bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The necessary component of regulatory interactions: identification and characterization of putative regulatory motifs. Case study on the control of seminal fluid proteins through regulatory hubs in *D. melanogaster*

Irina Mohorianu[1,2,3]

[1]School of Biological Sciences, University of East Anglia, Norwich Research Park, NR4 7TJ, UK.

[2]School of Computing Sciences, University of East Anglia, Norwich Research Park, NR4 7TJ, UK.

[3]Oxford Vaccine Group, Medical Sciences Division, University of Oxford

e-mail: irina.mohorianu@paediatrics.ox.ac.uk

Running title: Putative regulatory interactions based on the existence of motifs

Keywords: gene regulatory networks, regulatory hubs, transferome, miRNAs, transcription factors.

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Highly precise, yet flexible and responsive co-ordination of expression across groups of genes underpins the integrity of many vital functions. However, our understanding of gene regulatory networks (GRNs) is often hampered by the lack of experimentally tractable systems, by significant computational challenges derived from the large number of genes involved or from difficulties in the accurate identification and characterization of gene interactions. The proposed case study is based on a tractable experimental system: the genes encoding seminal fluid proteins transferred along with sperm (the 'transferome') in *D. melanogaster* fruit flies. These proteins resulting from the transferome genes are core determinants of reproductive success, yet we know little about the mechanisms underlying their tight, responsive and precise regulation. Using only genomic information, we identified potential regulatory motifs that linked the transferome genes in an interaction network. This analysis predicted the existence of variation in the strength of regulation across the transferome genes and revealed evidence for putative 'hubs' linked to either transcriptional or post-transcriptional control. We tested the role of post-transcriptional regulation in this gene set by directly manipulating the miRNA biosynthesis pathway. This affected the reproductive function of the transferome genes by abolishing the ability of males to respond to the threat of sexual competition. The results identified regulatory mechanisms that can underpin robust, precise and flexible regulation of important, fitness-related genes.

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1. Introduction

Gene regulatory networks

Genes rarely, if ever, function in isolation from one another. They are often interconnected within gene regulatory networks (GRNs) that regulate a specific pathway or function. Such GRNs are of vital importance in regulating ubiquitous aspects of development and organismal function. Genes may be regulated at the transcriptional, or post-transcriptional level via different mechanisms. Transcription factors (TFs) control the rate of gene transcription by binding specific DNA motifs, usually upstream of the coding region {Dai, 2012 #4136}. Post-transcriptional regulation can be achieved by microRNAs (miRNAs) {Bartel, 2004 #3776}, a particular class of small RNAs (sRNAs), which target mRNA transcripts, inhibiting translation into proteins. miRNAs are processed from a hairpin-like structure by Drosha and Dicer-1 enzymes (figure 1), and then loaded into the Argonaute protein, part of the RNA Induced Silencing Complex (RISC), which guides the miRNA to the target mRNA {Bartel, 2009 #4134}. In animals, miRNAs generally induce translational repression in their targets via matching of the miRNA 'seed' sequence (at positions 2-8 from the 5' end) to the 3'UTR region of the target mRNA{Brennecke, 2005 #4157}. Other small interfering (si)RNAs (e.g. 21nt siRNAs, repeat associated RNAs rasiRNAs, promoter associated pasRNAs and ~27-30nt piwi associated piRNAs) are processed by Dicer-2 and recruit different Ago proteins {Kim, 2009 #4158}; many details of their regulatory functions are not yet known {Kim, 2009 #4158}, however early studies indicate a role in transcriptional regulation.

Although our knowledge of gene regulation is rapidly growing, the identification and comparison of inter-relationships between co-regulated genes in GRNs poses significant challenges {Petralia, 2015 #4394}. For example, GRNs are often inferred from gene expression profiles, which may have a variable signal to noise ratio {Mitra, 2011 #4395; Penfold, 2011 #55}. GRNs can also be identified by using protein-protein interactions (e.g. {Filkov, 2005 #4396;Giot, 2003 #4634}), from steady state and manipulated datasets (e.g. knock outs) and also via the integration of gene expression with metabolomic data (e.g. {Gargouri, 2015 #4397}).

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

GRNs range from simple to the very complex, comprising many hundreds of genes and transcriptions factors {Milo, 2002 #4407}. There is a growing realisation of the valuable insight that can be gained by identifying and comparing GRNs across different cells and tissues over time (e.g. {Barabasi, 2004 #4400;Blais, 2005 #4402;Gaiteri, 2014 #4401;Linde, 2015 #4399}). In the study of evolutionary biology there is much interest in determining how core features of GRNs such as topology, composition, degree of connectivity, robustness to mutation, clustering and stability change under selection (e.g. {Ravasz, 2002 #4408;Luscombe, 2004 #4403;Ciliberti, 2007 #4404; Crombach, 2008 #4406; MacNeil, 2011 #4405}). A key, and so far unanswered question, is how selection acts in different environments to achieve network stability and indeed whether one can measure the degree of stability from characterising core network features (e.g. {Ciliberti, 2007 #4404}). The general emerging idea is that highly connected genes within networks are likely to be linked or co-regulated through one or multiple hubs that are essential for network organisation and hence themselves targets of selection. GRNs may also represent an efficient way to capture and maintain the effects of beneficial mutations, or to maintain selectively neutral ones {Crombach, 2008 #4406}.

An additional hurdle in the study of GRNs can be the difficulty in identifying an appropriate set of genes in which to study fundamental network features, both at the level of gene expression and the resulting phenotype. To facilitate the understanding of such a system, it should ideally (i) comprise a tightly linked network of genes, (ii) represent a set of genes within a defined biological process and/or localized expression, (iii) be genetically tractable for experimental testing, and (iv) produce a well-defined and measureable phenotype. The set of genes that encode the non-sperm components of the ejaculate in male *D. melanogaster* fruit flies {Findlay, 2008 #3633} (hereafter the 'transferome') fulfils these criteria. They represent a potentially valuable exemplar for testing features of GRNs because they (i) show co-ordinated expression {Bertram, 1992 #1210;Monsma, 1988 #493;Herndon, 1997 #1435}, (ii) have defined functions and easily measureable phenotypes {Chapman, 2001 #2713}, and (iii) can be subjected to controlled, experimental genetic manipulations.

