

# SCIENTIFIC REPORTS



OPEN

## MicroRNAs in Honey Bee Caste Determination

Regan Ashby<sup>1,2</sup>, Sylvain Forêt<sup>1</sup>, Iain Searle<sup>1,3,4,5</sup> & Ryszard Maleszka<sup>1</sup>

Received: 29 June 2015

Accepted: 12 November 2015

Published: 07 January 2016

The cellular mechanisms employed by some organisms to produce contrasting morphological and reproductive phenotypes from the same genome remains one of the key unresolved issues in biology. Honeybees (*Apis mellifera*) use differential feeding and a haplodiploid sex determination system to generate three distinct organismal outcomes from the same genome. Here we investigate the honeybee female and male caste-specific microRNA and transcriptomic molecular signatures during a critical time of larval development. Both previously undetected and novel miRNAs have been discovered, expanding the inventory of these genomic regulators in invertebrates. We show significant differences in the microRNA and transcriptional profiles of diploid females relative to haploid drone males as well as between reproductively distinct females (queens and workers). Queens and drones show gene enrichment in physio-metabolic pathways, whereas workers show enrichment in processes associated with neuronal development, cell signalling and caste biased structural differences. Interestingly, predicted miRNA targets are primarily associated with non-physio-metabolic genes, especially neuronal targets, suggesting a mechanistic disjunction from DNA methylation that regulates physio-metabolic processes. Accordingly, miRNA targets are under-represented in methylated genes. Our data show how a common set of genetic elements are differentially harnessed by an organism, which may provide the remarkable level of developmental flexibility required.

In honeybees, three contrasting adult phenotypes are produced from the same genome as a result of biological constraints that are manifested during development<sup>1–4</sup>. Two identical diploid embryos can develop either into a functionally sterile and short-lived female worker or into a highly reproductive and long-lived queen, depending on dietary intake during postembryonic development. In contrast, male drones develop from unfertilised haploid eggs laid by a queen into special cells. Although sex determination is genetically controlled<sup>5</sup>, drone larvae receive a distinct diet in larger quantities compared to that of worker larvae. This suggests that similarly to queens, nutrition provides important cues for their proper development<sup>4,6</sup>.

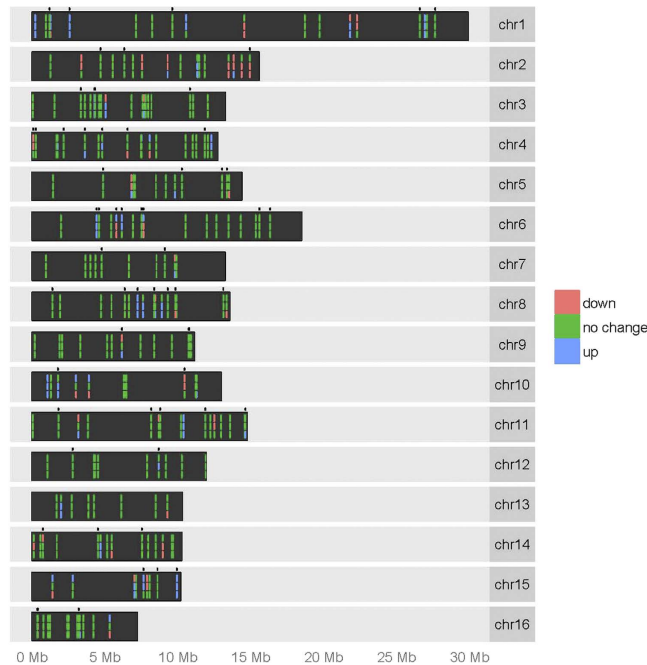
From a developmental perspective the phenomenon of conditional phenotypes is not trivial. It requires hierarchical global reprogramming of gene expression patterns related to reproduction, longevity, cell differentiation, organogenesis, apoptosis, contrasting behaviours, etc. The flexibility of the underlying processes is underscored by recent findings that the queen phenotype can be induced not only by feeding selected larvae with royal jelly, but also by other means<sup>7</sup>, notably by silencing *de novo* methylation in newly hatched larvae<sup>8</sup> and by interfering with the expression of genes belonging to the insulin pathway<sup>9</sup>.

To gain a greater understanding of the regulatory processes underlying caste specific phenotypic development in response to environmental cues, recent studies have investigated differences in the transcriptional landscape of the two female diploids to try and elucidate possible caste-biased gene expression profiles<sup>10–17</sup>. In general, larvae destined to be queens show enrichment in gene transcription associated with physio-metabolic pathways and ovary development, most likely associated with greater mass and reproductive capabilities of the larger fertile queen. In contrast, larvae destined to become the functionally smaller sterile workers show enrichment in genes associated with developmental maturation and caste biased structures such as enlarged mushroom bodies, suppressed ovary development, as well as structural differences in the legs and wings required for foraging. The transcriptional profile of both female diploids varies somewhat from the above generalisations at earlier stages of larval development (L1–L3)<sup>13</sup>, which precede the differential expression of juvenile hormone, a process believed to be critical to the

<sup>1</sup>Research School of Biology, Australian National University, ACT, 0200, Australia. <sup>2</sup>Centre for Research in Therapeutic Solutions, Faculty of Education, Science, Technology and Mathematics, University of Canberra, ACT, 2617, Australia.

<sup>3</sup>School of Biological Sciences, University of Adelaide, SA, 5005, Australia. <sup>4</sup>School of Agriculture, Food and Wine, University of Adelaide, SA, 5005, Australia. <sup>5</sup>The University of Adelaide and Shanghai Jiao Tong University Joint

International Laboratory for Agriculture and Health, University of Adelaide, SA, 5005, Australia. Correspondence and requests for materials should be addressed to R.A. (email: regan.ashby@canberra.edu.au) or S.F. (email: sylvain.fores@anu.edu.au)



**Figure 1. Location of miRNAs in the honeybee genome and their differential expression.** For each chromosome, the top line represents miRNA expression levels in queens vs workers, middle line represents drones vs workers, and the bottom line represents drones vs queens. Green shading indicates no statistically significant difference in miRNA expression between the two castes compared, red represents a down-regulation and blue represents an up-regulation. Dots located above each chromosome indicate the location of novel miRNAs describe for the first time in this study.

general body growth and ovary development of queens<sup>11,18</sup>. In contrast to the considerable analysis of differences in the transcriptional profile of diploid females, virtually no work has been undertaken on the transcriptional profile of the developing haploid male, a critical question addressed in the current article.

How such distinctly different transcriptional profiles are established during larval development, and how such profiles lead to caste specific phenotypic outputs, is not well understood, but appears to involve discrete changes in hormone signalling associated with dietary intake<sup>8</sup>. At the molecular level, the contrasting genome-wide methylation profiles in queens and workers<sup>17</sup> are essential for controlling the caste-specific gene expression patterns, and in particular the patterns of alternative splicing in response to dietary cues. The role of other epigenomic modifiers in bee development is less clear, but some unusual fatty acid components of royal jelly have already been implicated in such a role<sup>4,19</sup>. They exhibit powerful histone deacetylase inhibitory activities, which results in open chromatin conformation affecting the expression of hundreds of genes from the very moment every young larva gets a mouthful of royal jelly. This idea is supported by a recent study showing that extensive post-translational histone tail modifications occur in queen ovaries and in larval tissues<sup>20</sup>.

In more recent times, the role of non-coding microRNAs (miRNAs) has become an area of major interest in regards to the regulation of phenotypic and developmental plasticity. As part of a larger RNA induced silencing complex (RISC), miRNAs provide sequence-specific translational repression of target mRNA molecules. They may also be involved in RNA-directed DNA methylation, translational activation and alternative splicing<sup>21–24</sup>. Increasing evidence suggests that miRNAs confer genetic robustness by reinforcing transcriptional programs, preventing major phenotypic variation in response to environmental fluctuations [canalization, for review see<sup>25</sup>].

Work in several species has indicated that miRNAs are important in safeguarding the correct spatiotemporal transcriptional profile, ensuring proper tissue differentiation during the movement between developmental stages, for review see<sup>26</sup>. In honeybees, preliminary studies on the spatiotemporal patterning of miRNA expression in adults suggested that the miRNA landscape changes during the age-dependent switch in labour division from nurse to forager in female workers<sup>27–33</sup>. However, little work has been undertaken on the role of miRNAs in the establishment of caste specific gene-expression profiles associated with the phenotypic polymorphism observed between drones, queens and workers. This study investigates differences in the miRNA profiles observed between these three castes during a critical time window of developmental canalization, or the point of no return, and how this relates to differences in their transcriptional profile and ultimately their phenotypic output.

## Results/Discussion

**MiRNA profile varies significantly between castes.** In addition to the 218 known honeybee miRNAs deposited in miRBase<sup>34</sup> (<http://www.mirbase.org/>), we identified 82 novel miRNAs in our samples (Supplementary Table S1). Novel miRNAs were dispersed throughout the genome (Fig. 1), sitting primarily within intergenic and intronic regions (Table 1). The majority of honeybee miRNAs are located in intergenic

miRNA category	Intergenic	Intronic	UTR	CDS
Known	105	85 (54:31)	8 (4:4)	20 (13:7)
Novel	42	32 (28:4)	3 (3:0)	5 (5:0)
Total	147	117 (82:35)	11 (7:4)	25 (18:7)

**Table 1. Genomic location of novel miRNAs.** We identified 218 known miRNAs and 82 novel miRNAs. These miRNAs were predominately located in intergenic regions however a significant number were located in introns. For intronic, untranslated region (UTR) and coding regions (CDS), the first number in parenthesis indicates the number of miRNA on the same strand as the corresponding gene, and the second number indicates the number of the reverse strand. The known miRNAs are those already deposited in miRBase.

regions, but a substantial number are also found within protein coding genes, primarily in introns (Table 1). MiRNAs overlapping with protein coding genes tend to be found on the same strand suggesting that these miRNAs could be derived from mRNAs e.g. mirtrons<sup>35,36</sup>. A map of the chromosomal location of the identified miRNAs is shown on Fig. 1. MicroRNAs, both known and novel, are dispersed throughout the genome, with only a small proportion clustered in regions of 1Kb or less (Supplementary Figure S1)

After filtering out miRNAs with low counts (see methods), the expression of 164 distinct miRNAs were detected. Of these, 119 represent known or predicted sequences that have been previously deposited in miRBase, while 45 represent previously unknown sequences. The majority of the unknown miRNAs share little sequence similarity with known miRNAs from other species and therefore appear to represent miRNAs unique to *Apis mellifera*.

