSDMMGAKLSELIVQILKRFRSATDVA
PRHIVLYFSGISEGQWSLVADTYMRAIHTGIKSLSASYKPSLTALTVSKDHNERIYKA
NITGSRATEQNIPPGTVVDTKIVSPVINEFYLNAHSAFQGTAKTPKYALVYDDSNIPM
NVVEGMTHGLCYLHEIVTATVSVPVPLIVADRCAKRGHNVYIANSNQRDAVGSIKE
ANERLVNQGELQKVRYNA

>ancylostoma_duodenale.ANCDUO_12582

MCATQLLMVSTIALPEKRPPGTTGANTEFVANLTSLKLKPNVPFYKYDMRMYIVYK
GNDGKEHLKELTKQTKDDFPEQERKTGTVLVYKHLLKSHPNIFPQDGALLYDRAAV
LFSAQKIKLDGDEKVFTFPANLVPSAGEDAVGVRVVIKKVTDGFQVTSNDLQKVVN
VRDVEKDKGILEVLSLAMSQKGYMETSQFVTYGSGVHYLFDHRALGFKDHEVPEL
ADGKYMGIGLSKSVKVLEGEKGQPCGAYVVTDVTKGAFHMDDQNLLEKISQMIDP
GQSHFSVQAATQPFNQKNILQLIKGLYVRTTYGKKKTFPIGNLAQPANQLKFQTTD
GTQCTVEQYFKKHYNIQLKYPGMFTLSERHNPHTYYPVELLRVAPSQRVTLQQQT
PDQVATMIKACATLPQNRLHQTKLLKDALAIKEGNPHLSAAGISVVNGFTSVPGRVL
PSPSIVYGGNQLVKPVDNCKWNGDRSRFLEPARLHNWAVCATLTQNDSRRLNVK
EYIAKIEGRCRQRGMDVEPCAEIFNLQRQNFESLKEWYASQKAKNRRYLMFLTSD
GIKQHGRNQNQTLDNVVAKINEKLGGVNYNIMLGTRPTDVNKWISDKDRMFVGFEI
SNPPALSKVEIERGATYRMPSVLGWGANCAKNPQQYLGDYVYIEPRQDVSFHACC

>ACANT_exon

MDQLEGAMTNLTVKTVAMPAKRSPGQRGVKSEFLTNLTKLTLKPNVPFFKYDVRM
YMVYKGKDGGEHLKELTKQTKDDFPEQERKASSVIVYKHLVKSYPDVFTKGGALF
YDRAAILFSAQVQLKFDGEKKEFHLPASVLSNAGDNVGVRVVIKKVADSFQVTSND
ILKAVNVRDIERDRSILEVLNLAISQEGYLETLKFVAYGSTVQYLFDHQAFGFRDNEL
PELMDGKYAGIGLSKSVKVLEGDDGKCSPFVVADVTKSAFHADEQNLLEKISQMSI
FFDYRTGTSNFSVQIASRPNIIKDILRLIKGLYVKTLYGRTRTFPIGGISAAASSLRFQ
ATDGKQYTVEQYYLKQYNIQLKYPGLFTVSERHNPHTYYPVELLAVAPSQRVTQQ
QQTPSDVMALIKASATLPQQRLSQTRVMKNALKIAPGNSLLEAAGISVDKNFTKVV
GRVLPSPTILYGGSSIAKVDVCCKWSWDRAKFLKPANLSNWAVCVTLTQNDFRRV
QIKEFITRVEGRCQSHGMKVSPASEVFYLKHQTFDGLKEWYAEQKKKNRKYLLFIT
SDNIQQHDTIKLLEVQYQIVSQEVKAGTVTAVVLKNQNQTLDNVISKINQKLGGVNY
NIVLDSRLSEKRNWLSDNGVLFVGFEISNPPALSKMEIARGETYKMPSVLGWGAN
CSSNPQQYLGDYVYVEPRQSDMMGSKLADIVIDIIKRHQMATSVTLQHIIFYFSGISE
GQFSMIADMYMRAVYTAISCISLENEPKVTALAVSKDHNERIYKSNIVGKRAAEQNI
PGGTVIDTTIVSPVINEFYLNAHSAFQGTAKTPKYSLVADNSQISLDAIEGVTYGLCY
LHEIVTATVSLPVPLVVADRCAKRGHDVYVANLREKHVVVNSIKEANELLVNQGGL
KKLRYNA

>ACOST_exon_2

MDQLEGAMRNLTVKTVAMPAKRSPGKRGVKSEFLTNLTKLTLKPNVPFFKYDVRM
YVVYKGKDGGEHLKELTKQTKDDFPEQERKASSVIVYKHLIKSYPDVFTKDGALFY
DRAAILFSAQVQLKFDGEKKEFHLPASVLSSVDDNAGVRVIIKKVTDSFQVTSNDIV

KAVNVRDIERDKSILEVLNLAVSQEGYLETLKFVAYGSTVQYLFDHQAFGFRDDEL
PELMDGKYVGIGLSKSVKVLEGDDGKCSPFVVADVTKSAFHVDEQNLLEKISQMSI
FFDYETGKSNFSVQTASRPNVMKDILRLVKGLYVKTLYGKTRTFPIGGISAAASSLR
FQATDGKQYTVEQYYMKQYNIQLKYPGLFTVSERHNPQTYYPVELLAVAPSQRVT
QQQQTPSDVAALIRASATLPLHRLNQTGVMKKALKIAPGNFLLEAAGISVDKNFTKV
VGRVLPSPTILYGGSSITKVDVCCKWSWDRAHFLKPANLSNWAVCVMLSQNDFRR
VQINEFIGRVEVRCRSHGMQVSPASQVSYLKRQTYDGLKEWYAEQKKKNRKYLLF
ITSDNIQQHDTIKLLEVHYQIVSQEVKAGTVNAVVLKNQNQTLDNVISKINQKIGGVN
YNIVLDSGLSDNVRNWLSDNGVLFVGFEISNPSALSRMEIARGATNKMPSVLGWG
ANCSSNPQQYMGDYVYVEPRQSDMIGSKLAEIVIEIIKRHRMTTRVTLHHIIFYFSGI
SEGQFSMISDIYMRAVYTGISCLSLQNVLKVTALVVSKDHNERIYKSNIVGKRAADQ
NIPGGTVVDTVIVSPVINEFYLNAHSAFQGTAKTPKYSLVADNSQISLDAIEGITYGL
CYLHEIVTATVSLPVPLIVADRCAKRGHSVYVANLREKHVVVNSIQEANELLVNQGD
LRKLRYNA