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Functions and significance of the reproductive transferome

Seminal fluid proteins that comprise the transferome are of key importance across many animal taxa {Findlay, 2008 #3633;Ram, 2007 #3619;Sirot, 2014 #4180}. They are more than a buffer to maintain sperm osmotic potential {Arnqvist, 2005 #3181;Chapman, 2008 #4409}. In *D. melanogaster* these remarkable substances cause a profound remodelling of female behaviour, physiology, gene expression and fitness (e.g. {Chapman, 2001 #2713} {Gioti, 2012 #4133}). Individual seminal fluid proteins affect egg production, sexual receptivity, feeding and nutrient balancing, sleep patterns, sperm retention and usage, water balance and antimicrobial peptide production (reviewed in {Sirot, 2014 #4180}). These actions are fundamental to reproductive success {Wigby, 2005 #2995;Fricke, 2009 #3927;Fricke, 2009 #3752}. Seminal fluid components in *D. melanogaster* have been well characterized at the genetic, functional and structural levels {Ram, 2007 #3619}. Isotopic 15N labelling has defined a set of ~138 extracellular proteins secreted by the male accessory glands, ejaculatory ducts and bulb, plus non-sperm molecules from the testes that are transferred to females during mating {Findlay, 2008 #3633}.

The transferome as a GRN that responds to the socio-sexual context

Male *D. melanogaster* exposed to rivals prior to mating for at least 24h mate for significantly longer and transfer more of key seminal fluid proteins into females {Wigby, 2009 #3855}. Such responses are precise, robust and flexible {Bretman, 2011 #4155;Bretman, 2012 #4152}. They result in significantly increased male fitness {Bretman, 2009 #3863} because, in response to increased transfer of seminal fluid proteins, females lay significantly more eggs and become significantly less sexually receptive, effects that increase a male's representation of offspring in future generations {Bretman, 2009 #3863}. Hence ejaculate composition can be modified in a highly sophisticated manner in response to social and sexual context {Wigby, 2009 #3855;Sirot, 2011 #4156}. This is also underpinned by differential expression in transferome-encoding genes {Mohorianu, 2017 #4369}. Together these data support the idea that males calibrate responses to sexual competition with remarkable precision and suggest

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that the transferome genes are linked in a tight and highly co-ordinated regulation in response to the environment {Mohorianu, 2017 #4369}. However, little is yet known about how this is achieved.

We hypothesise that an effective way in which to regulate > 130 individual transferome components within a GRN is to manage them in 'sets' controlled by the same regulator. This could facilitate rapid and co-ordinated expression of groups of genes when required. This level of control may be achieved by transcription factors that enhance the transcription of sets of genes, or by small RNAs that bind to mRNA transcripts and repress the translation of functionally linked groups of proteins (e.g. {Tibiche, 2008 #4410}). We adopted a predictive approach to test these ideas. We first tested whether we could identify known sequence motifs shared between members of the transferome gene set. We used sequence analysis to detect motifs in the 3' and 5' UTRs of all transferome genes to test for regulation by microRNAs (miRNAs) or transcription factors (TFs), respectively. The results showed evidence for shared (putative) regulatory regions at either the 5' UTR, 3' UTR or both and variation in the number/type of shared regulatory sequences. The results also suggested the presence of regulatory 'hubs' controlling specific sets of transferome genes. We further investigated this prediction by manipulating small RNA biosynthesis pathway directly in order to measure the effect on the transferome phenotype of knocking down an upstream major component of miRNA (Drosha) biogenesis.

2. Methods

To detect regulatory signatures, we focused on genes encoding the *D. melanogaster* seminal fluid transferome components {Findlay, 2008 #3633} as described in v6.11 of the *D. melanogaster* genome build. This resulted in a working set of 136 transferome genes. Our strategy was first to ascertain whether the 5' UTR, or 3' UTR of these genes were enriched in motifs linked to TFs or miRNAs, respectively; next a more general window based approach was used for the identification of other motifs not related to known regulators. All analyses mapping miRNA seed sites or TF binding motifs to 3'UTR and 5'UTR regions were

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

performed on unique motifs at the transcript level. To account for the presence of different transcript isoforms corresponding to the same gene (which partially share portions of the UTRs) we also generated a collapsed version of the results, at gene-level (see tables S1-S4).

(a) Regulation of transferome genes by known miRNAs

A conservation analysis was first conducted to identify all miRNAs in the *D. melanogaster* genome. All mature miRNAs from the 12 *Drosophila* subspecies {Kozomara, 2014 #28} were mapped on the *D. melanogaster* genome and the miRNA loci then determined using criteria based upon the identification of miRNA hairpin-like secondary structures (specifically: adjusted minimal folding free energy (aMFE) < -20 and no branching adjacent to the miRNA/miRNA* duplex) a similar approach as in {Mohammed, 2018 #4633}. We then determined all 7 and 8nt seed regions for all mature miRNAs. miRNAs sharing seed regions (perfect identity) were collapsed under one entry. Seeds were mapped to the 3' UTRs of the transferome gene transcripts (with full length matching and no mis-matches or gaps allowed). The enrichment of miRNA usage was calculated by comparing the number of target genes for each miRNA seed site, on the transferome transcripts and on all *D. melanogaster* transcripts, using identical targeting criteria for both analyses. We used the Fisher exact test to evaluate whether the observed number of putative targets was in line with the expectation across the *D. melanogaster* genome or whether it was enriched/depleted for transferome transcripts.

(b) Regulation of transferome genes by known TFs

Using similar methodology as in (a), we searched for putative TF binding sites on the 5' UTRs of transferome transcripts using TF motifs (5-50nt long) from the 'Redfly' (http://redfly.ccr.buffalo.edu/), FlyTF (http://www.flytf.org/) and flyatlas (www.flyatlas.org/) databases. Due to the high redundancy in TF motifs, we collapsed motifs with identical sequences and merged their identifiers; the analysis was conducted on all unique motifs. The enrichment analysis was done using the same approach and thresholds as for the miRNAs.

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(c) Regulation by unknown regulatory elements (sliding window analysis)

The last step was to conduct a sliding windows analysis (with lengths varying from 9 to 21nt, in increments of 2nt; the overlap of consecutive sliding windows is L-1, where L is the length of the window) for an unbiased test for regulatory elements. The input for this analysis consisted of the 5' UTRs and the 3' UTRs of transferome gene sequences. Next, all input fragments were mapped against all other entries in each dataset, allowing up to 2 mismatches and eliminating self-matches. Fragments with low sequence complexity were also eliminated. To identify co-regulated regions we used only matches on the positive strand.