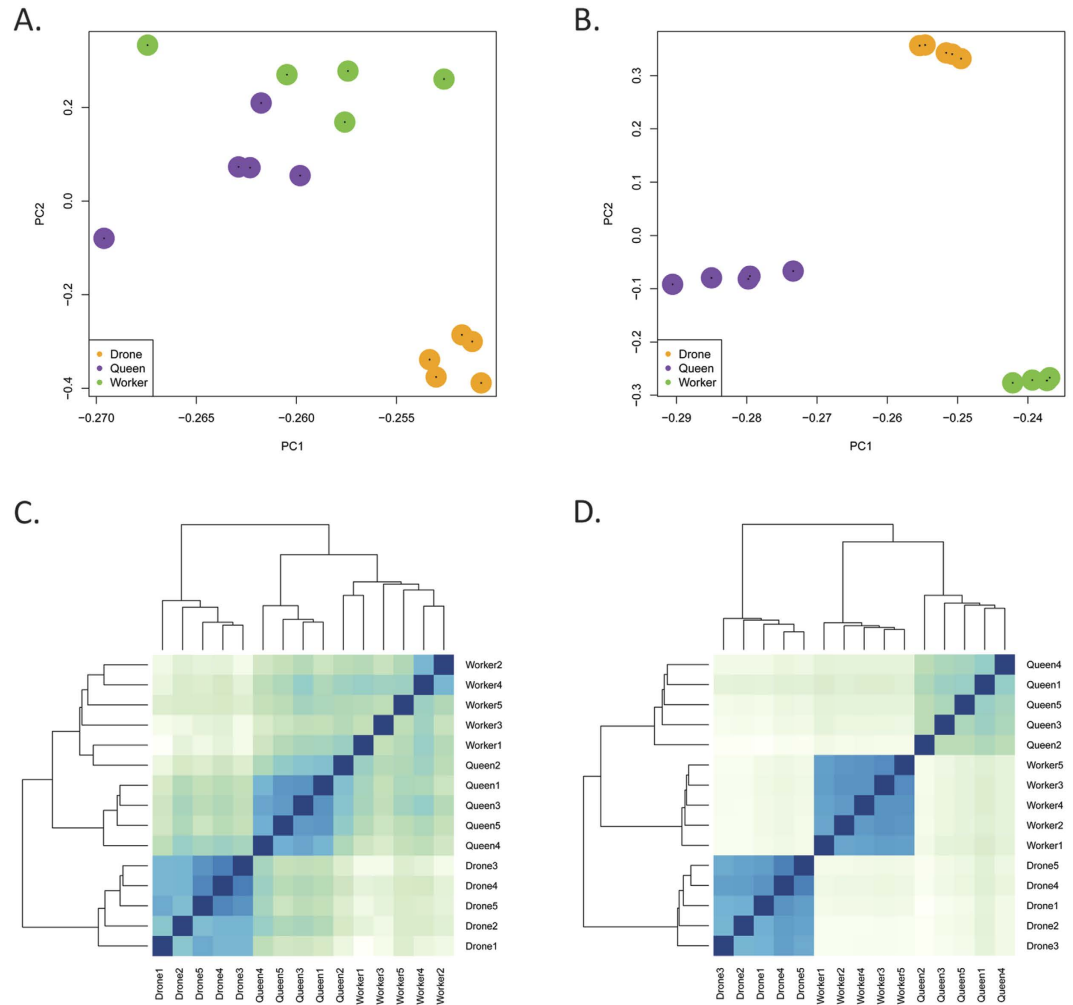
At larval stage L4/L5, the miRNA profile is clearly distinct for each of the three honeybee castes. Principle component analysis (Fig. 2A) and clustering analysis according to the Euclidian distance between gene expression vectors (Fig. 2C) illustrates that each caste groups in a distinctly separate manner. Of the 164 miRNAs detected, 120 show differential expression between at least two castes, with 27 showing differential expression amongst all three castes (Fig. 3A). Between diploid castes, 72 miRNAs show differential expression. Of these, 42 are up-regulated and 30 are down-regulated in workers relative to queen larvae. A comparison of queens relative to the haploid drone illustrates the differential expression of 87 miRNA (49 up-regulated and 38 down-regulated). Finally, the differential expression of 80 miRNAs (44 up-regulated, 36 down-regulated) is observed in workers relative to drones. Significant overlap is seen in the expression differences observed between drones relative to either of the female castes. This suggests that the haploid system, at this developmental stage, shows clear differences in its non-coding RNA profile to that of the diploid system.

**High success rates of PCR validation confirms differential expression of miRNAs.** The expression profile of 17 candidate miRNAs was further validated by StemLoop PCR (Table 2). Candidate miRNAs were chosen based on having; 1) a moderate to high expression level, 2) a differential expression with a false discovery rate of less than 5%, and 3) a fold change of more than 1.5 in at least one of the three comparisons between castes. A between-caste comparison of the sequencing data indicated 36 differential changes in the expression of the 17 candidate miRNAs chosen. In the majority of cases, differential expression represented a reduction in miRNA levels in one or both of the diploid females relative to drones (this pattern represents 20/36 changes observed). Of these changes, 31 were validated by StemLoop PCR (Table 2), representing an 86% validation rate. Of the 15 conditions in which the sequencing data indicated no differential expression should be observed, 13 were validated by PCR (Table 2), again resulting in a high validation rate of 87%.

**Transcriptional profile of the honeybee at larval stage L4/L5.** At this larval stage (L4/L5), when the commitment to a specific developmental trajectory is virtually irreversible, the transcriptional profile of the three honeybee castes is clearly distinct. Principal component analysis (Fig. 2B) and clustering according to the Euclidian distance between gene expression vectors (Fig. 2D) show each caste grouping separately.

**Gene Ontology (GO) and pathway analysis substantiate phenotypic distinctiveness of female castes.** Distinct differences in the gene enrichment profile are observed between all three castes at 96hrs of larval development, as determined by both Gene Ontology (GO) and pathway analysis (using the Kyoto Encyclopaedia of Genes and Genomes, KEGG). Queen larvae show gene enrichment in pathways associated with physiometabolic processes (Supplementary Table S2, sections 1.1 and 1.2), especially in the catabolism of the three major classes of macromolecules (proteins, carbohydrates and fats) as well as the biosynthesis of juvenile hormone (JH, Table 3), consistent with the significant increase in JH levels in queens at this developmental stage<sup>11</sup>. This is indicative of the increased growth rates seen in queen larvae during this period. In our bee colonies, 96 hr-old queen larvae have a body weight ranging from 200–310 mg, whereas the weight of worker larvae is typically 170–200 mg. For comparison, drone larvae of similar age can reach body mass of up to 400 mg. The largest changes are apparent in pathways associated with protein metabolism followed by a significant up-regulation in genes associated with the major energy pathways, including; the citric acid cycle (Table 4), pentose phosphate pathway, glyoxylate metabolism, glycolysis/gluconeogenesis, oxidative phosphorylation and pyruvate metabolism. Pathways associated with carbon fixation and metabolism are also significantly up-regulated, as well as genes associated with the metabolic pathways of lysosome activity, folate biosynthesis, the metabolism of propanoate and xenobiotics, as well as mitochondrial development.

In contrast to queens, workers show a profile indicative of an increased rate of developmental maturation and caste biased structural differences (Supplementary Table S2, sections 1.1 and 1.2). Strong enrichment is seen in

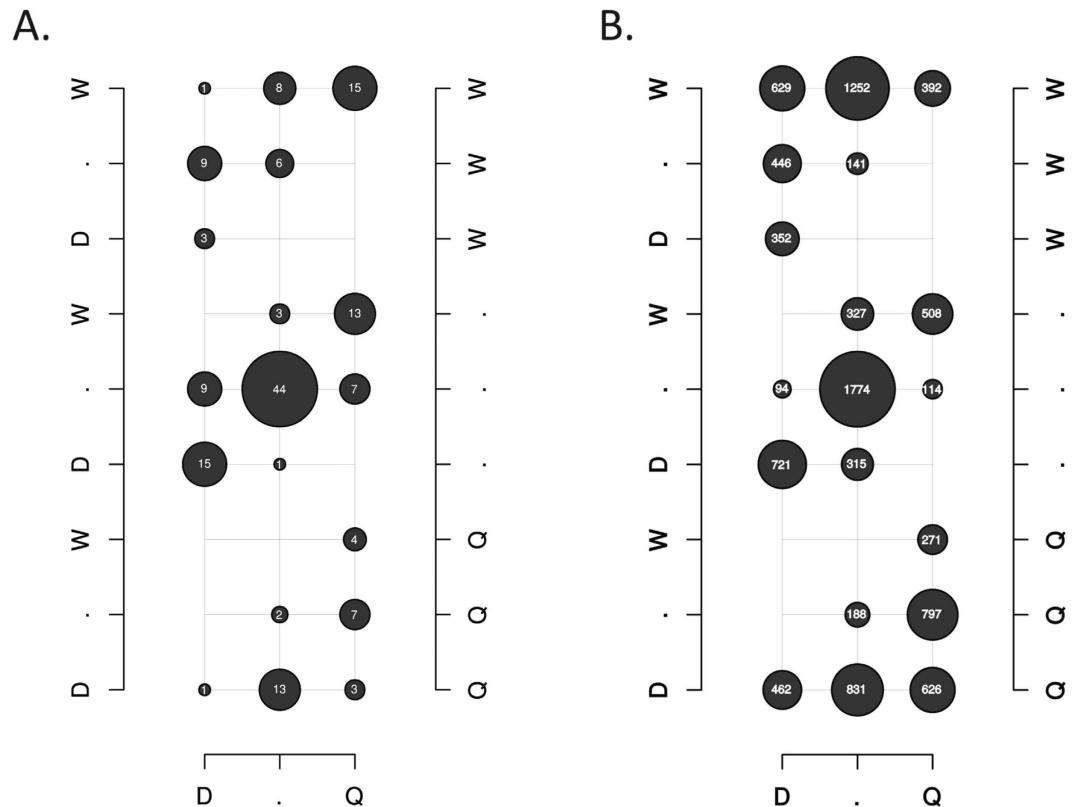


**Figure 2. Principal component analysis of the miRNA (A) and mRNA (B) samples.** Clustering using the Euclidian distance between the miRNA (C) and mRNA (D) expression vectors. The logarithm of the expression values was used.

genes associated with the major cell signalling pathways involved in embryonic development, including hippo, Wnt, MAPK/ERK, hedgehog, PI3K/AKT, Notch, VEGF and calcium signalling. In diploids, a number of these pathways also show differential methylation, especially PI3K/AKT<sup>17</sup>. Workers also show enrichment in genes associated with cell guidance (chemokine signalling), adhesion and microfilament formation (e.g. actin cytoskeleton), as well as pathways associated with transcription, translation and protein processing. During the fourth and fifth stage of larval development, workers show a reduction in ovariole numbers, through programmed cell death. Consistent with this observation, the associated apoptotic pathways show gene enrichment (e.g.  $\beta$ -catenin), similar to that reported previously<sup>11,13,16,37</sup>.