>DVIP_exon_fixed

MDQLRNAMGDLNMKTVAVPEKRPSGNRGMKTEFLTNITKLSLKPNVPFFKYDVRM
YVVYKGTDGREHLKELTKQTKDDFPEQQRKSLAVLVYKHLIKSYPDVFLKGVALFY
DRASVLFSAQRQIKLAAEKEEFIVPASILSNACGDAEKVCVVIKKVSDSFQVTSNDIM
KAVNVREFERDKNILEVLNLAVSQEGYLETTKFVVYGSNEHYLYDHRAYGFRDNEL
PELMDGKYMGIGLSKSVKVLEGENGKCGPFFVADVTKSAFHADEQNLLDKISQMS
VFLDRRTGAYNFSVDVACKPYNMKNILQLLKGLYVKTSYGKTRTFPIGNLAPAANIL
RFQASDGNQYTVEQYFKKHYNIKLKYPSLFTVSERHNPHTYYPVELLFVAPSQRVT
SQQQTPEDVAAIIKASATLPQHRLNQTKVMKDALKMVPGNTYLEAAGITVDKDFAK
VVGRVLPAPTVVYGRSETIAVNECKWNWDRSQFIQPGNLSNWAVCATLTPNDSG
RIKIREYISRVESRCRSHGMKIEAAAEIFYLKRQTFEELKEWYAAQKKKNRKYLMFIT
SDSIKQHDLIKLLEVQYQIVSQEVKASKVDAVMFKNQNQTLDNVVAKINEKIGGVNY
NITLNTAAGDNNCLSDSGVLFIGFEISNPPSSSKTEVIICTAHKIPSVLGWGANCASN
PQQFIGDYVYVEARQSDMMGSKLAGVVMDIIERHRLATKTAIRHLIFYFSGISEGQF
GMIPDSYMRAINTGLSSLSPQCVACVTALAVSKDHNERIYRANITGKRAADQNIPG
GTVVDTKIVSPVMNEFYLNAHSAFQGTARTPKYSLLADNSMMPLDVIEGLTYGLCY
LHEIVSATVSVPAPLIVADRCAKRGHNIYLANMKEKQVVVGTIKEANEKLVNRGDLK
KVRYNA

>H.contortus_HCOI02120600.t1

MADQLGAGMKQLSVSTVALPEKRDPGSRGTKTEYLTNLTNINLKPNVPFYKYDVR
MYIVYKGNDGREHLKELTKQTKDDFPEQERKIAAVGVFKYMCKNYKDVFPNDGAL
FYDRAAVLFSAQNELKLGGQERTLSIKPTANLVPAAGKDAAEVRVVIKKVTEGYQV
TSNDLIKAVNIRDCERDKGMLEVLNIALSQKGYLETSQFVTYGTGVHYLFDHRALGF
RDNELPSLMDGKYMGIGLTKSVKVLEGDGKKGSAYVVTDVTKGAFHIDEQSLLEKI
SQMSIFFDPRRGQSTFNVQAAMQPFNMKSILQLIKGLYVRTIYGKKKCFPIGNIASP
ANQISFETESGKATVEQYFKKHYNIQLKYPTLFTVSERHNPSTYYPVELLVVAPSQR
VTLQQQTPDQVASMIKASATLPAVRIQQTKVMKDALGITSGNAKLSSAGISVEDSFT
KVPGRVLPAPTIIYGGNQGIRPKDNCKWNGDQSKFIEPAQLTNWAVCATLTQNDS
RRLKIGDYIARVEARCRLRGMQVERAAEIFDLKKQTFEGLREFYAAQKQKDRKYLM
FITSDSIKQHDYIKLLEVEYQIVSQEVKGSKVDSVMFKNQNQTLDNVIAKINMKLGG
VNYVVSLGPRMDDPVSKWLNDEARLFVGFEISNPPALSKMEIERGATYKMPSVLG
WGANCAANPQHYLGDYKFIKARQSDMMGATLGELIVEILKKFKAATSRAPRHIVLY
FSGISEGQFSLVTDTYVRAIRTGIKSLSEAYNPNVTALAVSKDHNERLYKSKIVGER
ATDQNIPPGTVVDTKIVSPVINEFYLNAHSAFQGTAKTPKYSLLADDSNIPLDVIESM
THGLCYLHEIVTSTVSVPVPLIVADRCAKRGHNIFIANSGQGKAAVSSIEEANERLTN
HGELQKVRYNA

>N.americanus_NAME_model

MADQLKKRMGELTVDTVALPEKRAPGTLGAATEFVTNLTSLKLKPNVPFFKYDIRM
YIVYKSSDGKEHLKELTKQTKDDFPEQERKTGTVLVYKHLLKTNPSVFPQDGALLY
DRAAVLFSAQKQIKLDGEEKVFMLPASLVPSAGEDATGVRVVVKKVTEGFQVTSN
DLAKAVNVRDFEKDKGILEVLNLAVSQKGYMETSQFVTYGSGVHYLFDHRALGFR
ESGLFKIIKLLFKIIGLXXXVKVLEGEGESCTAYVVTDVTKGAFHIDDQNLLEKISQMS
MFIDPRSGQSHFNIQAAMQPINQKNILQLIKGLYVRTTYGKKRTFPIGNIAQAANQLK
FQTVEGTQCTVEQYFKKHYNIVLKHPGMFTVSERHSPHTYYPVELLRVAPSQRVTL
QQQTPDQVATMIKACATLPQNRLHQTKLLKDALDIKPGNPRLAVAGISVENGFTTV
PGRVLPPPSIIYGGNQLVKPIDNCKWNGDRSRFLEPARLYNWAVCATLTPNDSRRL
HIKEYIVRVEGRCRQRGMDVEPCSEIFNLQRQNFESLKEWYASQKEKDRRYLMFIT
SDHIKQHDLIKLLEIEYQIVSQEIKGSKVDAVLTRNQNQTLDNVIAKINEKLGGVNYNI
MLGSSPSDKANKWLYEKDRMFVGFEISNPPALSKAEIERGAAYKMPSVLGWGAN
CAKNPQQYLGDYVYIEPRQTDMMGAKLSEIIVHILKRFRAATDVAPRHIVLYFSGISE
GQWSLVADTYMRAIQTGIKSLSATYGPSLTALTVSKDHIERIYKSNITGNRATEQNIP
PGTVVDTKIVSPVINEFYLNAHSAFQGTTKTPKYALVYDDSNIPINAVEGMTHGLCY
LHEIITATVSMPVPLIVADRCAKRGHNVYIANSSQRNAVSCIKEANEKLVNQGALQK
VRYNA