3. Results and Discussion

Overall, we showed significant over-representation among transferome genes of 37 miRNA seed sequences and 42 TF binding motifs. This was accompanied by a significant under-representation of TF, miRNA and siRNA binding sites among the transferome set overall. These results reveal the tight nature of regulation of transferome genes and reveal how a diverse set of functionally important gene products can be regulated.

(a) Regulation of transferome genes by known miRNAs

We first evaluated the over-representation of miRNA target sites among the 3' UTRs of the transferome genes, when compared to the probability of occurrence in the entire set of *D. melanogaster* 3' UTRs. We found 37 miRNAs whose targets were significantly enriched amongst transferome transcripts (table S1a). The most significantly enriched target site was that of miR-4943-5p, which has seed sites in 80 transferome 3' UTRs (corresponding to 42 genes). In contrast to the typical pattern of miRNA biogenesis, the miR-4943 locus spans the sense strand of an exon/intron boundary in the gene CG5953, rather than from an intronic or intergenic region. Interestingly, this miRNA appears to be lineage-specific (i.e. restricted to *D. melanogaster*) and expressed at relatively low levels {Berezikov, 2011 #4135}. Further investigation of this putative miRNA may reveal its role in the regulation of so many transferome genes.

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In total, the targeted 3' UTRs of all enriched miRNAs correspond to 71 genes, approximately half of the transferome set. We observed no particular functional enrichment for the subset of 71 genes; instead these genes correspond to a broad range of processes within the transferome.

We next explored the presence of miRNA seed sites amongst transferome genes, regardless of any enrichment compared to the entire genome. We show the predicted target genes of each known miRNA (table 1b) and the number and identity of miRNA seed sites on every transferome 3' UTR (table S1c). The interactions between miRNAs that can target the transferome genes and their corresponding targets are presented as a Cytoscape network diagram {Shannon, 2003 #4622} (figure S1). It is clear from the node sizes that the majority of known miRNAs were predicted to target very few transferome genes. Indeed, 213 miRNAs had only 1-2 seed sites amongst all transferome 3' UTRs. However, it was also apparent that some miRNAs have putative target sites in many different genes, and so have the potential to act as regulatory 'hubs', controlling many different genes simultaneously. The miRNAs with the highest number of predicted target genes were miR-4943-5p (42 genes), miR-4953-3p (17 genes), miR-7-3p (14 genes), miR-315-5p (11 genes) and miR-9369-3p (10 genes) (figure 2). To investigate if the genes targeted by the same miRNA shared functional profiles, we performed a GO enrichment analysis on groups of $\geq$ 10 genes, using the list of 136 transferome genes as a reference set (table S5; g:Profiler http://biit.cs.ut.ee/gprofiler/index.cgi {Reimand, 2016 #4623}). We found no GO enrichment of terms for the targets of miR-4943, miR-4953 or miR-9369. However, significant enrichment of some biological process terms was found for miR-7 and miR-315 targets. Putative miR-7 targets were enriched for "organonitrogen compound metabolic process", which characterised 9 of the 14 genes (*Acp62F*, *trithorax*, *Peritrophin-A*, *ND-51L2*, *Ggt-1*, *CG10862*, *CG10585*, *CG31704*, and *CG4815*). The products of these genes are all predicted to be involved with protein processing (e.g. proteases, protease inhibitors, histone modification and chitin binding). However, there is as yet no evidence that these 9 genes are expressed in a co-ordinated fashion, or whether their products have pleiotropic effects. For miR-315 targets, 3 of 11 genes were associated

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with "nervous system development" - *wurstfest*, *trithorax*, and *Esterase*-6. The products of these genes have diverse functions in translational and transcriptional control, and pheromone processing {Baker, 2015 #4624;Petruk, 2006 #4625;Chertemps, 2012 #4626}.

Of the 136 transferome genes, 104 had at least one putative miRNA target site incident with a 3' UTR transcript. The genes with the highest number of miRNA target sites were *trithorax* (putative sites for 50 miRNAs), potentially suggesting chromatin remodelling {Schuettengruber, 2017 #4635}, and *wurstfest* (putative sites for 42 miRNAs). Since these genes encode transcriptional and translational regulators, respectively, they may also require tight regulation themselves. Indeed, there is evidence in mice that genes whose products are involved in a regulatory role (such as transcription factors) have more predicted miRNA target sites in their 3' UTRs than housekeeping or structural genes {Zare, 2014 #4627}. Another 9 genes were predicted to have >15 binding sites corresponding to different miRNAs. Amongst those genes were three whose products potentially play a role in cell development – *CG18135* which is known to interact with the unconventional myosin Myo10A {Liu, 2008 #4628}, *CG10433*, which when over-expressed in male flies leads to defective microtubule organisation {Liu, #4629}, and *β-tubulin at 85D* which has been shown to regulate salivary gland migration {Jattani, 2009 #4630}. Another two genes, *polyphemus* and *Niemann-Pick type C2b* encode products involved in the immune response {Gonzalez, 2013 #4631;Shi, 2012 #4632}. The remaining four genes with >15 miRNA sites have no experimentally confirmed functions, but may be involved in chitin-binding (*Peritrophin-A*), calcium ion binding (*regucalcin*) and protein-folding (*CG2852*). *CG18067* encodes a protein of unknown function.

To gain further insight into whether a subset of genes, whose products are involved in similar biological processes, could be regulated by miRNA 'hubs', we created a network diagram in Cytoscape {Shannon, 2003 #4622} of 19 genes which have a role in the post-mating response (PMR) of females (figure 3). We know that ejaculate proteins that affect sperm storage and female behaviour are precisely controlled by the male fly in response to sperm competition, so we reasoned that these genes may be co-regulated by the same

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

miRNAs. As for the entire transferome gene set, the most prolific miRNA amongst the PMR subset was miRNA-4943. Of the 19 genes chosen, 9 had target sites for miR-4943 (*Acp26Ab*, *Acp36DE*, *Acp53Ea*, *Acp62F*, *antr*, *Ebp*, *lectin-46Ca*, *lectin-46Cb*, and *SP*). Although the term 'post-mating behaviour' was not found to be significantly enriched in the GO analyses of miR-4943 targets described above, the fact that almost half of the PMR subset have miR-4943 target sites suggests that this miRNA is still an important regulator of sperm storage and post-mating response genes. Other potential PMR regulators were miR-972 and miR-289, which both had complementarity to *CG10433*, *Ebp*, *EbpII*, *lectin-46Ca*, and *SP*. miR-972 was also predicted to bind *antr*. It is also apparent (figure 3) that some PMR genes have target sites for an abundance of different miRNAs (e.g. *CG10433*, *Ebp* and *EbpII*), and thus instead of being regulated by a single 'hub', these genes may require very tight control, mediated by many different regulators.