Workers also show significant enrichment in neuronal pathways, including: cholinergic (Table 5), glutamatergic and dopaminergic synaptic development, axon guidance, retinal phototransduction, ommatidial rotation, olfactory learning, synaptic vesicle formation, peripheral nervous system development, potassium and calcium ion transmembrane transport and voltage-gated potassium channel complex formation. This distinctiveness in neuronal development/maturation in the worker larvae may be associated with the early stages of building a sophisticated nervous system required for their remarkable navigational skills and high mnemonic fidelity during adult life. Workers also show caste specific structural differences in the morphogenesis of both their legs and wings to allow them to undertake foraging tasks (e.g. unique structures for pollen and propolis collection). This is underscored by enrichment in genes associated with imaginal disc-derived leg morphogenesis, wing disc dorsal/ventral pattern formation and morphogenesis, and positive regulation of synaptic growth at neuromuscular junctions during this critical period of caste determination. Relative to queens, workers also show gene enrichment in pathways associated with; digestive activities (gastric acid secretion), hormone biogenesis (insulin and aldosterone), melanogenesis (believed to be critical for innate immunity)<sup>38</sup>, and circadian entrainment.

**GO and pathway analysis reveal contrasting regulation in haploid males versus diploid females.** Relative to workers, drones show a gene enrichment profile similar to queens, with an up-regulation in genes associated with physio-metabolic processes (Supplementary Table S2, sections 1.1 and 1.2). Relative to



**Figure 3. Summary of the number of differentially expressed miRNAs (A) and mRNA transcripts (B) between castes.** The bottom, left and right axes represent the differences between drones and queens, drones and workers, queens and workers respectively. The letters D, Q, W indicate higher expression in drones, queens and workers respectively, while “.” denotes no differential expression. For example, when comparing the expression of miRNAs between workers and queens in (A), there are 42 miRNAs up-regulated in workers, 30 up-regulated in queens, and 92 miRNAs that show no change between these two castes when analysing all 164 miRNAs that were detected. For workers, the 42 miRNAs that show increased expression are displayed across the three “W” rows from right to left (15 + 8 + 1 + 6 + 9 + 3 = 42). For queens, the 30 miRNAs that show increased expression are displayed across the three “Q” rows from right to left (4 + 7 + 2 + 3 + 13 + 1 = 30), while those miRNA that show no change in expression between these two castes are displayed across the three “.” rows from right to left (13 + 3 + 7 + 44 + 9 + 1 + 15 = 92). Individual numbers represents the expression profile of that grouping of miRNAs when looking across all three comparisons (drones and queens, drones and workers, queens and workers), e.g. the top right most grouping of 15 miRNAs was up-regulated in queens relative to drones, in workers relative to drones, and in workers relative to queens.

drones, workers again show enrichment in genes associated with developmental maturation, cell signalling, neuronal maturation and caste biased structural differences.

Interestingly, drones show an even greater level of gene enrichment in a number of physio-metabolic components relative to queens, indicating a possible higher rate of metabolic activity (Supplementary Table S2, section 1.1 and 1.2). Drones also show increased expression of genes associated with the metabolism of proteins, carbohydrates, carbon and methane, as well as structural constituents of ribosomes and cuticles. In contrast, queens show gene enrichment in pathways associated primarily with oxidative phosphorylation and protein processing and exportation relative to drones.

**Predicted miRNA targets are part of distinct gene pathways.** MicroRNA target prediction was initially undertaken using a minimum of two out of three *in-silico* prediction methods (see methods). From this initial analysis, the ontology terms (biological processes) that showed the greatest enrichment for predicted targets included; transcription factor activity, steroid hormone pathway, neuronal development (brain morphogenesis, olfactory learning and axon guidance/formation), ion channel formation (specifically potassium voltage-gated and glutamate-gated), and imaginal disc-derived wing morphogenesis (Supplementary Table S2, section 2.1).

This list was further refined by accepting only miRNA target pairs with a statistically significant Pearson correlation coefficient (Fig. 4A). Based on this analysis, 654 interactions were predicted to occur, involving 96 miRNAs and 526 target transcripts (Supplementary Table S3), 87% of which are differentially expressed, supporting the validity of the predicted targets. Of the predicted interactions, 449 involve a positive correlation, in which both the miRNA and the targeted transcript show a similar direction of change, while 205 show a negative correlation (Fig. 4B). A number of miRNAs are predicted to target multiple genes, the maximum being ame-mir-210 with

miRNA	High Throughput Sequencing - Fold Change			qRT-PCR - Fold Change		
	Queen vs. Drone	Worker vs. Drone	Worker vs. Queen	Queen vs. Drone	Worker vs. Drone	Worker vs. Queen
ame-bantam	-1.6	-1.8	-1.2	-1.3	-1.4	-1.1
ame-mir-1175	-2.5	-3.6	-2.1	-1.8	-2.0	-1.2
ame-mir-13b	-1.7	1.6	2.4	-2.1	1.4	2.8
ame-mir-2	-1.9	1.2	2.1	-1.6	-7.3	4.4
ame-mir-252a	-2.1	-1.4	1.6	-2.2	-2.2	1.0
ame-mir-276	1.8	1.0	-1.8	1.7	-2.1	-3.6
ame-mir-2765-5p	3.7	3.5	-1.2	11.5	8.6	-1.3
ame-mir-283	-1.9	-2.0	-1.2	-1.5	-2.6	-2.0
ame-mir-315	1.7	-2.1	-2.9	-1.2	-2.3	-2.0
ame-mir-3715	-3.7	-3.9	-1.2	-3.0	-3.0	-1.0
ame-mir-375	2.2	1.5	-1.6	2.2	1.2	-1.9
ame-mir-6001-3p	3.2	3.5	1.3	2.5	6.0	2.4
ame-mir-71	-2.0	-1.8	1.2	-3.8	-3.0	1.3
ame-mir-750	1.0	-2.5	-2.5	-1.5	-3.6	-2.4
ame-mir-87-1	-1.9	-1.6	1.4	-2.8	-3.5	-1.3
ame-mir-87-2	-2.0	-1.7	1.2	-2.9	-2.2	1.4
ame-mir-let7	1.4	-2.2	-2.7	3.2	-1.5	-3.5

**Table 2. Candidate validation of miRNA expression levels from drone, queen and worker larvae by StemLoop PCR.** Comparison of changes in miRNA expression levels obtained from Illumina high throughput sequencing data and StemLoop PCR. Expression changes are reported as fold changes, with a fold change of 1 equalling no change in miRNA expression, while a fold change of 2 equals a doubling in product. Minus values represent a reduction in miRNA expression. A significant increase in miRNA expression is denoted by green shading ( $p < 0.01$ ), while a significant decrease in miRNA expressed is denoted by pink shading ( $p < 0.01$ ).

Enzyme (JH Biosynthesis)	GB Number	Log Fold Change*	FDR	Function
Adenosylhomocysteinase	GB14324	-0.55	$1.86 \times 10^{-16}$	Hydrolysis of S-adenosylhomocysteine
Adenosine kinase	GB13063	-1.12	$1.53 \times 10^{-39}$	Phosphorylates adenosine
Acetyl CoA thiolase	GB18952	-0.54	$1.13 \times 10^{-19}$	Converts acetyl CoA to acetoacetyl CoA
	GB12956	-3.58	$1.00 \times 10^{-40}$	
Hydroxymethylglutaryl-CoA synthase	GB12497	-0.15	0.35	Condenses Acetyl-CoA with acetoacetyl-CoA to HMG-CoA
Hydroxymethylglutaryl-CoA reductase	GB18242	-0.31	$9.50 \times 10^{-5}$	Reduces HMG-CoA to mevalonate
Mevalonate kinase	GB14701	-0.67	$1.00 \times 10^{-7}$	Phosphorylates mevalonate
Phosphomevalonate kinase-like protein	GB11982	-0.27	0.14	Phosphorylates phosphomevalonate
Diphosphomevalonate decarboxylase	GB16296	-0.83	$1.19 \times 10^{-16}$	Converts phosphomevalonate to isopentenyl diphosphate (IPP)
Isopentenyl diphosphate delta isomerase	GB44492	-0.55	$1.57 \times 10^{-14}$	Converts IPP to dimethylallyl pyrophosphate (DMAPP)
Farnesyl pyrophosphate synthase	GB18623	1.40	$4.14 \times 10^{-61}$	Formation of farnesyl diphosphate from condensation of IPP and DMAPP
	GB12385	-0.60	$5.69 \times 10^{-12}$	
	GB15337	1.22	$3.15 \times 10^{-75}$	
	GB18623	-1.76	$2.28 \times 10^{-11}$	
Citrate synthase	GB12573	-0.18	$2.00 \times 10^{-3}$	Citrate synthesis (mitochondria)
ATP citrate lyase	GB10992	0.19	$5.28 \times 10^{-5}$	Synthesis of acetyl CoA from citrate
Short-chain dehydrogenase/reductase	GB12522	1.87	$7.20 \times 10^{-4}$	Oxidation of farnesol to farnesal
	GB14607	-1.20	$5.48 \times 10^{-217}$	
Juvenile hormone acid methyltransferase	GB10517	-5.76	$1.00 \times 10^{-40}$	Transfer of a methyl group to farnesoic acid
Methyl farnesoate epoxidase (CYP15A1)	GB15634	-3.52	$1.00 \times 10^{-41}$	Oxidation of methyl farnesoate to juvenile hormone III

**Table 3. Increased expression of enzymes associated with juvenile hormone (JH) biosynthesis in queens relative to workers.** Note, no genes within the JH pathway were predicted to be a miRNA target. \*Log Fold Change represents the expression level in workers relative to queens, as determined by high throughput sequencing. A negative value indicates a down-regulation in gene expression in workers relative to queens, a positive value indicates the opposite. FDR = False discovery rate adjusted P-value.