>Nbrasiliensis.NBR_exon

MVDQLQGRVEKLSVSNVALPEKRAPGSAGVKQDFVCNLTALKLKPNVPFYKYDIRI
YVVFEKSDGTEHQKELTKQTKDDFPEQERKTASVYVFKTLCKMYKDIFPQDSALFY
DRAAVLFSAQTQLKLGGEEKQFKLPAKGVPSTGADAKGILVVIKKVTDKFQLTSND
LMKAVNVRDCEREKEILEVLTLAVSQKGYLETSQFVTYGSGVHYLFDHRALGFRDN
ELPELMDGKYMGIGVTKSVKVLEGEDGTGATPFVVTDVTKGAFHIDDQNLMEKISQ
MSIFFDQRTGQSSFSQRTAMQPHNQKTILQQIKGLYVRTTYGAKPRTFPIGNIARPA
SELKFEAADGSTFTVEQYFKKQYNIQLKYPWMFTVSERHKPTSYYPVELLRVAPSQ
RVTLQQQTPDQVASMIKASATLPQTRLHQTKVMKDALDINSRNPYLASAGITVNND
FATVTGRILPTPTILYGERQEVKPVDNCKWNGDRSRFLEPAYLSNWAVCATLTQFD
ARKIQVKDYVSRIEARCRNRGMQVDPVAEIFHLQNQSFEGLKEWYAAQKQKGRRY
LMFITSDGIKQHDLIKLLEIEYQIVSQEVKGSKVDAVLHRNQNQTLDNVVAKINMKLG
GVNYNVMLGRPNEPSSKWLTDKNRLFIGFEISNPPPLSKAEMERGATYKMPSVLG
WGANCASNPQQYIGDYLYIQARQSDMMGQKLSEVVVDILKRFREATSVAPQHMVL
YFSGISEGQFSLVTDTYMKAIHTGISSLSANYKPHITALAVSKDHNERIYKSKISGTR
ASEQNIPPGTVVDTKIVSPVINEFYLNAHSAFQGTAKTPKYSLLADDSKVPLDIVER
MTHGLCFLHEIVTSTVSVPVPLIVADRCAKRGHNVFIANSGQGRSAVGSIDEANAKL
VNHGELQKVRFNA

>ODENT_exon

MADQLKKAMGELSVSTVALPEKRPPGTMGARTEFVANVTSLKLKPNIPFFKYDIRM
YVVYKGADGKEHLKELTKQTKDDFPEQERKNATVLVYKSLVKNNSKVFPPEGALF
YDRAAVLFSAGTQIKLDGDEKQFMMPASLVPSAGEDAVGVRVVIKKVTEGFQVTS
NDLAKAVNVRDIEKDKGLLEVLNLAMSQKGYLETSQFVTYGSGVHYLFDHRALGF
RDQEVPELMDGKYMGIGLTKAVKVLEGDKGQSCGAFVVTDVTKGAFHMDDQNLL
EKISQMSMFIDPRSGQSHFNVQSAMQPFNQKAILQLIKGLYVRTTYGKKRTFPIGNI
AQPASQLKFQTVDGKQCTVEQYFKQHYNIQLKYPAMFTVSERHNPHTYYPVELLR
VAPSQRVTLQQQTPDQVATMIRACATLPQNRLQQTRVLKDALGIKDGNPHLSAAGI
SVVNGFTSVPGRVLPSPSIVYGGNQLAKPIDNCKWNGDRYRFLEPASLRNWAVCV
TLTPNDSRRLHVKDYVARIEGRCRQRGMEVEPCSEIFTLQRQNFDSLKEWYVSQK
AKNRRYLMFITSDNIKQHDLIKLLEIEYQIVSQEVKGSKVDAVLTKNQNQTLDNVVAK
VNQKLGGVNYNVMLGANPNEAVNKWISEKDRMFVGFEISNPPALSKVELERGGTY

KMPSVLGWGANCAKNPQQYLGDYLYIEPRQTDMMGAKLSELIVQILKRFRGATEV
APRHIVLYFSGISEGQWSMVADTYVRAIQTGIKSLSPNYKPNLTALTVSKDHNERLY
RANITGNRASEQNVPPGTVVDTKIVSPVINEFYLNAHSAFQGTAKTPKYALVYDDS
HIPIDVVEGMTHGLCYLHEIVTATVSVPVPLIVADRCAKRGHNVYIANSNQRNAVGS
IDEANARLVNQGELQKVRYNA

>Tcirc.Tcirc_exon_trim

MADQLSGGMGKLSVAAVALPEKRAPGSLGTKLDFVTNLTGIKLKPNVPYYKYDVR
MYIVYKGNDGREVLKELTKQTKDDFPEQERKMAAVAIYKHLVKSYKDIFPQDGQFF
YDRAAVLFSAQREMKLGGPEKVITLPASLSPTAGSDAAGIRVVIKKVTDGYQVTSN
DLMKAVNVRDCERDKGILEVLNLAVSQKGYMETSQFVTYGTGVHYLYDHRALGFR
DNELPDLMDGKYMGIGLTKAVKVLEGDQGKSASAFVVTDVTKGAFHIDEQNLLEKI
SQMSIFFDPRTGQSTFSVKAAMQPHNMKSILQLIKGLYVRTTYGRKRTFPIGNLAAA
PNALKLQTSDGVQCTIEQYFKKQYNVQLKYPGLFTVSERHNPHNYYPVELLTVAPS
QRVTLQQQTPDQVASMIKASATLPSNRLHQTKVMKEALDITPRNAKLASAGINVED
GFTTVPGRVLPTPTILYGGSQTLKPVDNCKWNGDRSRFLEPAQLTNWAVCATLTQ
NDARRLQIKDYVARVESRCRAKGMQVEAAAEIFTLTKQNFDGLREFYAAQKKKNR
KYLLFITSDGIKQHDLIKLLEVEYQIVSQEVKGSKVDSVMFKNQNQTLDNVIAKINMK
LGGVNYNVVLGSKPNDPASKWLNDKDRLFVGFEISNPPALSKMEIERGATYKMPS
VLGWGANCAANPQHYIGDYVYIKPRQSDMMGAKLSELIVEILKKFRGATSLAPRHIV
LYFSGISEGQFSLVTDTYMKAINTGITSLSANYRPSVTALAVSKDHNERLYKSNISG
SRANEQNIPPGSVVDTKIVSPVINEFYLNSHSAF