Overall, our results indicate that several miRNAs are predicted to regulate multiple transferome genes, thereby acting as regulatory 'hubs'. Groups of genes with seed sites for the same miRNA are not necessarily enriched for a particular function, suggesting that their co-ordinated regulation impacts on diverse reproductive processes. In addition, we observed considerable redundancy in miRNA seed sites for individual genes, i.e. genes with seed sites corresponding to numerous different miRNAs. This suggests that some transferome genes may require particularly tight regulation, potentially because they themselves are transcriptional or translational regulators {Zare, 2014 #4627}.

(b) Regulation of transferome genes by known TFs

Next, we evaluated the over-representation of TF binding motifs among the transferome 5' UTRs, when compared to all *D. melanogaster* transcripts (table S2a). In total, 29 unique TF motifs were significantly enriched in the transferome transcripts. These 29 motifs are the binding sites of a potential 30 different transcription factors. The motifs were distributed among a total of 27 genes. GO enrichment analysis of the 27 targeted genes revealed a significant over representation of genes encoding proteins involved in microtubule based processes (*wurstfest*, *α-Tubulin at 84D*, *β-Tubulin at 85D*, *α-Tubulin at 84B*, *Cytoplasmic*

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dynein light chain 2). This may suggest the potential for co-ordinated structural changes in transferome cells, potentially associated with secretory function. This would be interesting to test directly, using targeted genetic manipulations.

We considered the total number of TF motifs that were present in the 5' UTRs of all transferome transcripts, regardless of enrichment compared to all *D melanogaster* 5' UTRs, and presented it as a Cytoscape network in figure S2. Overall, we observed binding motifs for 76 known TFs amongst 43 genes in the transferome set (table S2b). Of these genes, 30 had 5' UTR motifs for Abd-B. Abd-B is known to be expressed in the secondary cells of the male accessory gland, and suppression of the Abd-B activator iab-6 in males affects egg-laying and receptivity in his mates {Gligorov, 2013 #4329}. The abundance of Abd-B binding motifs amongst transferome genes provides further support that this transcription factor plays an important role in the regulation of male seminal proteins.

A GO analysis (table S6) revealed the targeted genes of Abd-B were significantly enriched for the term "microtubule cytoskeleton" – α-*tubulin84B*, α-*tubulin84D*, β-*tubulin85D*, *Cytoplasmic dynein light chain 2*, *trithorax* and *CG2852*. These analyses again suggest that the co-ordinated regulation of microtubule function by TFs plays a key, and previously unknown, role in ejaculate secretion.

Of the 43 genes with at least one TF binding site motif (table S2c), the most targeted gene was *CG10433*, with 51 unique motif sequences in the 5' UTR. The 51 motifs represent binding sites for potentially 45 different transcription factors. *CG10433*, described in section 3a as encoding a protein involved in microtubule organisation {Liu, 2014 #4629}, was also one of the genes with the most 3' UTR miRNA seed sites. Indeed, of the 11 genes with >15 miRNA seed sites, 5 are also represented amongst the 11 genes with >15 TF binding motifs. In addition to *CG10433*, the other 4 genes were *CG18135*, *wurstfest*, *Peritrophin-A*, and *trithorax*. In addition, *Cytoplasmic dynein light chain 2* had 16 TF binding motifs, and 13 miRNA seed sites. This result suggests that these 5 genes, whose products have all been shown to play key roles in cell development are particularly tightly regulated. It would be very interesting to determine their functions as secreted proteins in the seminal fluid.

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Interestingly, TF binding sites were distinctly under-represented in the 5'UTRs of the 19 PMR genes. Only two PMR genes had putative TFBSs - *CG10433* and *Acp53Ea*. As mentioned above, the product of *CG10433* has a role in microtubule organisation, but has also been shown to reduce female receptivity to remating when overexpressed in males {Liu, 2014 #4629}. *Acp53Ea* had binding motifs for only one TF – Pannier. An explanation for the under-representation of TFBSs amongst PMR genes, is that binding motifs could be more prolific in the promoter regions of the PMR genes, rather than the 5'UTRs, and are therefore not captured by this study. Alternatively, it is interesting to consider that TFs may not be the primary regulators of PMR gene expression, and that these particular genes are regulated post-transcriptionally.

(c) Regulation of transferome genes by unknown elements

The sliding window analysis of the 5' and 3'UTRs of all the transferome genes showed evidence of potential co-regulation of the transferome genes corresponding to known TF, miRNA as presented above, but also potential novel co-regulatory sequences for both sets. Within these, there was also evidence for significant variation in the degree of shared regulation, with some genes showing regulatory similarities with > 50 transferome genes. Genes predicted as tightly regulated (sharing regulatory sequences at the 5' and 3' regions) included protein phosphatase Y regulator 1; Odorant receptor (Or)82a; Serpin 77Bc, 38F; male specific RNA 57Db. Examples with high 3' and low 5' regulatory similarities: Serpin 28F; Odorant binding protein (Obp)22a; Lectin 46Ca; Accessory gland protein (Acp)24A4 and Andropin). The reverse situation (tight 5' but not 3' regulation) occured in Met75Ca; NUCB1; Serpin 77Bb. To determine the probability of obtaining these results by chance we chose 130 genes at random and repeated analyses using the same parameters (5' and 3' 2kb regions, low complexity discarded). This procedure was then iterated 100 times. For 5' regions we detected ~4 times (standard deviation, sd=0.5) more putative shared regulatory regions in the transferome than in randomly selected genes. Similarly, for the 3' UTR there were 1.5 (sd=0.12) times more putative regulatory regions, rising to 2.7x more such regions (sd =0.33). The results suggest that sets of transferome genes vary in the degree to which they