68 interactions. Similarly, many transcripts are targeted by more than one miRNA, with the transcription factor mblk-1 (GB50048) showing the largest number of predicted interactions (seven). Mblk-1 is a transcription factor expressed within neuronal tissue, specifically the mushroom bodies, and appears to be modulated by the Ras/MAPK pathway<sup>39</sup>. Both neuronal tissue and cell signalling pathways such as Ras/MAPK appear to be primary targets of the miRNA system, as discussed below.

Enzyme (Citric Acid Cycle)	GB Number	Log Fold Change*	FDR	Predicted Targeting miRNAs	Function
Citrate synthase 1, mitochondrial-like	GB52073	-0.18	$2.00 \times 10^{-3}$		Catalyses the formation of citrate from acetyl-CoA and oxaloacetate
ATP citrate lyase	GB54216	0.19	$5.28 \times 10^{-5}$		Catalytic conversion of citrate and CoA to acetyl-CoA and oxaloacetate
Aconitate hydratase	GB43618	0.93	$6.29 \times 10^{-57}$	miR-316, miR-6063	Catalyses the isomerisation of citrate to isocitrate via cis-aconitate
Isocitrate dehydrogenase	GB45258	-1.23	$2.74 \times 10^{-168}$		Isomerisation of isocitrate to 2-oxoglutarate
	GB48527	-0.81	$8.47 \times 10^{-26}$		
2-oxoglutarate dehydrogenase, mitochondrial-like	GB50958	0.38	$7.65 \times 10^{-7}$	miR-279b	Enzyme complex that catalyses the conversion of 2-oxoglutarate to succinyl-CoA
Dihydrolipoamide dehydrogenase 1	GB51335	-0.50	$7.92 \times 10^{-11}$		
Dihydrolipoalysine-residue succinyltransferase	GB44430	0.66	$2.15 \times 10^{-15}$		
Succinyl coenzyme A synthetase alpha subunit	GB49632	-0.94	$2.48 \times 10^{-126}$		Catalyses the reversible reaction of succinyl-CoA to succinate
Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial-like	GB47042	0.47	0.10	miR-6056	Oxidation of succinate to fumarate, with the transfer of electrons from succinate to ubiquinone (coenzyme Q)
Fumarate hydratase, mitochondrial-like	GB51042	-0.62	$2.60 \times 10^{-24}$	miR-6059, miR-3741	Catalyses the conversion of fumarate to malate
Malate dehydrogenase, mitochondrial-like	GB54720	-1.21	$2.00 \times 10^{-2}$	miR-12	Catalyses the conversion of malate to oxaloacetate
	GB42526	-1.69	$1.00 \times 10^{-60}$		
Phosphoenolpyruvate carboxykinase	GB51494	-4.36	$1.00 \times 10^{-60}$		Catalyses the conversion of oxaloacetate (OAA) to phosphoenolpyruvate (PEP)
Dihydrolipoamide S-acetyltransferase	GB53566	-0.12	0.54	miR-3741, miR-6059, miR-3792	Catalyses the overall conversion (acetyl transfer) of pyruvate to acetyl-CoA
Pyruvate dehydrogenase (lipoamide) beta	GB55496	-0.26	$2.00 \times 10^{-2}$		
Pyruvate carboxylase, mitochondrial-like	GB40280	0.57	$1.92 \times 10^{-43}$	miR-315	Catalyses the carboxylation of pyruvate to oxaloacetate (OAA)
Dihydrolipoamide dehydrogenase 1	GB51335	-0.50	$7.92 \times 10^{-11}$	miR-279b	Oxidises dihydrolipoamide to lipoamide

**Table 4. An example of increased expression of enzymes associated with metabolic activity (citric acid cycle) in queens relative to workers.** \*Log Fold Change represents the expression level in workers relative to queens, as determined by high throughput sequencing. A negative value indicates a down-regulation in gene expression in workers relative to queens, a positive value indicates the opposite. FDR = False discovery rate adjusted P-value.

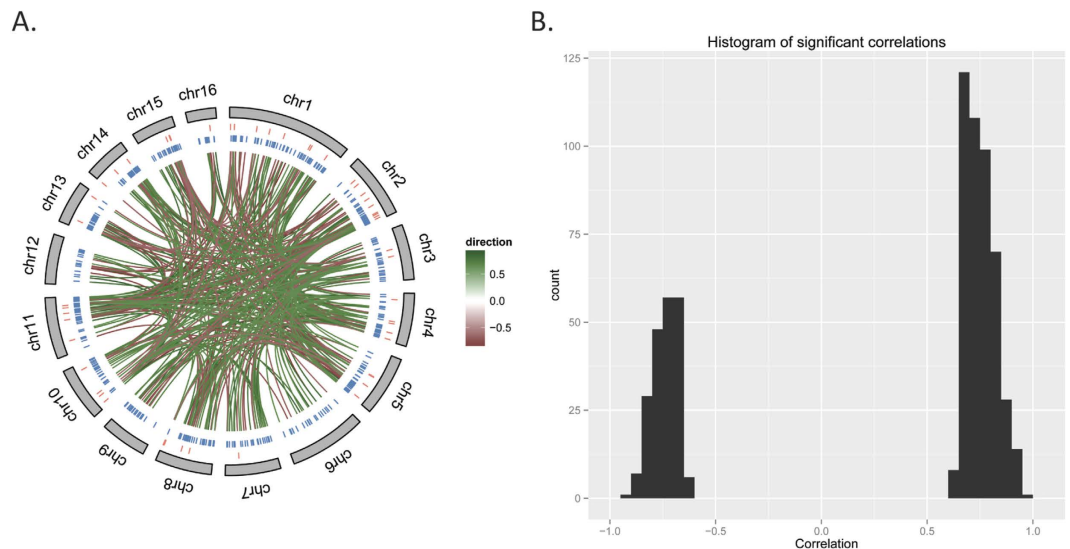
Of the 526 predicted transcripts that show a correlation with their targeting miRNA, 371 show a positive correlation, 190 show a negative correlation, while 35 show both. A similar breakdown was seen for the 96 miRNAs involved, with 73 showing a positive correlation to their targets, 71 showing a negative correlation and 48 showing both. Interestingly, it is often assumed that a negative correlation should predominate. That is, a detectable reduction in transcript numbers would be observed following the up-regulation of the targeting miRNA, if cleavage of the target transcript occurred. However, miRNAs, for the most part, only moderately repress their targets, and appear to confer genetic robustness by preventing the overexpression of target genes during fluctuations in environmental cues, for review see<sup>25</sup>. Under this framework, a positive correlation would not be unexpected, as reported recently in other biological contexts<sup>40,41</sup>. For example, in response to an environmental cue such the ingestion of royal jelly, the transcriptional output of specific developmental genes may increase. However, overexpression of such developmental pathways would be deleterious to the organism and may therefore be prevented by increased miRNA expression, buffering translational output. In this scenario, although both the target transcript and miRNA would be up-regulated, the miRNA is still acting to reduce (buffer) the rate/amount of translation of the targeted gene relative to that which would occur in the absence of the miRNA. This senior could also occur when the targeting miRNA interferes with the translational machinery rather than causing target-specific mRNA cleavage (for a review of possible translational suppression methods, see<sup>42</sup>). In this case, transcription levels could rise or remain unchanged, yet translation levels may be suppressed.

Analysis of GO terms under the categories of biological processes and molecular function indicates that miRNAs in the honeybee are predicted to overwhelmingly target transcription factors within pathways associated with neuronal development (Supplementary Table S2, sections 2.1 and 3.1; Supplementary Table S3). Specifically, gene enrichment was seen in pathways associated with; axon guidance and development, synapse formation (target recognition, vesicle formation, synaptic growth, formation of neuromuscular junctions, dopaminergic and cholinergic synapse formation), telencephalon development, visual behaviour, neuronal specific ion channels, and learning and memory. This is consistent with the reported enrichment of miRNAs within neural tissue of many species, where they appear to play a critical role in synaptic plasticity, axon guidance, as well as learning and memory, for review see<sup>43-45</sup>. For example, mir-124 is involved in neuronal cell differentiation through a negative feedback loop with the membrane protein ephrin-B1, a member of the ephrin signalling pathway<sup>46</sup>. In the developing mouse brain,