>acanthocheilonema_viteae.nAv.1.0.1.t07849RA

MGQTSITGPPVCLTEKIAPGTRKIELLDVVTNVWGLIPHENIPIYRYDFRVMEEYPPK
KGSPLLKEVTKQTKNDYLTVDRKTKCMVIYQILLKREEQFFGTLDSLIYDRASTLYSL
RKLPFSKDEKEEIFFVKPNELPMNIVGENCIMVHVHIKPCKDDFQLTMNDLSSCVSN
NPDQINRSLQQFLEILAMQEVFFMEGRFVSYGTGECYLMDPNQFGFGERDVPELQ
EGKYVAIGAAKGVRIIEGPRGKKGINAALVIDVKKAAFHIDNQNLFKKAEDILRKSVD
LTRRIDQQSITVLNKALKGLYVRCNYGKNRAFVIAGVSKENARTSKLVTKDGEMSV
EKYFGMKYSMKLKYPFLPLIIERFPPKNNFYPIEVLSVCENQRVSKGQQTSFQVQT
MVKACAIPPSLRLQQTNVLSKAMNLGTFDRNRWMEKCNVAVTNNLELKARVLPMP
AIEYRTNGWVKPSEKTSWEDGKNQYLIPAVCKKWCAVALMGPRERLNEYQFRNYI
RTFLQHCRRHGMEMDDPFLCEYIQRSKQEDIEPLITRAKHSGAVFIHFITADELNYH
AHMKYVESQEQIVTQDLKASTAVAVVTQNKRQTLDNIVNKTNVKLGGMNYSVHLE
TNCDIWLTKPGLLIVGLDIAHPVFSNVSKRDRNSIPSVVGYSANIKKHPLDFIGGYRY
GKANMEEMADDTIQHIFSDILRYFNANRGKPPTHLFVIRD

>anisakis_simplex.ASIM_0001785301mRNA1

MEINCSKMELTDESAEERLVMPSKKISGMRLKAADSIAVVMNGYEIDLSNVPEKVY
KHELKLMAVKQDNQLRDLTRGPRNDVSITLRRRVLWSVYTTVLKNYKDFFGSDTK
MYFYDCGVTLYSINKILREDGEKEFRLKMDQLSSNSRDFLGVKVVGIVAKLLACEEV
YLRDFDESISECIDERQRSLQQFLEVAINQKLYHDQDHLLFGNKAYHRAMPHDKQL
KGGKLLQSGFAKNIRVVGDSVENAVIVAQLDAKKSAFFKEQNLVALIASLCNNRFTL
LNDANTRRRVAKQLKGLVVRTNHLPKSQRIFSIFGMTKESASKVIIVIDDEEISIEEYY
NHKYQIGLQYTNLPCIVERRFQAKNYYPMECLDVCKGQRVENKKQTPDLVEELIAG
CRLLPNRLKEENEKQRRAASITNQNPYFRRLGMLYPPAIIYGNNDRVEPNSNGHW
RLNRQHLFYSPANPPRRWVVFIFEDAVSKELFDRFLYSYIERAQSHGIKLQRPSRIE
TIDRVDMDYLMDKMKMMRKNCVEYVMFITKDKRDPVHDKMKLTEVQASVVTQHIF
SGTIQRSIGNRGAEMTLDNLIMKMNLKLGGISHAIAASTPFMRCNHLNEDICKRIWL
RPTRMFIGLDMSHSSPLSFYERQAGFHASEPTVVGMAYTCGSEFGMRGLYWMQE

PRVYTIQSLKEHLVEALNQFKEENKSYPEHVIVFRGGISDGQFQKVMTLEANAFRA
AFEALPSSTPPNKIRLSLICVQTNSNYRLLLDNLQVDGNALQQNVPSGTCVDSSIVH
PTQAEFILVAHKSIMGTARPIRCTVLVDDPRMGLDEVEGITNCLCYMHGIVTSPISVP
AHLYAASNLAKRGRNNWKIVCNNGDDDVSIASGDGSHNQFHNDGAPDFFTNISSH
LAPKLKHKFWA

>ascaris_lumbricoides.ALUE_0001385501mRNA1

MPIAVQDMPVRFAEKVSPGRRHAETVDLISNVWGVIPKKNVPIYRYDVRILEEFPPK
ASGAPTKEVTKQCRDDFPSIERKNRCVAVFLRLLEREEAFFGKRESVVYDRASILY
TLDRLQIENDETKTFIVTPNELPEGSVSTDCVRVLFNIKQCTEDFQLTTSDLKQGVS
LEGERINRSLQQFFELLASQEAFFTEGRFVTYGTGQSFLFEPYDFGFRDQDMPPLP
DGKYIGIGASKGVKLIEGPSGPGGIHAALVMDVKKAAFHIEQQSVAEKVAMIFNVLS
MSVDPRQIPQLKKLLKGLYVRCEYGKQRVFMITNIATQNALQMRFRCDDMMVTVA
DYFASKYDIRLKYPQLPLVIERRPSGESYYPMEKLIVCENQRVTQTQQSSAQVQAM
IKACATLPIHRIRQTTAMTRAMKLDGTELNRWMREYSVNVTKNLTLKGRVLPPPRIE
YGRNERTIVNPERTSWLANRNHYLLPAKCEKWHVVALVGPSERFSNDKLRAYVRA
FMNQCRNRGMQMADPMVVDYVRGAREQEVDVRMQKAKQMGATFVHFVTSDML
KFHGHIKLVEMQLQIVTQDLTTRTASQAPQKWQTLDNIVNKTNLKLGGINFGLILEN
EIFRFSAQKWLMNEGRLVVGIDVAHPPLAAVRGIDRTKVPSVVGYSSNCKKFPLEFI
GGYRYATANMEELTDNSIRDVIVDSIRKFQVNRGKLPDHLFILRDGISEGQYKYVVV
SEVEGVKKACGLVGGIGYRPNITYIVATKLHNMRLFKKNINQQDKATGQNIKPGTVV
DKHIVNPVLNEFYLNSHSAFQGTAKTPRYTILFDTAKVPSDEIQAIVYALAYNHQIVN
AAISLPAPIVIAARMASRGRSNYAVQFGEGSDSTEGGRERNIAELNANMGYMDKPL
SDCRFNA