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

345 are regulated, with some having tight regulation, as indicated by the presence of known and
346 unknown regulatory regions in the 3' and 5'UTRs. The functional significance of this is not
347 yet clear.
348
349 **4. Conclusions**
350 The results showed evidence for the presence of regulatory elements that modulate the
351 expression of seminal fluid transferome genes in *D. melanogaster*. Cross referencing the 5'
352 and 3' UTRs of transferome genes to known databases showed evidence for under
353 representation of regulatory motifs in general coupled with significant over-representation of
354 motifs for specific TF binding sites and miRNA seed sequences. Interestingly, several
355 miRNAs were predicted as putative regulatory hubs, with seed sequences mapping to
356 multiple transferome genes. We also observed variation in the degree of regulation across the
357 transferome genes, with some sets of genes putatively regulated via mechanisms operating at
358 both 5' and 3' UTRs. The prediction of transcriptional regulation of transferome genes via
359 known TFs was consistent with published reports (e.g. {Gligorov, 2013 #4329}). The
360 prediction involving miRNAs is novel and was supported by experimental validation.
361 Silencing of miRNA biosynthesis by *drosha* knockdown altered the expression of the
362 transferome phenotype and resulted in males that were no longer able to respond to
363 competition with male rivals by reducing the probability of remating by their current mates.
364 The results indicated that cross referencing of regulatory regions to existing databases
365 and unbiased methods for detecting regulation of unknown origin has the potential to reveal
366 signatures of gene regulation. This variation in number or type of regulatory interactions
367 would be interesting to study further. The potential fitness benefits of multiple layers of
368 regulatory control can be studied by manipulating individual regulatory components.
369 We propose that the layers of gene regulation, mediated by specific TFs and miRNAs
370 is important to facilitate a robust and precise response in many tens of different genes. The
371 next steps are to test this hypothesis experimentally on a genome-wide scale and to determine

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

whether this is an emergent property of efficient GRNs. Whether there is any functional significance to the potential for regulation by TFs versus miRNAs is not yet apparent, but will be important to resolve.

Competing interests. We declare we have no competing interests**.**

Funding. We thank the BBSRC (BB/L003139/1; research grant to Tracey Chapman, Tamas Dalmay and Irina Mohorianu) for funding.

References

1. Dai Q, Smibert P, Lai EC. 2012 Exploiting *Drosophila* genetics to understand microRNA function and regulation. *Curr. Top. Genet. Dev.* **99**, 201-235.
2. Bartel DP. 2004 MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-297.
3. Bartel DP. 2009 MicroRNAs: Target recognition and regulatory functions. *Cell* **136**, 215-233.
4. Brennecke J, Stark A, Russell RB, Cohen SM. 2005 Principles of microRNA-target recognition. *PLoS Biol*. **3**, 404-418.
5. Kim VN, Han J, Siomi MC. 2009 Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **10**, 126-139.
6. Petralia F, Wang P, Yang J, Tu Z. 2015 Integrative random forest for gene regulatory network inference. *Bioinformat*. **31**, i197-i205.
7. Mitra S, Das R, Hayashi Y. 2011 Genetic networks and soft computing. *IEEE/ACM Trans. Comp. Biol. Bioinformat.* **8**, 94-107.
8. Penfold CA, Wild DL. 2011 How to infer gene networks from expression profiles, revisited. *Interface Focus* **1**, 857–870.
9. Filkov V. 2005 Identifying gene regulatory networks from gene expression data. *Handbook of Computational Molecular Biology, Chapman & Hall/CRC Computer and Information Science Series*, **27**.
10. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, et al. 2003 A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727-1736.
11. Gargouri M, Park JJ, Holguin FO, Kim MJ, Wang H, et al. 2015 Identification of regulatory network hubs that control lipid metabolism in *Chlamydomonas reinhardtii*. *J. Exp. Bot.* **66**, 4551-4566.
12. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, et al. 2002 Network motifs: simple building blocks of complex networks. *Science* **298**, 824-827.
13. Barabasi AL, Oltvai ZN. 2004 Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101-113.
14. Blais A, Dynlacht BD. 2005 Constructing transcriptional regulatory networks. *Genes Dev*. **19**, 14991511.
15. Gaiteri C, Ding Y, French B, Tseng GC, Sibille E. 2014 Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. *Genes Brain Behav*. **13**, 13-24.
16. Linde J, Schulze S, Henkel SG, Guthke R. 2015 Data-and knowledge-based modeling of gene regulatory networks: an update. *EXCLI Journal* **14**, 346.
17. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL. 2002 Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551-1555.

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

18. Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, et al. 2004 Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**, 308-312.
19. Ciliberti S, Martin OC, Wagner A. 2007 Innovation and robustness in complex regulatory gene networks. *Proc. Natl Acad Sci USA* **104**, 13591-13596.
20. Crombach A, Hogeweg P. 2008 Evolution of evolvability in gene regulatory networks. . *PLoS Comput. Biol.* **4**: e1000112.
21. MacNeil LT, Walhout AJ. 2011 Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res* **21**, 645-657.
22. Findlay GD, Yi X, MacCoss MJ, Swanson WJ. 2008 Proteomics reveals novel Drosophila seminal fluid proteins transferred at mating. *PloS Biol.* **6**, e178.
23. Bertram MJ, Akerkar GA, Ard RL, Gonzalez C, Wolfner MF. 1992 Cell type-specific gene-expression in the *Drosophila melanogaster* male accessory-gland. *Mech. Dev.* **38**, 33-40.
24. Monsma SA, Wolfner MF. 1988 Structure and expression of a *Drosophila* male accessory gland gene whose product resembles a peptide prehormone precursor. *Genes Dev.* **2**, 1063-1073.
25. Herndon LA, Chapman T, Kalb JM, Lewin S, Partridge L, et al. 1997 Mating and hormonal triggers regulate accessory gland gene expression in male *Drosophila*. *J Insect Physiol.* **43**, 1117-1123.
26. Chapman T. 2001 Seminal fluid-mediated fitness traits in *Drosophila*. *Heredity* **87**, 511-521.
27. Ram KR, Wolfner MF. 2007 Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction. *Int. Comp. Biol.* **47**, 427-445.
28. Sirot LK, Wong A, Chapman T, Wolfner MF. 2014 Sexual conflict and seminal fluid proteins: a dynamic landscape of sexual interactions. In: Rice WR, Gavrilets S, editors. *Sexual Conflict* CSHL Press.
29. Arnqvist G, Rowe L. 2005 *Sexual Conflict.* Krebs JR, Clutton-Brock T, Editors. Princeton, NJ: Princeton University Press. 330 pp.
30. Chapman T. 2008 The soup in my fly: evolution, form and function of seminal fluid proteins. *PLoS Biol.* **6**, e179.
31. Gioti A, Wigby S, Wertheim B, Schuster E, Martinez P, et al. 2012 Sex peptide of *D. melanogaster* males is a global regulator of reproductive processes in females. *Proc. Roy. Soc. B* **279**, 4423-4432.
32. Wigby S, Chapman T. 2005 Sex peptide causes mating costs in female *Drosophila melanogaster*. *Curr. Biol.* **15**, 316-321.
33. Fricke C, Perry J, Chapman T, Rowe L. 2009 Conditional economics of sexual conflict. *Biol. Lett.* **5**, 671-674.
34. Fricke C, Wigby S, Hobbs R, Chapman T. 2009 The benefits of male ejaculate sex peptide transfer in *Drosophila melanogaste*r. *J. Evol. Bio*l. **22**, 275-286.
35. Wigby S, Sirot LK, Linklater JR, Buehner N, Calboli FCF, et al. 2009 Seminal fluid protein allocation and male reproductive success. *Curr. Biol.* **19**, 1-7.
36. Bretman A, Westmancoat JD, Gage MJG, Chapman T. 2011 Multiple, redundant cues used by males to detect mating rivals. *Curr. Biol.* **21,** 617-622.
37. Bretman A, Westmancoat JD, Gage MJG, Chapman T. 2012 Individual plastic responses by males to rivals reveal mismatches between behaviour and fitness outcomes. *Proc. Roy. Soc. B* **279**, 2868-2876.
38. Bretman A, Fricke C, Chapman T. 2009 Plastic responses of male *D. melanogaster* to the level of sperm competition increase male reproductive fitness. *Proc. Roy. Soc. B* **276**, 1705-1711.
39. Sirot LK, Wolfner MF, Wigby S. 2011 Protein-specific manipulation of ejaculate composition in response to female mating status in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **24**, 9922-9926.