Gene (Cholinergic Synapse)	GB Number	Log Fold Change	FDR	Predicted Targeting miRNAs	Function
Acetylcholinesterase 1	GB43191	2.23	$9.15 \times 10^{-24}$	miR-6063, miR-6065, miR-3749, miR-6012, miR-3751	Hydrolysis of acetylcholine
Acetylcholinesterase 2	GB41856	2.21	$1.80 \times 10^{-28}$	miR-3718a, miR-3749, miR-3741, miR-3776, miR-3768, miR-3049-3p	
Voltage-dependent calcium channel type D subunit alpha-1	GB55480	2.23	$5.29 \times 10^{-12}$	miR-3741, miR-3739, miR-965-5p, miR-210, miR-3768	Mediate entry of calcium ions
Inositol 1,4,5,-tris-phosphate receptor	GB46583	0.39	$7.96 \times 10^{-9}$		Inositol triphosphate mediated calcium channel
Adenylate cyclase type 8	GB54593	0.39	$2.34 \times 10^{-5}$		Membrane-bound enzyme that catalyzes the formation of cAMP
Adenylate cyclase 3	GB45150	1.08	$2.28 \times 10^{-3}$		Membrane-associated enzyme that catalyzes the formation of cAMP
Phospholipase C	GB44174	-0.43	$1.00 \times 10^{-4}$	gi 323388982 ref NC_007075.3 :4427403-4427428:-, gi 323388984 ref NC_007073.3 :76047-76070:+, miR-3755, miR-6039, miR-210, miR-6012	Catalyzes the hydrolysis of phosphatidylinositol-4,5-bisphosphate to generate inositol 1,4,5-trisphosphate (IP3) and diacylglycerol (DAG)
Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta isoform	GB43074	1.72	$4.69 \times 10^{-60}$	miR-6056, miR-6059, miR-3730, gi 323388987 ref NC_007070.3 :9617943-9617963:-, miR-3746, miR-3741	Phosphorylates phosphatidylinositol 4,5-bisphosphate to generate phosphatidylinositol 3,4,5-trisphosphate (PIP3)
Protein kinase C	GB43222	2.90	$6.78 \times 10^{-46}$	miR-3751, miR-6066, miR-3792, gi 323388987 ref NC_007070.3 :9617943-9617963:-, miR-928, miR-3718b, gi 323388984 ref NC_007073.3 :251491-251510:+, miR-3776, miR-3775, miR-3739, miR-6042, miR-965-5p, miR-6059, miR-3741, gi 323388987 ref NC_007079.3 :1762825-1762846:-, miR-6056	Protein phosphorylation
Dual specificity mitogen-activated protein kinase kinase dSOR1	GB41067	0.76	$3.01 \times 10^{-3}$	miR-3741, miR-3777	Dual-specific protein kinase that sits down-stream of Raf in the MAPK pathway
Mitogen-activated protein kinase ERK-A	GB51503	0.10	$2.58 \times 10^{-6}$		Protein kinase that sits down-stream of Raf in the MAPK pathway
High-affinity choline transporter 1	GB43293	2.48	$5.62 \times 10^{-7}$	gi 323388982 ref NC_007075.3 :15585205-15585227:-, miR-6056, gi 323388972 ref NC_007085.3 :3190449-3190472:-, miR-6063, gi 323388987 ref NC_007070.3 :9617943-9617963:-, miR-6059, gi 323388984 ref NC_007073.3 :76047-76070:+	Imports choline from extracellular space
Guanine nucleotide-binding protein beta subunit 2	GB54181	1.63	$1.84 \times 10^{-163}$	miR-3049-5p, miR-6048, miR-6049, gi 323388976 ref NC_007081.3 :2775657-2775678:-, miR-210, miR-3775, miR-3776, gi 323388987 ref NC_007070.3 :1183784-1183805:+, gi 323388982 ref NC_007075.3 :15585205-15585227:-, miR-6056, miR-3740, miR-3741, gi 323388982 ref NC_007075.3 :15584822-15584844:-, miR-3751, miR-3717, gi 323388984 ref NC_007073.3 :251491-251510:+, miR-3790	Involved in the recruitment, assembly and/or regulation of a variety of signaling molecules
Guanine nucleotide-binding protein G(o) subunit alpha	GB54642	1.34	$2.92 \times 10^{-10}$	miR-282, miR-6063	
Guanine nucleotide-binding protein G(q) subunit alpha	GB43282	1.52	$7.91 \times 10^{-51}$	miR-263a, miR-14	
Potassium voltage-gated channel subfamily KQT member 1	GB51854	0.42	$5.97 \times 10^{-4}$	gi 323388972 ref NC_007085.3 :379513-379535:+, miR-3761, gi 323388984 ref NC_007073.3 :76047-76070:+, miR-3741, gi 319965553 ref NW_003378466.1 :4745-4762:+, miR-3717	Voltage-gated potassium channel, normally associated with cardiac cell repolarisation
cAMP-dependent protein kinase 3	GB48362	0.78	$4.24 \times 10^{-5}$		cAMP-dependent phosphorylation of target proteins
Calcium/calmodulin-dependent protein kinase II	GB49535	1.25	$4.59 \times 10^{-22}$	miR-3790, gi 323388979 ref NC_007078.3 :10730925-10730949:+	Calcium/calmodulin-dependent serine/threonine protein kinase
Phosphatidylinositol 3 kinase 21B	GB42200	0.95	$7.74 \times 10^{-22}$		Phosphorylation of the inositol ring of phosphatidylinositol
Muscarinic acetylcholine receptor	GB51689	0.94	$1.56 \times 10^{-4}$	miR-34	Cholinergic receptor family
Cyclic AMP response element-binding protein A	GB47941	0.62	$6.97 \times 10^{-4}$	miR-92a, miR-92b	Transcription factor

**Table 5. An example of gene enrichment in pathways associated with neuronal development (cholinergic synapse) in workers relative to queens.** Log Fold Change represents the expression level in workers relative to queens, as determined by high throughput sequencing. A negative value indicates a down-regulation in gene expression in workers relative to queens, a positive value indicates the opposite. FDR = False discovery rate adjusted P-value.





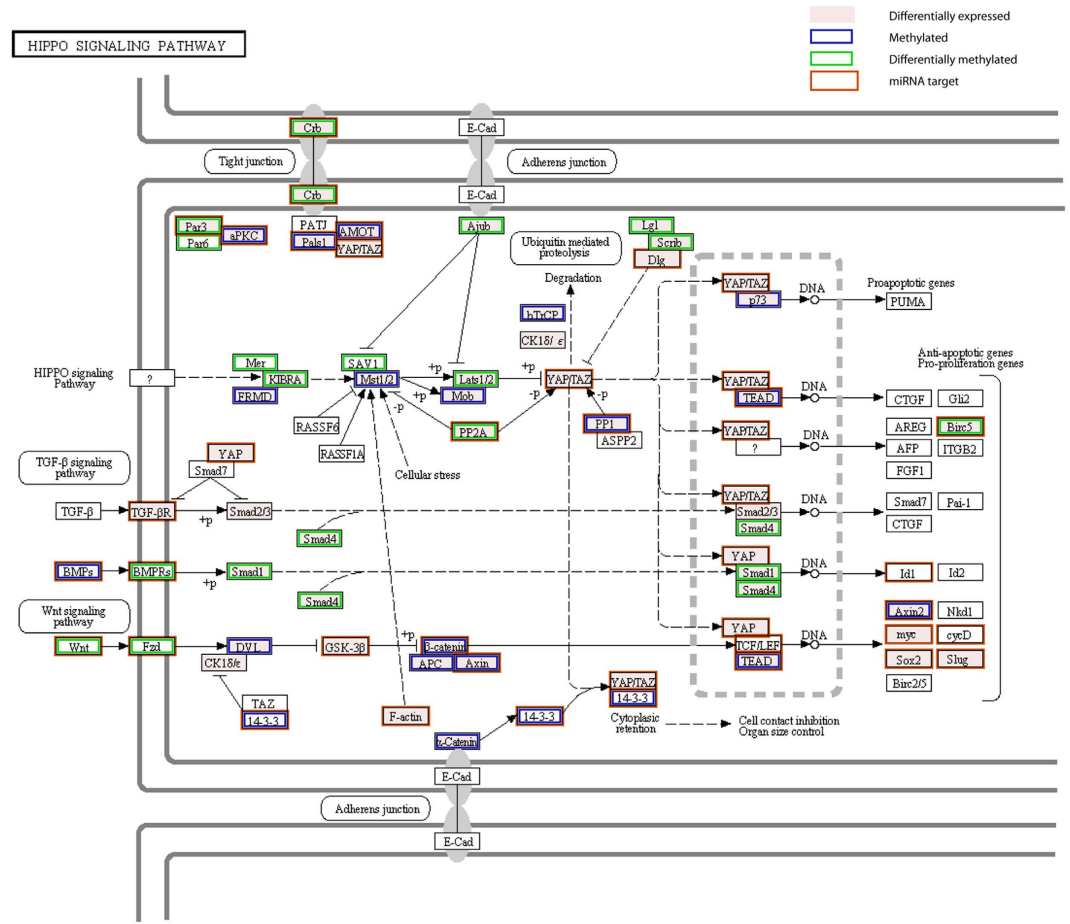
**Figure 4. Chromosomal distribution of miRNAs and their predicted mRNA target genes.** (A) Genome-wide view of the statistically significant correlations between miRNAs and their predicted targets. The outer circle represents the chromosomes, the next inner track represents the location of the miRNAs, followed by their targets and the links between the two. The links are coloured according to the value of the Pearson correlation coefficient. (B) Distribution of statistically significant Pearson correlation coefficients between miRNAs and the mRNA level of their target genes, indicating that more than half of the correlations are positive. A positive correlation indicates that the direct of change observed in the expression of the miRNA and its target transcript was the same. In contrast, a negative correlation indicates the change in expression of the miRNA and its target was in the opposite direction.

ephrin-B1 holds cells in their progenitor state, while mir-124 promotes their differentiation through inhibition of ephrin-B1 expression. In workers, the expression of ame-mir-124 and components of the ephrin signalling pathway (ephrin receptor tyrosine kinase (GB41629)<sup>47</sup>, ephrin-B2a (GB40333)<sup>47</sup>, ephrin protein (GB43375) and Pak1 (GB48749)) were significantly higher relative to queens, suggestive of significant neuronal maturation occurring in workers at L4/L5.

Other miRNA target pathways that show significant gene enrichment include; cell signalling (Notch, Wnt and steroid hormone signalling, as well as tyrosine kinase activity), apoptosis, modulation of transcription factor activity, and cellular differentiation/migration (chemotaxis, cell cycle arrest, cell migration, cell delamination, tube morphogenesis and anterior/posterior patterning). Pathways associated with both wing and leg imaginal disc morphogenesis also shows significant enrichment.