>ascaris_suum.ASU_04951

MLMKTDQGNHLITSVMAEVEELARRAERMEIQDMPVRFADKVSPGRRHAETVDLI
SNVWGVIPKKNVPIYRYDVRILEEFPPKASGAPTKEVTKQCRDDFPSIERKNRCVA
VFLRLLEREKAFFGKRESVVYDRASILYTLDRLQIENDETKTFIVTPNELPEGSVSTD
CVRVLFNIKQCTEDFQLTTSDLKQGVSLEGERINRSLQQFFELLASQEAFFTEGRF
VTYGTGQSFLFEPYDFGFRDQDMPPLPDGKYIGIGASKGVKLIEGPSGPGGIHAAL
VMDVKKAAFHIEQQSVAEKVAMIFNVLSMSVDPRQIPQLKKLLKGLYVRCEYGKQR
VFMITNIATQNALQMRFRCDDMMVTVADYFASKYDIRLKYPQLPLVIERRPSGESY
YPMEKLIVCENQRVTQTQQSSAQVQAMIKACATLPIHRIRQTTAMTRAMKLDGTEL
NRWMREYSVNVTKNLTLKGRVLPPPQIEYGRNERTIVNPERTSWLANRNHYLLPA
KCEKWHVVALVGPSERFSNDKLRAYVRAFMNQCRNRGMQMADPMVVDYVRGA
REQEVDVRMQKAKQMGATFVHFVTSDMLKFHGHIKLVEMQLQIVTQDLTTRTASQ
APQKWQTLDNIVNKTNLKLGGINFGLILENEIAQKWLMNEGRLVVGIDVAHPPLAAV
RGIDRTKVPSVVGYSSNCKKFPLEFIGGYRYATANMEELTDNSIRDVIVDSIRKFQV
NRGKLPDHLFILRDGISEGQYKYVVVSEVEGVKKACGLVGGIGYRPNITYIVATKLH
NMRLFKKNINQQDKATGQNIKPGTVVDKHIVNPVLNEFYLNSHSAFQGTAKTPRYTI
LFDTAKVPSDEIQAIVYALAYNHQIVNAAISLPAPIVIAARMASRGRSNYAVQFGEGS
DSTEGGRERNIAELNANMGYMDKPLSDCRFNA

>Brugia_malayi.Bm4557

MAGSSAVEKVRMQMEQASVSRLPVRLADKIAPGIRKSELLDVVTNVWGLISHENIPI
YRYDFRVLEEYPPKKDSPSFKEVTKQVRNDYVAVDRRAKCFAVYQALLRREKQFF
GAADTLIYDRASILYSLRKLSFPRDEVRKTFFLNKNELPANIVSDDCVKVHVNIKPCK
EDFQLAMHDLKNCVSNNPDEINSSLQQFLEILAMQEVFFMQGRFVSYGTGECYLM
DPSQFGFGDRDMPELQEGKYVAIGAAKGVRIIEGLGNKRGINAGLVLDAKKAAFHI
DNQHLLEKVESIFRRRAEIGHGIDQQSILILSKALKGLYVRCSYGKNREFAIAGISKE

NAHTSKLALKTGEMSVEKYFMTKYSIKLKYPKLPLIMERCQPKNNLYPIEVLVVCEN
QRVSKGQQTSSQVQTMIRACATEPSVRLMQTNRLSQAMKLDNSNQHKWIGKCN
MTITDNFMFPARVLPQPVIEYHTNGWINPNEKTVWMDGKKYYLIPAVCEKWYAVAL
MGPRERFHENQFCTYIKMFLERCRKHGMQIRDPIGYEYIRRSKEQDIESLIIKAKKS
GATFIHFVTADELNYHARMKYIESQEQILTQDLKASTALGIVMKRQYQTLDNIVNKT
NIKMGGLNYSVHLEENCDRWLGRPGFLIVGLDIAHPSYSRVPNKDRNTVLSVVGYS
ANIKKHPLDFIGGYRFAKAQMEELMDDAIQQIFSDLLRYFNANRGTPPTHLFVIRDGI
SVGQYKYVMNTEVEQIKHACQLVGGQNYRPHITFIVLTKMHNLRIYKKNIHKQERA
AQQNIKPGTVIDKHVVNPVLSEFYLNSHSTFQGTAKTPRYTLLFDTSKMEADEMQG
IVHALAYNFQIVNMAVSLPSPVMIASRMAKRGRCNYVAMFGDGSESSDNGKNNEK
DAVELNKQLSYINKPLEVVRFNA

>Dirofilaria_immitis.nDi.2.2.2.t06656

MAGCSGVEKVRTQLERTSVSEPPVYLTEKIAPGSRAYESLDVVTNVWGLIPHKNIF
VYRYDFRLFEEYPPKGRSPVMKEVTKQTRNDYLSVDRKTKCLTIYQTLLKRQKQFF
GDVHSLIYDRASILYSLRQLPFPKDEQRVAFFLNPNELPEDMVAEDCVKIHVYVKPC
KEDFHLSMKDLKSGVSNNPDEINRNLQQFLEILTMQEVFFMEGRFISYGSGECYLL
EPDQFGFGERDVPELREGKYVAVGASKGVRIIEGPQEGGMNAALIIDVKKAAFHID
NQSLLEKIEHILGKRGDLMRGIDQQSIAILNKALKGLYVLCNYGKKRAFMVAGISKKN
ARTSKLITKSGEVSVEEYFKMKYSMKLKYPAMPLVMERCQPKNNFYPIEVLSVCVN
QRVSIGQQTSSQVQTMIRACAAVPSLRLHQTNTLSKAMKLDNAKGNRWITNCSVGI
TNNLAFSARVLPAPAIEYGVNGWIKPDEKTSWKVEKNQYLIPAICKNWYAAALMGP
RERMNENLFRNYIRIFLQHCREHGMEMGDPLGCEYIRRAKQQDVEPLIAKAKSLGA
TFIHFVLADELNYHAHIKYIESQEQIVTQDLKATTALAVTLQHKRQTLDNIVNKTNIKM
GGLNYSVHLETNCDQWITKPQFLIVGLDIAHPAISMVSKRDQNIVPSVVGYSANIKK
HPLDFIGGYRYAKAQMEELVDDTIQQIFSDLLRYFNANRGKPPMHLFIIRDGVSIGQ
YKYIMNIEVEQIKKACQMVGGPGFRPHITFIVLTKMHGLRIYKKNINKQEGSAKQNIK
PGTIIDKYVVNPVINEFYLNSHSTFQGTTKIPRYTLLFDTSEMKADVMQGIVHALAFD
FQIVNMAVSRPSPVMIASRMAKRGRCNYVAMYGDEMENSDNGKGIEKDLVELNN
KLSYVSKPLEAVRFNA