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

40. Mohorianu I, Bretman A, Smith DT, Fowler E, Dalmay T, et al. 2017 Genomic responses to socio-sexual environment in male *Drosophila melanogaster* exposed to conspecific rivals. *RNA* **23**, 1048-1059.
41. Tibiche C, Wang E. 2008 MicroRNA regulatory patterns on the human metabolic network. Open Syst. Biol. J. **1**, 1.
42. Kozomara A, Griffiths-Jones S. 2014 miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucl. Acids Res*. **42**, D68-D73.
43. Mohammed J, Flynt AS, Panzarino AM, Mondal MMH, DeCruz M, et al. 2018 Deep experimental profiling of microRNA diversity, deployment, and evolution across the *Drosophila* genus. *Genome Res*. **28**, 1-14.
44. Chapman T, Bangham J, Vinti G, Seifried B, Lung O, et al. 2003 The sex peptide of *Drosophila melanogaster*: female post-mating responses analyzed by using RNA interference. *Proc. Natl. Acad. Sci. USA* **100**, 9923-9928.
45. R Development Core Team 2015 R: A Language and Environment for Statistical Computing. In. *R Foundation for Statistical Computing*. Vienna, Austria.
46. Berezikov E, Robine N, Samsonova A, Westholm JO, Naqvi A, et al. 2011 Deep annotation of *Drosophila melanogaster* microRNAs yields insights into their processing, modification, and emergence. Genome Res. **21**, 203-215.
47. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. 2003 Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. **13**, 2498-2504
48. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, et al. 2016 g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucl. Acids Re*s. **44**, W83–W89.
49. Baker CC, Gim BS, T. FM. 2015 Cell type-specific translational repression of Cyclin B during meiosis in males. *Dev*. **142**, 3394-3402.
50. Petruk S, Sedkov Y, Riley KM, Hodgson J, Schweisguth F, et al. 2006 Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. *Cell* **127**, 1209-1221.
51. Chertemps T, François A, Durand N, Rosell G, Dekker T, et al. 2012 A carboxylesterase, Esterase-6, modulates sensory physiological and behavioral response dynamics to pheromone in *Drosophila*. *BMC Biol*. **10**, 56.
52. Zare H, Khodursky A, Sartorelli V (2014) An evolutionarily biased distribution of miRNA sites toward regulatory genes with high promoter-driven intrinsic transcriptional noise. *BMC Evol. Biol*. **14**, 74.
53. Liu R, Woolner S, Johndrow JE, Metzger D, Flores A, et al. 2008 Sisyphus, the *Drosophila* myosin XV homolog, traffics within filopodia transporting key sensory and adhesion cargos. *Dev*. **135**. 53-63.
54. Liu C, Wanga J-L, Zheng Y, Xionga E-J, Li J-J, et al. 2014 Wolbachia-induced paternal defect in *Drosophila* is likely by interaction with the juvenile hormone pathway. *Insect Biochem. Mol. Biol*. **49**, 49-58.
55. Jattani R, Patel U, Kerman B, Myat MM. 2009 Deficiency screen identifies a novel role for beta 2 tubulin in salivary gland and myoblast migration in the *Drosophila* embryo. *Dev. Dynam*. **238**, 853-863.
56. Gonzalez EA, Garg A, Tang J, Nazario-Toole AE, Wu LP. 2013 A glutamate-dependent redox system in blood cells is integral for phagocytosis in *Drosophila melanogaster*. *Curr. Biol*. **23**, 2319–2324.
57. Shi X-Z, Zhong X, Yu X-Q. 2012 *Drosophila melanogast*er NPC2 proteins bind bacterial cell wall components and may function in immune signal pathways. *Insect Biochem. Mol. Biol*. **42,** 545-556.
58. Gligorov D, Sitnik JL, Maeda RK, Wolfner MF, Karch F. 2013 A novel function for the Hox gene Abd-B in the male accessory gland regulates the long-term female post-mating response in *Drosophila*. *PLoS Genet*. **9**. e1003395.

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

525 59. Fedorka KM, Winterhalter WE, Ware B. 2011 Perceived sperm competition intensity
526 influences seminal fluid protein production prior to courtship and mating. *Evolution*
527 **65**, 584-590.
528 60. Liu H, Kubli E. 2003 Sex peptide is the molecular basis of the sperm effect in *Drosophila*
529 *melanogaster*. *Proc. Natl. Acad. Sci. USA* **100**, 9929-9933.
530

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures

Figure 1. miRNA biogenesis. The miRNA biosynthesis pathway in *D. melanogaster* to indicate the drosha manipulation applied in the experimental manipulations.

Figure 2. Four examples of miRNAs as putative regulatory hubs. Secondary structures of four miRNAs with 100% complementarity between the 7-mer seed region (indicated by curly brackets) and the 3'UTRs of multiple transferome-encoding genes. (a) The 7-mer seed region of miR-4953-3p has matches to the 3'UTR of 17 genes (b) The seed-site of miR-7-3p matches 14 genes (c) The seed site of miR-4943-5p matches 42 genes (d) The seed-site of miR-315-5p matches 11 genes. The gene names are coloured according to broad functional categories: postmating behaviour/response (green); protein processing (blue); response to stimulus (purple); metabolic process (yellow); development/cellular organisation (red); gene expression (orange); unknown function (grey).