A number of the miRNAs found to be differentially expressed in our current study form critical components of the major cell signalling pathways in other species, including ame-Bantam, ame-mir-315 and ame-Let-7. Outside of *Drosophila*, the role of miRNAs in insect development has not been extensively studied. However, some insight can be drawn from other species, including *Drosophila*. For example, Bantam is an important target of the Hippo and EGFR/MAPK signalling pathways<sup>48–50</sup>, both critical in embryonic development and the control of cell proliferation and survival. In *Drosophila*, mir-315 expression affects the conserved Wingless (Wg or Wnt) signalling pathway<sup>51</sup>, which modulates tissue patterning and cell differentiation in many species. Mir-let-7 is a major target of the steroid pathways underlying temporal cell fate determination during the movement between developmental stages. Let-7 is one of the few consistent miRNA markers associated with the behavioural shift of worker bees from nurses to foragers<sup>27,29,31</sup>. Royal jelly has been reported to contain significant quantities of ame-mir-Let-7<sup>32</sup>, consistent with the idea that let-7 levels are higher in queens relative to workers at this critical larval stage. In *Drosophila*, the developmental progression from larvae to pupae has been shown to involve the modulation of the Let-7-C polycistronic miRNA cluster (Let-7, mir-100 and mir-125), as well as mir-8 and mir-34<sup>52</sup>, by the lipophilic hormones 20-hydroxy ecdysone (20E) and juvenile hormone (JH). Each of these miRNAs was also observed to be differentially expressed in our current study, with Ame-mir-34 showing one of the largest sets of predicted transcript targets (20). The majority of these predicted targets are associated with neuronal development or ion channel formation. One such transcript is semaphorin 2a (GB46091). In both vertebrates and invertebrates, semaphorins are involved in axon guidance and ultimately target selection and synapse formation<sup>53</sup>. In *Drosophila*, semaphorin 2a acts as a chemorepulsive axon guidance cue critical for neuronal pathfinding and synaptic arborisation<sup>54</sup>. Semaphorin 2a (GB46091) mRNA levels were significantly elevated in workers relative to queens, further suggestive of greater levels of neuronal development and maturation occurring in workers at this critical larval stage.

As mentioned above, movement between developmental stages in both *Drosophila* and the honeybee involves the activity of the steroid hormone 20E, which acts through the ecdysone receptor (EcR). In our current study, EcR mRNA levels were elevated in workers relative to queens. EcR expression and activity levels have been shown to be negatively regulated by mir-14<sup>55</sup>. Consistent with this role, ame-mir-14 levels were elevated in queens relative to workers.



**Figure 5. Regulation of the Hippo pathway in honeybee larvae by miRNAs and DNA methylation.** The Hippo pathway plays important roles in growth and development; it is enriched in differentially expressed genes, differentially methylated genes and miRNA targets. A pink shade indicates differential expression between any of the three larval castes. A blue outline indicates DNA methylation in queen or worker larvae, a green outline denotes differential DNA methylation between these. Predicted miRNA targets are outlined in brown. Based on KEGG map04390 ([http://www.genome.jp/kegg-bin/show\\_pathway?map04390](http://www.genome.jp/kegg-bin/show_pathway?map04390)).

It is noteworthy that broadly similar GO categories were identified by another parallel analysis where instead of looking at the targets at the intersection of different miRNA target prediction software, we looked at the intersection between the GO terms predicted by each software separately (see methods and Supplementary Table S4).

**DNA methylation and miRNAs target different developmental pathways.** When we considered miRNA target pairs predicted by at least two software approaches, we found that predicted targets were significantly underrepresented amongst methylated genes (Fisher exact test,  $p < 1 \times 10^{-6}$ ). When we looked at the predicted miRNA target pairs generated by each individual software, however, this underrepresentation was only confirmed with RNAhybrid ( $p < 1 \times 10^{-21}$ ), but not PITA ( $p = 0.24$ ) nor Miranda ( $p = 0.14$ ). Although this possible underrepresentation requires further analysis, DNA methylation appears to be predominantly associated with physiometabolic processes, a functional category that was not heavily targeted by miRNAs. Instead, miRNAs appear to overwhelmingly target neuronal maturation and caste biased structural differences in the honeybee. This suggests a distinct evolutionary division in the developmental processes that are regulated by DNA methylation and miRNAs.

Some overlap in the regulatory patterns of methylation and miRNAs was observed with respect to the major cell signalling pathways. The Hippo signalling pathway is one such striking example (Fig. 5). At the level of the whole larvae, this pathway was significantly enriched in differentially expressed genes between castes (Supplementary Table S2), many of which are predicted targets of miRNAs (Supplementary Table S3). Based on sequencing data from larval heads, this pathway is also enriched in differentially methylated genes between queens and workers (Fig. 5). The Hippo pathway plays an important role in the control of organ growth and cell proliferation<sup>56</sup>. Our results suggest that the development of morphological differences between castes requires a complex fine-tuning of this pathway by miRNAs and DNA methylation.

**Insights into gender-related disparity in immune systems in a social organism.** It is often argued that gender-specific immune responses exist because females gain fitness through increased longevity, whereas

males gain fitness by increasing mating rates. These contrasting investment strategies result in a higher susceptibility of males to pathogens and other stressors<sup>57</sup>. Our results add some support to the notion that females have stronger immune responses than males. For example, in both queen and worker larvae, genes encoding the antimicrobial peptides Defensin and Apidaecin belong to the most highly expressed category, whereas in males these peptides are expressed at extremely low levels. In the social context it is conceivable that drone larvae are partly shielded against stressors at the colony level by 'social immunity' that is provided by hygienic behaviour of adult nurse bees. These collective behaviours compensate for the smaller immune proteome in honeybees relative to solitary insects<sup>58</sup>. An extra level of protection for the queen larva is provided by royal jelly that has strong antimicrobial properties. The contribution of specific miRNAs to gender-associated immunologic processes in the honeybee is less straightforward because many immune-related miRNAs also regulate growth and apoptosis that are differentially manifested in the three castes. The highly conserved mir-1175 whose expression is affected by bacterial infection in adult honeybees<sup>59</sup> and in mosquitoes<sup>60</sup> shows differential expression during larval growth across all three castes. Several other miRNAs, postulated to be involved in immune regulation<sup>52,60–62</sup>, are also differentially expressed in our study, including; Let-7, mir-8, mir-12, mir-33, mir-34, mir-190, mir-279, mir-283 and mir-989.

**MicroRNAs in caste determination.** Our data clearly show that the expression profile of miRNAs is distinctly different between castes at this critical stage of larval development. These non-coding RNAs may therefore form an integral part of the molecular framework underlying developmental canalisation in the honeybee. The most important insight from our study is their apparent complementary role to that of DNA methylation based on miRNA target prediction. Our findings indicate that the interactions between miRNAs and their target transcripts form a highly interconnected network that appears to be biased towards regulating cellular processes involved in neuronal differentiation, cell signalling and the formation caste-specific structures. These processes are distinct from those controlled by DNA methylation, which have been shown to predominantly modulate physio-metabolic and information processing genes, whose products are known to regulate the rate of mass-transforming processes and the general growth of an organism<sup>2,11,17</sup>. The striking differences in both miRNAs and mRNAs profiles between queens, workers and drones occur during a critical larva-to-pupa transition when larval imaginal discs begin to differentiate into various outside body parts. The robustness of this process is essential to create the desired adult phenotypes associated with different life challenges. The streamlined behaviours of queens and drones are primarily geared towards reproduction, whereas sterile workers perform a wide range of tasks and undergo behavioural maturation that prepares them for complex hive building and foraging tasks. One possibility is that during the initial hours of larval development, a high level of transcriptional noise, resulting either from differential feeding (environment) or a haploid genome, is necessary to alter developmental trajectory. However, such alternative trajectories need to be rapidly stabilised to ensure phenotypic constancy<sup>63</sup>. A role of miRNAs in this context would be to buffer fluctuations in target gene expression, not only to ensure robustness of the desired phenotype, but also to suppress alternative ones. On the other hand, DNA methylation is a highly flexible epigenomic modifier, responding to both internal and external stimuli. It may therefore play a role in generating alternative patterns of gene expression and in doing so act as a driver of developmental plasticity. Indeed, silencing *de novo* methylation in newly hatched larvae results in a strong bias for queen phenotype<sup>8</sup>.

## Conclusion

Our study provides important insights into the possible role of miRNAs in shaping the conditional phenotypes associated with male and female castes in *Apis mellifera* and the associated flexible behaviours that arise from them in response to the integration of environmental and genomic signals. The distinct differences in behaviour, physiology, longevity and reproductive capacity between castes is driven by changes in the epigenetic landscape of specific transcriptional networks in response to diverse cues. We provide evidence that miRNAs may play a significant role in this process by targeting non-methylated genes associated primarily with neuronal development and thus complementing the regulatory effects of the DNA methylation toolkit. We propose that future studies on the epigenetic plasticity of different cell types, such as larval imaginal discs and their epigenomic signatures are required to achieve the in-depth understanding of novel developmental outcomes.

## Material and Methods

**Sample Collection.** Honeybee (*Apis mellifera* var. *ligustica*) larvae were collected from a single hive located at the Australian National University (ANU) in Canberra. To obtain specimens in similar developmental stages, newly capped queen, worker and drone larvae were removed from the frame and frozen in liquid nitrogen. Both queen and worker larvae were  $96 \pm 1$  hr-old at the time of collection, whereas drone larvae were  $96 \pm 5$  hr-old. This developmental phase corresponds to stage L5 after the 4<sup>th</sup> molt. At this point the commitment to a particular developmental trajectory is irreversible<sup>64</sup>. 25 larval samples for each caste were split into five biological replicates consisting of five larvae each.

**RNA extraction.** As previously described<sup>65</sup>, total RNA was extracted from larval samples using Trizol reagent, following the manufacturer's protocol (Invitrogen Life Technologies, Australia). Briefly, larval samples were homogenised in 500  $\mu$ l of Trizol reagent (3:1 Trizol to sample ratio). One hundred microliters of chloroform were then added and allowed to incubate for 3 min at room temperature, before the samples were centrifuged at 12,000 g for 10 min at 4 °C. Following centrifugation, the upper phase was collected, with total RNA precipitated using isopropanol. The pellet was washed briefly in 75% ethanol and dissolved in 50  $\mu$ l double-distilled water by incubating for 2–3 min at 80 °C. RNA integrity was determined by gel electrophoresis and quantified using a NanoDrop spectrophotometer.