>enterobius_vermicularis.EVEC_0000434401mRNA1

MSHIDVECVEKQVGELKLRSRVLLPGKTSPGQRCEGRIRVQTNVFGIKPLEDVPFF
RYDFRVLEEYPSKKQDPVFKEVTKQTRNDYVSIERKNKCVAVYLMVMHQEANFFG
DIGAFVYDRASTLYSLDKLKLADGEVKSFTVVARSLPKDLFSEDCLRVIVKVKQCAQ
EFQLSSMDLKSGFNLNPDLVSRSLQQFYELLLSQDAFFTEGRFVCYGSGENYLYN
PLDFGFSAQETPGLPDGKYIGVGAIKGVKIVEGSQDRPTLSFSVDVKKAAFHIELQS
IAEKVSDICNCVGVRLDDRHVKMLSKMLKGLYVRCNYGKRRTFIISGISEQSVDSFK
LEKDGKMIGMARYFMDKYKIALKFGKLPLVVEKTSRGTHYYPMELLFVCENQRVNL
SQQSSRQVQQMIRACATVPSMRKRQMLDMVSALNLDGGRKNRWCREFQVHLAE
RTYELEARVLQKPTIVYGGDGRVEMKESGAWTITQNTAYLLPSVCERWIIIALVSAS
DRFSRNDLENYTALFLERCRSRGISIRKPMEVIHIPRAREQDVDIAFKNAEKMSARFI
HFVTSENLKYHEKIKFLESQYQILTQDLMTKNAAVIRKRPQTLDNIVHKTNLKLGGIN
FDLHLESEEAQKWIGRKDRLIIGMDLTFTGVPSKDKSSKSPSVIGYAMNCHAHPLD
FTGGYRFCWTINEEVGDKTFCDIVSESLQLAKKNRRVPVHLVILRDGVSEGQYKYA
IEKEVQDVKDACARVGGAKYKPYITFVLATKRHQVRIYRSKIDPSRRAVDQNIPPGT
VVDTEVVNPVYNEFYLNSHAAIQGTAKTPRYNVLYDNSSMSSDEVQGMVYALTFN
MQIVNQPCSLPAPLIIADQMAARGRSNFIAFCGRSSSSVGSLDLDHINSELGYMGK
ELSNYRFNA

>litomosoides_sigmodontis.nLs.2.1.2.t06959RA

MAAGFDVEKITVQMGQASVREPPIYLAEKVPPGTRKFESLDVVTNVLGLIPQKNIPV
YRYDFCVLEEFPPKKGSPFFKEVSKQTKNDYLTVDRKTKCLLVYQALLKREKQFFG
NMDSLIYDRASILYSLRKLPMPKDEQRVTFFVMPDELPVNIVCEDCIKIHTHVKPCK
EDFQLTMSDLKSCVSNNLDNINRSLQQFLEILVMQEVFFMEGRFVSYGTGECYLM
DPNQFGFGDRDAPLLEEGKYVAIGAAKGVRIIEGPRGKEGINAALVFDVKKAAFHID
NQNLLEKVEHILRKRVDLTRGIGQQSITVLNNALKGLYVRCDYGKRRAFIIAGVSKE
NARTSKLVTKDGEMTIEKYFETKYSMKLKYPTLPLIMEKCQPKSNFYPIEVLSVCEN
QRVSKGQQTSSQVQTMIRACATLPSIRLQQTNVLSKAMKLDNSNRNKWMDKCNV
TVTSNLEFKARVLPMPAIEYRTNGWVKPCEKTSWTDGRNQYLIPAVCKNWCAVAL
MGQRERFNEHQFRNYIQIFLKHCRRHGMEMGDPLVLEYIHRSKQQDIESLVLKAKR
LGATFIHFVTADELSYHAHMKYIESQEQIVTQDLKASTAIAVAMQNKAQTLDNIVNKT
NIKMGGLNYSVHLETDCDGWLMKTGLLIVGLDIAHPVWMSSRKDQNGVPSVVGYS
ANIKKHPLDFIGGYRYGKANAEEMVDDTIQHIFADVLQYFNANRGKPPVHLFVIRDG
ISVGQYKYVMKYEVENIKKACQMVGGPGFRPHITFIVLTKMHRLRIYKKNIRKQDRG
AEQNIKPGTVIDEYVVDPTINEFYLNSHSAFQGTTKTPRYTLLFDTSGMKADVMQGI
VYALAYDFQIVNMAVSRPSPVMIASRMAKRGRSNYMAMYGDDSENSDNGNGIEK
DLVELNNRLSYISKPLEVVRFNA

>loa_loa.EN70_11666

MNDLKSCVSNNPDEINHSLQQFLEILAMQEVFFMEGRFVSYGAGECYLMYPNQFG
FGERDTPELEEGKYVAVGAAKGVRIVEGPRGEGGINAALVIDVKKAAFHVDNQCLL
EKVECILRRRVILMRGIDHLSIAILSKALKGLFVRCNYGKNRAFTIGGVSKENARTSK
LVSRTGEMSVEKYFEMKYSVKLKYPTLPLIMERCQTKSNFYPMEVLIVCENQRVSK
GQQTPSQVQTMIRACATVPSLRLQQTNTLSQAMKLNSSNQNKWMAKCNVAVTNN
LTFTARVLPTPSIEYRTNGWIKPSEKTSWTVGKYQYLIPGVCRNWYAVALMGPRER
FNEHQFRKYMDIFLQHCRLHGMEMRDPLKYVYIPHAKQQNVEPLITEAKSLGATFI
HFVTADELNYHAHIKYIESQEQVVTQDLKASTALSVTTQNKRQTLDNIVNKTNIKLG
GLNYSVHLETNCDRWLTKTGLLIVGLDIAHPYFAVPRKDRSCVPSVIGYSANIKKHP
LDFIGGYRYGKAEVEELTDDTIQQIFSDILRYFNANRGEPPKHLFVIRDGISTGQYKY
VMNTEVEQIKKACQMVGGPGFRPHITFIVLTKMHSLRIYRKNIRKQERAVEQNIKPG
TIVDKHVVNPVLNEFYLNSHSAAQGTTKTPRYTLLFDTSGMEADVMQGIVHALAYN
FQIVNMAVSLPCPVMIAARMAKRGRCNYIAMYGDESENSDNGSGVEKDLIELNNQ
LSYISKPLEAVRFNA