Figure 3. miRNA regulation of the post-mating response. Network of putative interactions between miRNAs (colourless nodes) and transferome genes (green nodes) whose products have a function in sperm storage and the post-mating response (PMR) of females. The size of the node is proportional to the number of edges.

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SI Tables

Table S1. (a) miRNAs whose seed sites are significantly over- or under-represented (Fisher exact test, $p < 0.05$) amongst the 3'UTRs of the seminal fluid protein transferome transcripts, when compared to all *D. melanogaster* 3'UTR sequences. The proportion of 3'UTRs with the seed site for a given miRNA is shown for the transferome transcripts (% transferome) and all *D. melanogaster* transcripts (% all). The direction of enrichment is indicated (1 = over-represented and -1 = under-represented in transferome 3'UTRs). **(b)** miRNAs with seed sites present in the 3'UTR of transferome genes. The total number of genes with seed sites for each miRNA, and the identities of those genes are shown in the second and third columns. **(c)** Transferome genes with seed sites for miRNAs within the 3'UTRs. The number and identities of miRNAs with seed sites in a given transferome 3'UTR are presented in the third and fourth columns.

Table S2. (a) Transcription factor binding motifs which are significantly over- or under-represented (Fisher exact test, $p < 0.05$) amongst the 5'UTRs of the seminal fluid protein transferome transcripts when compared to all *D. melanogaster* 5'UTR sequences. The proportion of 5'UTRs with a binding motif for a given transcription factor is shown for the transferome transcripts (% transferome) and all *D. melanogaster* transcripts (% all). The direction of enrichment is indicated (1 = over-represented and -1 = under-represented in transferome 5'UTRs). **(b)** Transcription factors with binding motifs present in the 5'UTR of transferome genes. The total number of genes with motifs for each transcription factor, and the identities of those genes are presented in columns 2 and 3. **(c)** Transferome genes with binding motifs for transcription factors within the 5'UTRs. The number and identities of transcription factors with motifs in a given transferome 5'UTR are presented in columns 2 and 3.

Table S3. (a) for each miRNA which could target one of the transferome transcripts, we counted the number of targeted transcripts (#transcripts) and presented their identity. **(b)** For

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

each transferome transcript, we counted the number of miRNA seed sites present on the 3' UTR (#miRNA column), able to induce translational repression {Bartel, 2009 #4134}; the identity of these miRNA seed sites is presented in the third column.

Table S4. (a) for each transcription factor (TF) which could target one of the transferome transcripts, we counted the number of targeted transcripts (#transcripts column) and presented their identity. The *D. melanogaster* TFs were downloaded from the REDfly database. **(b)** For each transferome transcript, we counted the number of TF motifs present on the 5' UTR (#tf column), able to induce transcriptional silencing {Bartel, 2009 #4134}; the identity of these TF motifs is presented in the third column.

Table S5. Gene ontology analysis of groups of transferome genes sharing the same putative miRNA seed sequence in the 3'UTR regions. Analyses were performed using the g:profiler software {Reimand, 2016 #4623}.

Table S6. Gene ontology analysis of groups of transferome genes sharing the same putative TF binding motif in the 5'UTR regions. Analyses were performed using the g:profiler software {Reimand, 2016 #4623}.

SI Figures

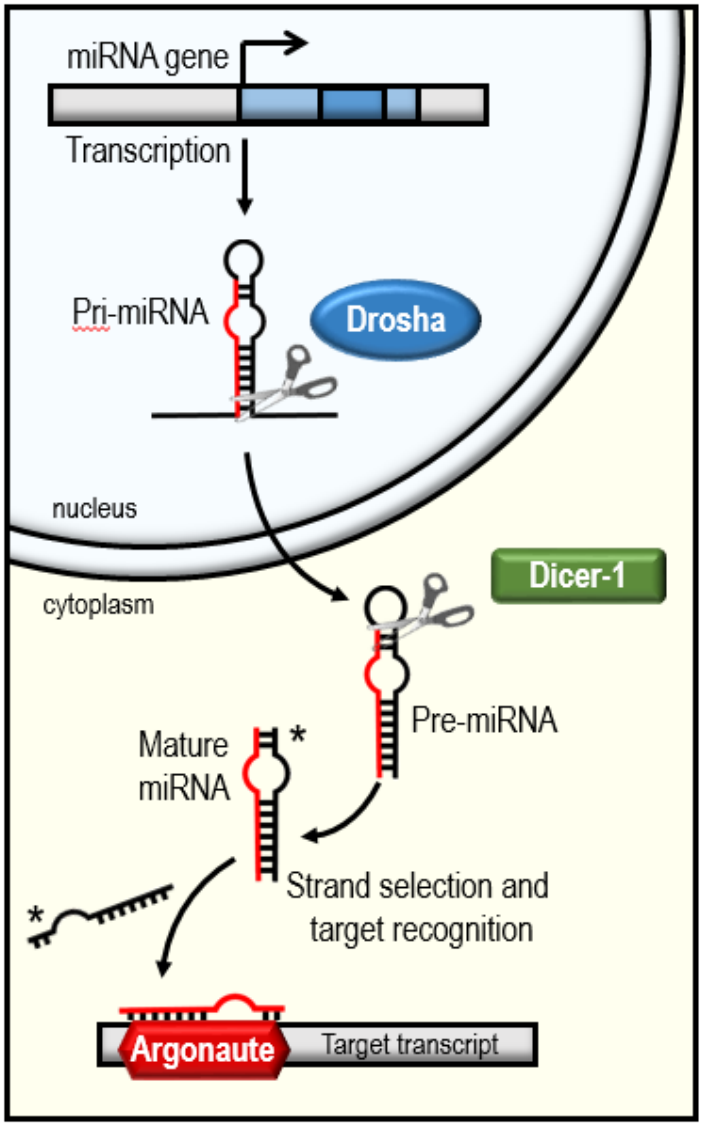
Figure S1. miRNA regulation of the transferome genes. Network of putative interactions between miRNAs (colourless nodes) and transferome genes whose products have a function in: development/cellular organisation (red), gene expression (orange), post-mating behaviour/response (green), response to stimulus (purple), metabolic process (yellow), protein processing (blue), unknown function (grey). Node size is directly proportional to the number of edges.