**RNA Libraries Preparation for Illumina high throughput sequencing.** Briefly, small RNAs were size selected and purified from 10 µg of total RNA ( $n = 5$  per caste) using a 15% acrylamide:bis-acrylamide PAGE, 8M urea gel with the aid of a miRNA Marker (#N2102S, NEB). The size selected small RNAs were eluted from the gel by incubation in a 0.3M NaCl solution over night at 4 °C, before precipitation using the isopropanol/ethanol. The precipitated pellet was dissolved in 6.5 µL RNAase free, DEPC-treated, distilled water. Library preparation, which involved adaptor ligation, reverse transcription and PCR amplification, were carried out following the manufactures protocol for small RNA library preparations (#E7300S/L, New England BioLabs (NEB), Geneseach, Australia). Product verification and size selection of amplified cDNA libraries was undertaken using a 6% PAGE gel. Gel bands corresponding to 140–150 nucleotides in size, which corresponds to adaptor-ligated constructs formed from 21–30 nucleotide RNA fragments, were excised from the gel and extracted by overnight incubation in 0.3M NaCl at 4 °C. Library constructs were then precipitated with 3M NaOAc, chilled isopropanol, before being washed in 75% ethanol, with the final pellet dissolved in 10 µL 10 mM Tris-HCl, pH 8.5. Libraries were validated on a 2100 Bioanalyser (Agilent Technologies, Australia), using a high sensitivity DNA LabChip. mRNA libraries were prepared using the NEB ultra-directional kit (#E7420S). Sequencing was carried out at the Biomolecular Research Facility (John Curtin School of Medical Research, ANU, Canberra) on an Illumina HiSeq 2500 platform. Transcriptome and miRNA sequence information and raw counts have been submitted to the GEO NCBI database (accession number GSE61253).

**Validation of miRNAs by Stem-Loop PCR.** Stem-Loop qRT-PCR validation of miRNA expression, using Universal ProbeLibrary Probe 21 (Roche Diagnostics Australia, Castle Hill, Victoria), was adapted from the protocol of Varkonyi-Gasic, Wu<sup>66</sup>. Total RNA was isolated from larval samples ( $n = 5$  per caste) that were independent of those processed for sequencing, using the Trizol methodology detailed above. Total RNA purity was determined by gel electrophoresis and quantified using a NanoDrop spectrophotometer. Reverse transcription was performed in a 15 µL reaction volume containing; 15 ng of total RNA, 50 nM Stem-Loop RT Primer (Supplementary Table S5), which was first combined and heated to 65 °C for 5 minutes, 1x SuperScript VILO Reaction Mix (Invitrogen Life Technologies, Australia), 1x SuperScript VILO Enzyme Mix, and 20 units of RNaseOut (Invitrogen). The final reaction mixture (15 µL) was heated to 16 °C for 30 minutes, 42 °C for 30 minutes, with the reaction terminated by incubating at 85 °C for 5 minutes and then cooled to 4 °C on an Eppendorf Master thermocycler.

RT-PCR was performed on a StepOnePlus Real-Time PCR System (Applied Biosystems (Life Technologies), Melgrave, Vic, Australia), using the Universal ProbeLibrary probe assay detailed in Varkonyi-Gasic, Wu<sup>66</sup>. The 15 µL PCR reaction mixture consisted of; 1 µL of cDNA, 50 nM forward and universal reverse primers (Supplementary Table S5), 10 nM Universal Probe 21, and 1xFastStart Universal Probe Master Mix (Roche Diagnostics Australia). Cycling conditions were; initial denaturation at 95 °C for 5 minutes, followed by 40 cycles consisting of; 95 °C for 10 seconds, 60 °C for 30 seconds and 72 °C for 10 seconds. Specificity of the PCR products was determined by agarose gel electrophoresis, melt curve analysis and sequencing.

**Bioinformatics analyses.** Throughout this study, the honeybee genome assembly version 4.5 and the official gene set version 3.2<sup>58</sup> were used.

**Prediction of novel miRNAs.** Small RNA reads were first mapped to the honeybee rRNA and tRNA sequences to remove potential contamination from these transcripts. Novel miRNAs were predicted using the mirDeep2 software v2.0.0.5<sup>67</sup> with default parameters, discarding reads smaller than 18bp. Predictions were run first on the 15 (5 drones, 5 queens, 5 workers) samples separately, and then on all the reads from the different samples combined together. These 16 sets of predictions were then merged and manually curated. MirDeep2 was also used to calculate the read counts of each miRNA transcripts.

**Mapping mRNA.** RNA-seq reads were mapped onto the honeybee genome using the TopHat software v2.0.9<sup>68</sup> with the following parameters: —read-mismatches 5 —read-gap-length 4 —read-edit-dist 6 —read-realign-edit-dist 5 —min-intron-length 40 —min-segment-intron 40 —library-type fr-firststrand —b2-very-sensitive, and using the —GTF parameter to provide the gene models of the official gene set v3.2. Counts per transcripts were calculated using the htseq-count tool from the HTSeq v0.5.4p3 package<sup>69</sup> with the following parameters: —stranded =yes—mode= intersection-nonempty.

**Differential gene expression and gene set enrichment analysis.** Differentially expressed miRNAs and mRNAs were identified using the EdgeR package from Bioconductor release 2.12<sup>70</sup>. Only the miRNAs and mRNAs with at least 5 counts in at least 5 samples were considered. Differentially expressed genes were called at a stringent false discovery rate of 1%<sup>71</sup>. GO annotations were carried out with blast2go<sup>72</sup> and KEGG annotations with KASS<sup>73</sup>. Gene set enrichment analyses in GO terms or KEGG pathways were carried out with GoSeq<sup>74</sup> to correct for selection bias caused by gene length using a false discovery rate of 5%.

**Prediction of miRNA targets.** We searched for targets in the 3' UTR of genes models. For genes lacking a predicted 5' UTR, 500bp downstream of the stop codon were selected. We used three different target prediction methods (PITA<sup>75</sup>, Miranda<sup>76</sup> and RNAHybrid<sup>77</sup>) with a maximum free energy of -10, -20 and -27 kcal/mol respectively. As miRNA target predictions tend to have a high false positive rate, we refined our predictions by calculating the Pearson correlation coefficient between the expression levels of each miRNA and their targets as describe in<sup>78</sup>. Only miRNA target pairs significantly correlated with a false discovery rate lower than 5% were considered. As it has been reported that taking the intersection of different target prediction methods does not necessarily increase the true positive rate<sup>79,80</sup>, we used two different approaches to combine our target predictions. In the first approach, only targets predicted by two of the three methods were considered (Supplementary Table S3). In

the second approach, we considered the predictions separately and then looked at the overlapping enriched GO categories (Supplementary Table S4).