>onchocerca_flexuosa.OFLC_0000331401mRNA1

MFFPRGRFVSFGSGETYLMDPSQFGFSERDMPELEGGKYIAIGASKGVRIIEGPQE
GGINAAMVIDVKKAAFHADNQDLLEKVECLLRKRSDLKRGVDQQSIAILNKALKGLY
VLCNYGKKRAFTVTGVSKENARTSKLVTKSGEMSVEKYFEMKYSMKLKYPTLPLIM
ERSQPKNNFYPIEVLSVCENQRVSKGQQTSSQVQTMIRACATVPSLRLQQTNALS
KAMKLDATGENKWMKNCSVVVTNNLMFPARVLPSPDIEYRINGWVKPSEKTSWLT
GKNQYLIPAVCNKWYAVALIGPRERMNENLFRNYIRIFLQHCRQHGMEMSEPLGC
EYIRRANQQDIEPLIIKAKNLGATFIHFVTADELNYHGHMKYIESKEQVVTQDLKATT
AVAVAVQNKRQTLENIVNKTNIKLGGLNYSVHLETNCDKWLTKAGFLVVGLDIAHP
AVSMVSRKDRNFVPSVIGYSANIKKHPLDFIGGYRYCKAEMEELVDDTMQEVFSYI
LRYYKASRGEPPNHLFVIRDGVSAGQYKYVMNTEVQQIKKACQMVGGPNFCPHIT
FIVLTKMHNVRLYKKNIHKQERPAEQNIKPGTIIDKHVVNPVLNEFYLNSHLAFQV

>onchocerca_ochengi.nOo.2.0.1.t05983RA

MHVIREKMKKAAFHMDNQDLLEKVECVLRKRIDLKRGVDQQSIAVLSRALKGLYVL
CNYGKKRAFTVAGVSKENARTSKLVTKSGEMSVEKYFEMKYSMKLKYPTLPLIME
RCQPKNNFYPIEVLSVCENQRVSKGQQTSSQVQTMIRACATAPSLRLQQTNALSK

AMKLDATGENRWMKNCNVVVTNNLTFPARVLPTPAIEYRINGWVKPNEKTSWMT
GRNQYLIPAVCNKWYAVALMGPRERMNENLFRNYIRIFLQDCRQHGMQMSEPLG
CEYIRRANQQDVEPLIIKAKNLGATFIHFVTADELNYHGYMKYIESREQVVTQDLKA
TTAVAVAMQNKRQTLENIVNKANIKLGGLNYSVHLETNCDNWLTKAGFLIVGLDIAH
PAVSMVSKKDRNIVPSVIGYSANIKKHPLDFIGGYRYGKAEMEELVDDTMQEIFSYI
LRYYKTSRGEPPKHLFVIRDGVSAGQYKYVMNTEVQQIKKACQMVGGPNFCPHIT
FIVLTKMHNVRLYKKNIRKQERPAEQNIKPGTIIDKHVVNPVLNEFYLNSHSAFQGTT
KIPRYTLLYDTSDMKADVMQGIVHALTYNFQIVNMAVSCPSPVMIAARMAKRGRCN
YVAMYGDESENSDNGSGIEKDLTQLNNQLSYISKPLEAVRFNA

>Onchocerca_volvulus.OVOC1650

MAAKYSDVEKVRMELGQASVSELPVRLAKKVAPGTRACESLDIMTNVWGLISRENI
PVYRYDFRVLEEYPPKKGTPAFKEVTKQTRNDYLAVDRKTKCLAIYQTLLKHQKQF
FGDMYSLIYDRASILYSLRKLPIPEDERRVEFFLKPNELPKNTVADDCLQVHVFVKP
CKEDFQLTMNDLKSGVSNDPDKINRSLQQFLEILAMQEVFFMQGRFVSFGSGETY
LMDPSQFGFGERDMPELEEGKYIAIGASKGVRIIEGQREGGINAALVIDVKKAAFHI
DNQDLLEKVECVLRKRIDLKRGVDQQSIAVLSRALKGLYVLCNYGKKRAFTVAGVS
KENARTSKLVTKSGEMSVEKYFEMKYSMKLKYPTLPLIMERCQPKNNFYPIEVLSV
CENQRVSKGQQTSSQVQTMIRACATAPSLRLQQTNALSKAMKLDATGENRWMKN
CNVVVTNNLTFPARVLPTPAIEYRINGWVKPNEKTSWMTGRNQYLIPAVCNKWYA
VALMGPRERMNENLFRNYIRIFLQDCRQHGMQMSEPLGCEYIRRANQQDVEPLIIK
AKNLGATFIHFVTADELNYHGYMKYIESREQVVTQDLKATTAVAVAVQNKRQTLENI
VNKANIKLGGLNYSVHLETNCDNWLMKAGFLIVGLDIAHPAVSMVSKKDRNIVPSVI
GYSANIKKHPLDFIGGYRYGKAEMEELVDDTMQEIFSYILRYYKTSRGEPPKHLFVI
RDGVSAGQYKYVMNTEVQQIKKACQMVGGPNFCPHITFIVLTKMHNVRLYKKNIRK
QERPAEQNIKPGTIIDKHVVNPVLNEFYLNSHSAFQGTTKIPRYTLLYDTSDMKADV
MQGIVHALTYNFQIVNMAVSCPSPVMIAARMAKRGRCNYVAMYGDESENSDNGS
GIEKDLTQLNNQLSYISKPLEAVRFNA