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

609 **Figure S2.** Transcription factor **(**TF) regulation of the transferome genes. Network of putative
610 interactions between TFs (colourless nodes) and transferome genes whose products have a
611 function in: development/cellular organisation (red), gene expression (orange), post-mating
612 behaviour/response (green), response to stimulus (purple), metabolic process (yellow), protein
613 processing (blue), unknown function (grey). Node size is directly proportional to the number
614 of edges.
615

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

616 **Figure 1.**



617

618

619

620

621

622

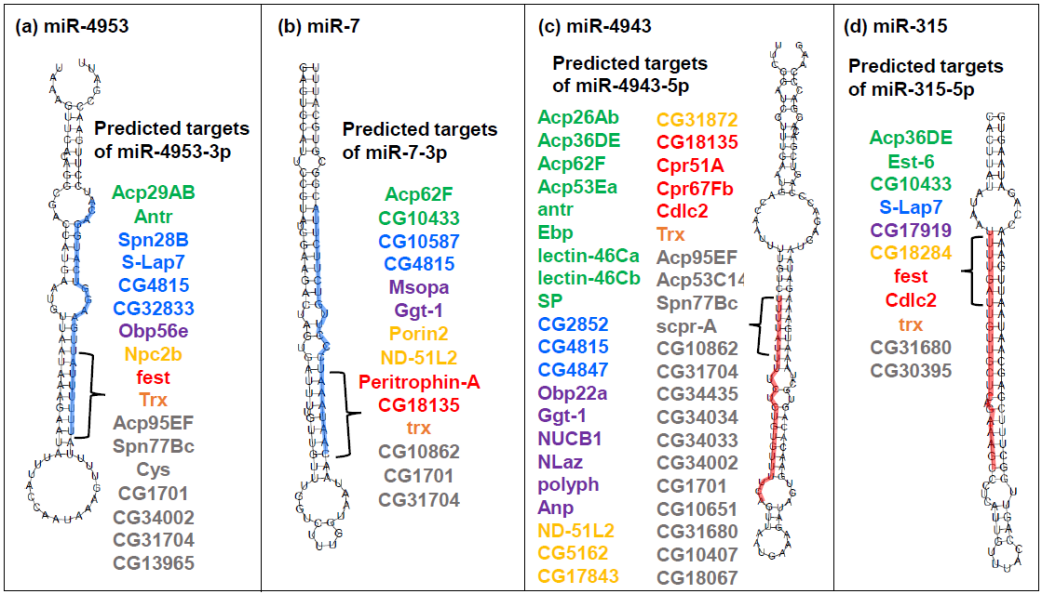
623

624

625

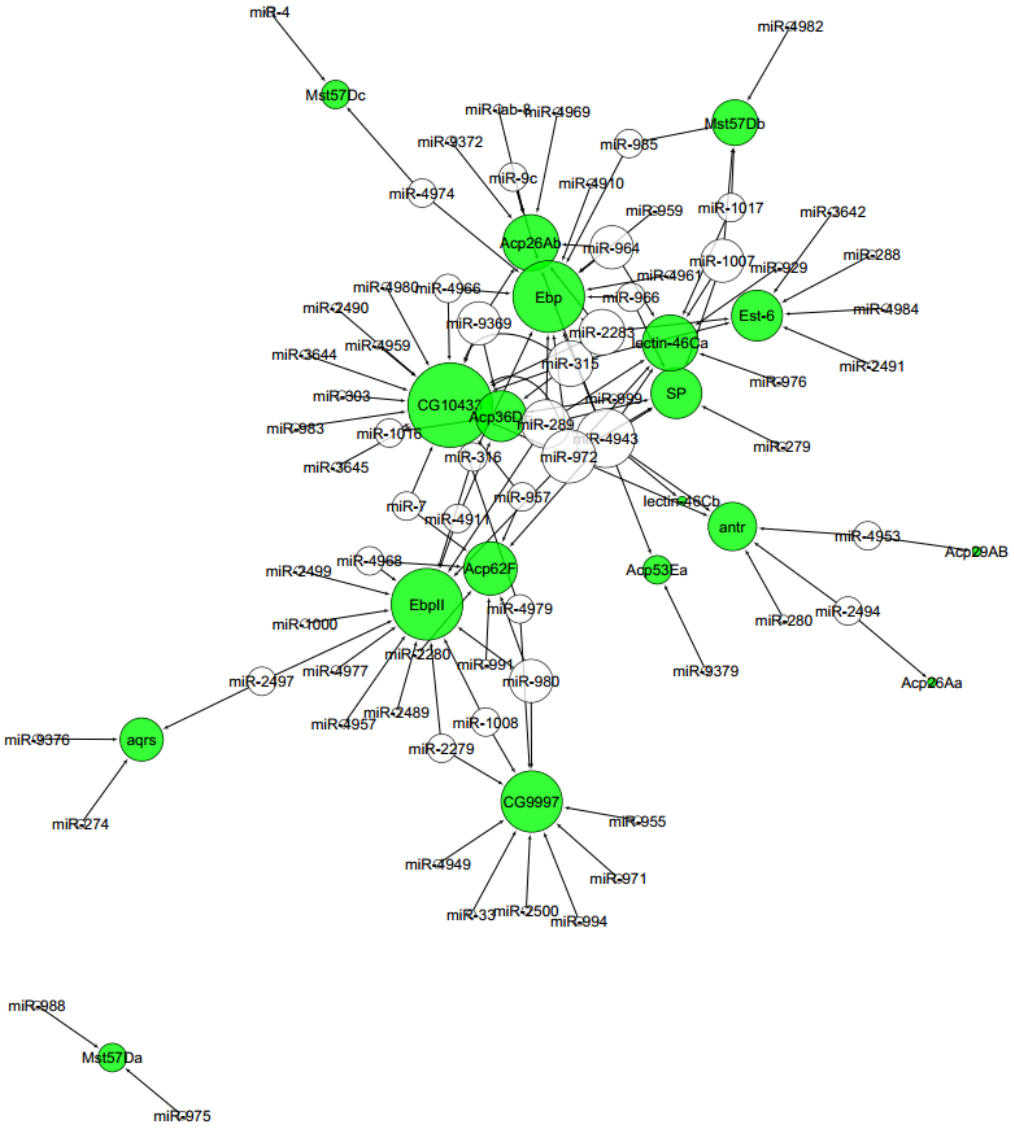
bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2.



bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

650 **Figure 3.**



651

bioRxiv preprint doi: https://doi.org/10.1101/379610; this version posted July 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

652