**DNA methylation.** DNA methylation analysis was carried out as described in Forêt *et al.*<sup>17</sup>.

**Data Access.** Sequence data and raw counts have been submitted to GEO NCBI (accession number GSE61253).

## References

- Miklos, G. L. G. & Maleszka, R. Epigenomic communication systems in humans and honey bees: From molecules to behavior. *Horm Behav* **59**, 399–406 (2011).
- Maleszka, R. Epigenetic integration of environmental and genomic signals in honey bees: the critical interplay of nutritional, brain and reproductive networks. *Epigenetics* **3**, 188–192 (2008).
- Weaver, N. Physiology of Caste Determination. *Annu Rev Entomol* **11**, 79–102 (1966).
- Maleszka, R. The social honey bee in biomedical research: realities and expectations. *Drug Discovery Today: Disease Models* **12**, 7–13 (2014).
- Gempe, T. & Beyreuther, M. Function and evolution of sex determination mechanisms, genes and pathways in insects. *Bioessays* **33**, 52–60 (2011).
- Hrassnigg, N. & Crailsheim, K. Differences in drone and worker physiology in honeybees (*Apis mellifera*). *Apidologie* **36**, 255–277 (2005).
- Kucharski, R., Foret, S. & Maleszka, R. EGFR gene methylation is not involved in Royalactin controlled phenotypic polymorphism in honey bees. *Scientific reports* **5**, 14070 (2015).
- Kucharski, R., Maleszka, J., Foret, S. & Maleszka, R. Nutritional control of reproductive status in honeybees via DNA methylation. *Science* **319**, 1827–1830 (2008).
- Patel, A. *et al.* The Making of a Queen: TOR Pathway Is a Key Player in Diphenic Caste Development. *PLoS one* **2**, e509–e515 (2007).
- Corona, M., Estrada, E. & Zurita, M. Differential expression of mitochondrial genes between queens and workers during caste determination in the honeybee *Apis mellifera*. *The Journal of experimental biology* **202**, 929–938 (1999).
- Barchuk, A. R. *et al.* Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC developmental biology* **7**, 70 (2007).
- Evans, J. D. & Wheeler, D. E. Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera*. *Proc Natl Acad Sci USA* **96**, 5575–5580 (1999).
- Cameron, R. C., Duncan, E. J. & Dearden, P. K. Biased gene expression in early honeybee larval development. *BMC genomics* **14**, 903 (2013).
- Evans, J. D. & Wheeler, D. E. Expression profiles during honeybee caste determination. *Genome biology* **2**, doi: 10.1186/gb-2000-2-1-research0001 (2001).
- Chen, X. *et al.* Transcriptome comparison between honey bee queen- and worker-destined larvae. *Insect biochemistry and molecular biology* **42**, 665–673 (2012).
- Cristino, A. S. *et al.* Caste development and reproduction: a genome-wide analysis of hallmarks of insect eusociality. *Insect Mol Biol* **15**, 703–714 (2006).
- Foret, S. *et al.* DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 4968–4973 (2012).
- Wirtz, P. & Beetsma, J. Induction of caste differentiation in the honey bee (*Apis mellifera* L.) by juvenile hormone. *Ent Exp Appl* **15**, 517–520 (1972).
- Spannhoff, A. *et al.* Histone deacetylase inhibitor activity in royal jelly might facilitate caste switching in bees. *Embo Rep* **12**, 238–243 (2011).
- Dickman, M. J., Kucharski, R., Maleszka, R. & Hurd, P. J. Extensive histone post-translational modification in honey bees. *Insect biochemistry and molecular biology* **43**, 125–137 (2013).
- Meister, G. Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet* **14**, 447–459 (2013).
- Vasudevan, S., Tong, Y. & Steitz, J. A. Switching from repression to activation: microRNAs can up-regulate translation. *Science* **318**, 1931–1934 (2007).
- Orom, U. A., Nielsen, F. C. & Lund, A. H. MicroRNA-10a binds the 5' UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell* **30**, 460–471 (2008).
- Iwasaki, S. & Tomari, Y. Argonaute-mediated translational repression (and activation). *Fly* **3**, 204–206 (2009).
- Posadas, D. M. & Carthew, R. W. MicroRNAs and their roles in developmental canalization. *Current opinion in genetics & development* **27C**, 1–6 (2014).
- Asgari, S. MicroRNA functions in insects. *Insect biochemistry and molecular biology* **43**, 388–397 (2013).
- Behura, S. K. & Whitfield, C. W. Correlated expression patterns of microRNA genes with age-dependent behavioural changes in honeybee. *Insect Mol Biol* **19**, 431–439 (2010).
- Chen, X. *et al.* Next-generation small RNA sequencing for microRNAs profiling in the honey bee *Apis mellifera*. *Insect Mol Biol* **19**, 799–805 (2010).
- Greenberg, J. K. *et al.* Behavioral plasticity in honey bees is associated with differences in brain microRNA transcriptome. *Genes, brain, and behavior* **11**, 660–670 (2012).
- Li, L. *et al.* Differences in microRNAs and their expressions between foraging and dancing honey bees, *Apis mellifera* L. *Journal of insect physiology* **58**, 1438–1443 (2012).
- Liu, F. *et al.* Next-generation small RNA sequencing for microRNAs profiling in *Apis mellifera*: comparison between nurses and foragers. *Insect Mol Biol* **21**, 297–303 (2012).
- Shi, Y. Y. *et al.* Epigenetic modification of gene expression in honey bees by heterospecific gland secretions. *PLoS one* **7**, e43727 (2012).
- Weaver, D. B. *et al.* Computational and transcriptional evidence for microRNAs in the honey bee genome. *Genome biology* **8**, R97 (2007).
- Griffiths-Jones, S. The microRNA Registry. *Nucleic acids research* **32**, D109–111 (2004).
- Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M. & Lai, E. C. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* **130**, 89–100 (2007).
- Ruby, J. G., Jan, C. H. & Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**, 83–86 (2007).
- Thompson, G. J., Kucharski, R., Maleszka, R. & Oldroyd, B. P. Genome-wide analysis of genes related to ovary activation in worker honey bees. *Insect Mol Biol* **17**, 657–665 (2008).
- Di Pasquale, G. *et al.* Influence of pollen nutrition on honey bee health: do pollen quality and diversity matter? *PLoS one* **8**, e72016 (2013).
- Park, J. M., Kunieda, T. & Kubo, T. The activity of Mblk-1, a mushroom body-selective transcription factor from the honeybee, is modulated by the ras/MAPK pathway. *J Biol Chem* **278**, 18689–18694 (2003).

40. Nunez, Y. O. *et al.* Positively correlated miRNA-mRNA regulatory networks in mouse frontal cortex during early stages of alcohol dependence. *BMC genomics* **14**, 725 (2013).
41. Lopez-Gomollon, S., Mohorianu, I., Szitty, G., Moulton, V. & Dalmay, T. Diverse correlation patterns between microRNAs and their targets during tomato fruit development indicates different modes of microRNA actions. *Planta* **236**, 1875–1887 (2012).
42. Gu, S. & Kay, M. A. How do miRNAs mediate translational repression? *Silence* **1**, 11 (2010).
43. Earls, L. R., Westmoreland, J. J., Zakharenko, S. S. & Non-coding, R. N. A. regulation of synaptic plasticity and memory: Implications for aging. *Ageing research reviews* **17**, 34–42 (2014).
44. Iyengar, B. R. *et al.* Non-coding RNA interact to regulate neuronal development and function. *Frontiers in cellular neuroscience* **8**, 47 (2014).
45. Iyer, A. N., Bellon, A. & Baudet, M. L. microRNAs in axon guidance. *Frontiers in cellular neuroscience* **8**, 78 (2014).
46. Arvanitis, D. N., Jungas, T., Behar, A. & Davy, A. Ephrin-B1 reverse signaling controls a posttranscriptional feedback mechanism via miR-124. *Mol Cell Biol* **30**, 2508–2517 (2010).
47. Vidovic, M., Nighorn, A., Koblar, S. & Maleszka, R. Eph receptor and ephrin signaling in developing and adult brain of the honeybee (*Apis mellifera*). *Dev Neurobiol* **67**, 233–251 (2007).
48. Herranz, H., Hong, X. & Cohen, S. M. Mutual repression by bantam miRNA and Capicua links the EGFR/MAPK and Hippo pathways in growth control. *Curr Biol* **22**, 651–657 (2012).
49. Nolo, R., Morrison, C. M., Tao, C., Zhang, X. & Halder, G. The bantam microRNA is a target of the hippo tumor-suppressor pathway. *Curr Biol* **16**, 1895–1904 (2006).
50. Thompson, B. J. & Cohen, S. M. The Hippo pathway regulates the bantam microRNA to control cell proliferation and apoptosis in *Drosophila*. *Cell* **126**, 767–774 (2006).
51. Silver, S. J., Hagen, J. W., Okamura, K., Perrimon, N. & Lai, E. C. Functional screening identifies miR-315 as a potent activator of Wingless signaling. *Proc Natl Acad Sci USA* **104**, 18151–18156 (2007).
52. Luhur, A., Chawla, G. & Sokol, N. S. MicroRNAs as components of systemic signaling pathways in *Drosophila melanogaster*. *Current topics in developmental biology* **105**, 97–123 (2013).
53. Culotti, J. G. & Kolodkin, A. L. Functions of netrins and semaphorins in axon guidance. *Current opinion in neurobiology* **6**, 81–88 (1996).
54. Winberg, M. L., Mitchell, K. J. & Goodman, C. S. Genetic analysis of the mechanisms controlling target selection: complementary and combinatorial functions of netrins, semaphorins, and IgCAMs. *Cell* **93**, 581–591 (1998).
55. Varghese, J. & Cohen, S. M. microRNA miR-14 acts to modulate a positive autoregulatory loop controlling steroid hormone signaling in *Drosophila*. *Genes & development* **21**, 2277–2282 (2007).
56. Varelas, X. The Hippo pathway effectors TAZ and YAP in development, homeostasis and disease. *Development* **141**, 1614–1626 (2014).
57. Zuk, M. The sicker sex. *PLoS pathogens* **5**, e1000267 (2009).
58. Elsik, C. G. *et al.* Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC genomics* **15**, 86 (2014).
59. Lourenco, A. P., Guidugli-Lazzarini, K. R., Freitas, F. C., Bitondi, M. M. & Simoes, Z. L. Bacterial infection activates the immune system response and dysregulates microRNA expression in honey bees. *Insect biochemistry and molecular biology* **43**, 474–482 (2013).
60. Winter, F., Edaye, S., Huttenhofer, A. & Brunel, C. Anopheles gambiae miRNAs as actors of defence reaction against Plasmodium invasion. *Nucleic acids research* **35**, 6953–6962 (2007).
61. Freitak, D., Knorr, E., Vogel, H. & Vilcinskas, A. Gender- and stressor-specific microRNA expression in *Tribolium castaneum*. *Biology letters* **8**, 860–863 (2012).
62. Garbuzov, A. & Tatar, M. Hormonal regulation of *Drosophila* microRNA let-7 and miR-125 that target innate immunity. *Fly* **4**, 306–311 (2010).
63. Maleszka, R., Mason, P. H. & Barron, A. B. Epigenomics and the concept of degeneracy in biological systems. *Brief Funct Genomics* **13**, 191–202 (2014).
64. Weaver, N. Effects of larval age on dimorphic differentiation of female honey bee. *Annals Entomological Society of America* **50**, 283–294 (1957).
65. Foret, S., Kucharski, R., Pittelkow, Y., Lockett, G. A. & Maleszka, R. Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. *BMC genomics* **10**, 472 (2009).
66. Varkonyi-Gasic, E., Wu, R., Wood, M., Walton, E. F. & Hellens, R. P. Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant methods* **3**, 12 (2007).
67. Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W. & Rajewsky, N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic acids research* **40**, 37–52 (2012).
68. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**, R36 (2013).
69. Anders, S., Theodor, P. & Huber, W. HTSeq — A Python framework to work with high-throughput sequencing data. *bioRxiv* **31**, 166–169 (2015).
70. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
71. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* **57**, 289–300 (1995).
72. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
73. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic acids research* **35**, W182–185 (2007).
74. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology* **11**, R14 (2010).
75. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nat Genet* **39**, 1278–1284 (2007).
76. Enright, A. J. *et al.* MicroRNA targets in *Drosophila*. *Genome biology* **5**, R1 (2003).
77. Rehmsmeier, M., Steffen, P., Hochsmann, M. & Giegerich, R. Fast and effective prediction of microRNA/target duplexes. *Rna* **10**, 1507–1517 (2004).
78. Jayaswal, V., Lutherborrow, M. & Yang, Y. H. Measures of association for identifying microRNA-mRNA pairs of biological interest. *PLoS one* **7**, e29612 (2012).
79. Alexiou, P., Maragkakis, M., Papadopoulos, G. L., Reczko, M. & Hatzigeorgiou, A. G. Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics* **25**, 3049–3055 (2009).
80. Ritchie, W., Flamant, S. & Rasko, J. E. Predicting microRNA targets and functions: traps for the unwary. *Nat Methods* **6**, 397–398 (2009).

## Author Contributions

All authors were involved in the design and analysis of experiments and the preparation of the manuscript. Experiments were undertaken by RA.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Ashby, R. *et al.* MicroRNAs in Honey Bee Caste Determination. *Sci. Rep.* **6**, 18794; doi: 10.1038/srep18794 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>