>syphacia_muris.SMUV_0000813301mRNA1

MDQDTNDLVKKVAAMKVNDLQFATKRSEGLESQAKCVVASNIWEIRLLKDVPFYR
YEICVLEEYPTKDEKHAFKELTKKTRDDYQTVERKNKCVMVFETLLQKERAFFGNA
MSYVYDMASIVFSLNKLKIDGEKKFSIERGSLPEGLFPLDCIQVIVSVKPCVEDFVIT
AKECFSHSYLLKQFQALVQVYDLLLSQEAFFSKDVFVSYGASESYLLEPWNFGFDK
GLPDFPGGKYVGVGCCKAVRVIEGQGAPTLSVCLDVKKAIFHKDCQNLVEKTTEIC
KSGINNLEAWQVKNLEHFLKGLHVRCCYGKFRTFSIKGLSTGNAEKSRVQIREKSV
SIVQYYKEKYNKVLQYKHLPLVIEKTSKGYNYYPMELLEVCENQRVHNAQQLPVEA
QQLIRACAIPPSERLRQIACIARAMKLNNGKSNSWLKEFGLEISKKNLQVTARVIRKP
NVRYGMGAVFEIEDKGEWKMSPNTELLIPAACDLWCVYALIANSDTFNHCDLKSYV
QFFLDRCQKRGMKMKNPVEVALLHRPAEADLEKCFQLANKSHVKFIHFITTQSLPI
HGRMKYLELKYRIITQDVLSKTAIAVKRKPLTLDNIINKVNVKLGGMNFELFPASLNF
KTWFLSNEKLVISYKLSPVHLKTGKSVKSLSVIGLSSNCQKHPQKFSCGYRYTHSN
KDGETAVCDIVAESLLLVKKNRQIPDHIIIFRSGIPSAAYDYALKKEVVDIEEGCSKAF
GKSYQPKITFVVVTTAHNVRLFRPESSGKGRTFDQNVPVGTVVDRQLVSNQLTEF
YLNSHAAMQGTSKIPRYDVLFDTFGMSSDAIQSLAFTLCFDLQIVNHACSLPAPVMI
ARDMATRGENDISIFWQSWLKGSESCDIQCLNEQLGHMSKPLADLFLTT

>thelazia_callipaeda.TCLT_0000080401mRNA1

FSQRIRDSEEVKHESKRIADRKIKPSSTCLAQKVAPGTRAGEYIEVVTNVWGMIPQ
KNIPIFRYDFRVLEEYPPKKGSPAMKEVTKQTRNDYIAVERKSKCVIVYQTLLKREK
QFFGPPDSIVYDRAAILYSLKKLPITGDESQITFYLDPKELPANTVGDECVKIHVNVK

PCKDDFQLTMNDLKNSVSNNPDEVNRSLQQFLEILAMQEVFFTHGRFVSYGTGQC
YLLDPDQFGFKAQDMPELQEGKYVAIGASKGVRVIEGPKGSGGINAALVIDVKKAA
FHIDNQDLVEKAGNILKRKTELNRGLDQQSIAVLNKSLKGLYVRCGYGKERTFAIGG
ISKDTPRSLKLVTKTGEMSVEQYFFKKYAIRLKYATLPCIMERSQPRNKFYPMEVLS
VCENQRVSQGQQTSNQVQAMIRACATLPSRRISQTAALTEAMKLRGQIGNRWLK
GCDVTITDTLTLTARVLPAPSIEYSTNGRLKPNEKTSWSVGNNHYLIPATCKNWFAV
ALMGPRDKLNETQFRTYIRTFLLHCRQHGMEMSEPLGCEYIYRTKQQDVESLISKA
KRMGATFIHFVTADELDYHAQMKYVESQEQILTQDLKASTAAAVVLQNKWQTLDNI
INKTNVKLGGLNYSVHLEKNCDDWLLKPGLLVVGLDIAHPPVSAVLGKDRHVVPSVI
GYSANIKKHPLDFIGGYRYGKAEMEELMDESIHDIFVDIMRYFKNNRGKPPTHLFVI
RDGVSEGQYKYVLNTEIEQIKKACNTVGGSFYRPHITFTVLTKMHSIRLFKKNIAHQ
AKAVEQNIKPGTVIDKHVVNPVLNEFYLNSHSAFQGTTKAPRYTLLFDSSKMSADE
MQGIVHALAYDFQIVNMAVSRPSPVMIASRMAKRGRSNYIAMYGDDDSQHSSGSA
VEADLVNLNNQLSYISKPLEGVRFNA

>wuchereria_bancrofti.WBA_0000646201mRNA1

MFLSYVAVDRKMKCFAVYQALLKREKQFFGAVDTLIYDRASTLYSLRKLSFPRDEV
RKTFFLNKNELPMNIVGGDCVRVHVNIKPCKEDFQLSTHDLKNCVSNNPDEINNSL
QQFLEILAMQEVFFMEGRFVSYGTGECYLMDPSQFGFGDRDMPELQEGKYIAIGA
AKGVRIIEGLGGKRGISAGLVLDAKKAAFHIDNQDLLEKVESIFRRRADLSRGIDQQ
SILILSKALKGLYVRCTYGKNREFVIAGISKENARTSKLVHKTGEMSVENYFVMKYSI
KLKYPILPLIMERCQPKNNLYPIEVLVVCENQRVSKGQQTSSQVQTMIRACATDPS
VRLKQTNILSKAMKLDNSNQHKWMGKCNMTITDNLMFPARVLPQPTIEYHTNGWV
NPNEKTVWMDGRNYYLIPAVCEKWYAIALMGPRERLREDQFGTYIKMFLERCRKH
GMQIKDPIGYEYIRRAKEQDVESLMIKAKKSGATFIHFVTADELNYHANIRYFSCDK
WLGRPGFLIVGLDIAHPAYSMLPNKDRNSVLSVVGYSANIKKHPLDFIGGYRFAKAE
MEELIDDTMQQIFSDLLRYFNANRGKPPTHLFVIRDGISVGQYKYVMNTEVEQIKQA
CQSVGGRNYRPHITFIVLTKMHNLRIYKKNIHKQERAAQQNIKPGTIVDKHVVNPVL
SEFYLNSHSTFQGTTKIPRYTLLFDTSRMEADEMQGIVHALAYNFQIVNMAVSLPSP
VMIASRMAKRGRCNYVAMFGDESENSDNGRGVEKDVVELNNQLSYISKPLEVVRF
NA