bioRxiv preprint doi: https://doi.org/10.1101/022657; this version posted August 11, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

Allele specific expression and methylation in the bumblebee, *Bombus terrestris*

Zoë N. Lonsdale[1], Kate D. Lee[2], Maria Kyriakidou[1], Harindra E. Amarasinghe[3], Despina Nathanael[1], Catherine J. O'Connor[4], and Eamonn B. Mallon[1]

[1]**Department of Genetics and Genome Biology, University of Leicester, Leicester, U.K.**
[2]**Bioinformatics and Biostatistics Support Hub (B/BASH), University of Leicester, Leicester, U.K.**
[3]**Academic Unit of Cancer Sciences, University of Southampton, Southampton, U.K.**
[4]**Cardiff School of Biosciences, Cardiff University, Cardiff, U.K.**

Corresponding author:
Eamonn B. Mallon[1]

Email address: ebm3@le.ac.uk

ABSTRACT

The social hymenoptera are emerging as models for epigenetics. DNA methylation, the addition of a methyl group, is a common epigenetic marker. In mammals and flowering plants methylation affects allele specific expression. There is contradictory evidence for the role of methylation on allele specific expression in social insects. The aim of this paper is to investigate allele specific expression and monoallelic methylation in the bumblebee, *Bombus terrestris*. We found nineteen genes that were both monoallelically methylated and monoallelically expressed in a single bee. Fourteen of these genes express the hypermethylated allele, while the other five express the hypomethylated allele. We also searched for allele specific expression in twenty-nine published RNA-seq libraries. We found 555 loci with allele-specific expression. We discuss our results with reference to the functional role of methylation in gene expression in insects and in the, as yet unquantified, role of genetic cis effects in insect allele specific methylation and expression.

INTRODUCTION

Epigenetics is the study of heritable changes in gene expression that do not involve changes to the underlying DNA sequence (Goldberg et al., 2007). Social hymenoptera (ants, bees, and wasps) are important emerging models for epigenetics (Glastad et al., 2011; Weiner and Toth, 2012; Welch and Lister, 2014; Yan et al., 2014). This is due to theoretical predictions for a role for an epigenetic phenomenon, genomic imprinting (parent of origin allele specific expression), in their social organisation (Queller, 2003), the recent discovery of parent-of-origin allele specific expression in honeybees (Galbraith et al., 2016), and data showing a fundamental role in social insect biology for DNA methylation, an epigenetic marker (Chittka et al., 2012).

In mammals and flowering plants, allele specific expression is often associated with methylation marks passed from parents to offspring (Reik and Walter, 2001). However DNA methylation is involved in numerous other cellular processes (Bird, 2002). There is contradictory evidence for the role of methylation on allele specific expression in social insects. Methylation is associated with allele specific expression in a number of loci in the ants *Camponotus floridanus* and *Harpegnathos saltator* (Bonasio et al., 2012). Recently, we found evidence for allele specific expression in bumblebee worker reproduction genes (Amarasinghe et al., 2015) and that methylation is important in bumblebee worker reproduction (Amarasinghe et al., 2014). However, other work on the honeybee *Apis mellifera* found no link between genes showing allele specific expression and known methylation sites in that species (Kocher et al., 2015).

The presence of allele specific expression does not necessarily mean an epigenetic process is involved. Allele specific expression is known to be caused by a number of genetic as well as epigenetic processes (Palacios et al., 2009). The genetic process usually involves cis effects such as transcription factor binding

bioRxiv preprint doi: https://doi.org/10.1101/022657; this version posted August 11, 2017. 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 4.0 International license.

sites, or less often, untranslated regions which alter RNA stability or microRNA binding (Farh et al., 2005).

The aim of this paper is to investigate allele specific expression and methylation in the bumblebee, *Bombus terrestris*. The recently sequenced genome of the bumblebee, *Bombus terrestris* displays a full complement of genes involved in the methylation system (Sadd et al., 2015). An extreme form of allele specific expression involves monoallelic expression, where one allele is completely silenced. In the canonical mammal and flowering plant systems, this is often associated with monoallelic methylation. In this paper, we examined the link between monoallelic methylation and monoallelic expression in the bumblebee, *Bombus terrestris* using an integrative approached previously used in human epigenetic studies (Harris et al., 2010). Namely, we compare two types of whole methylome libraries and an RNA-seq library from the same individual. In humans, this integrative approach has been independently validated by clonal bisulphite sequencing (Harris et al., 2010). MeDIP-seq is an immunoprecipitation technique that creates libraries enriched for methylated cytosines (Harris et al., 2010). Methyl-sensitive restriction enzymes can create libraries that are enriched for non-methylated cytosines (MRE-seq) (Harris et al., 2010). Genes found in both libraries are predicted to be monoallelically methylated, with the putatively hypermethylated allele being in the MeDIP-seq data and the putatively hypomethylated allele in the MRE-seq data (Harris et al., 2010). Monoallelic expression was identified in these loci from the RNA-seq library. If only one allele was expressed then we knew that these loci were both monoallelically methylated and monoallelically expressed in this bee. We confirmed this monoallelic expression in one locus using qPCR.

We then more generally searched for allele specific expression by analysing twenty nine published RNA-seq libraries from worker bumblebees (Harrison et al., 2015; Riddell et al., 2014). We identified heterozygotes in the RNA-seq libraries and measured the expression of each allele. We then identified loci that showed significant expression differences between their two alleles.

MATERIALS AND METHODS

Samples

Data from twenty-nine RNA-seq libraries were used for the allele specific expression analysis (six from Harrison *et al.* (Harrison et al., 2015), and twenty-three from Riddell *et al.* (Riddell et al., 2014). The Riddell bees came from two colonies, one commercially reared bumblebee colony from Koppert Biological Systems U.K. and one colony from a wild caught queen from the botanic gardens, Leicester. The Harrison bees were from four commercially reared colonies obtained from Agralan Ltd. A Koppert colony worker bee was used for the MeDIP-seq / MRE-seq / RNA-seq experiment. Bees from three different Koppert colonies were used for the qPCR analysis. Samples are outlined in Table 1. Colonies were fed *ad libitum* with pollen (Percie du sert, France) and 50 % diluted glucose/fructose mix (Meliose – Roquette, France). Before and during the experiments colonies were kept at 26°C and 60% humidity in constant red light.

Next generation sequencing

MeDIP-seq, MRE-seq and RNA-seq

RNA and DNA was extracted from a single five day old whole bee (Colony K2). DNA was extracted using an ethanol precipitation method. Total RNA was extracted using Tri-reagent (Sigma-Aldrich, UK).

Three libraries were prepared from this bee by Eurofins genomics. These were MeDIP-seq and MRE-seq libraries on the DNA sample and one amplified short insert cDNA library with size of 150-400 bp on the RNA sample. Both the MeDIP-seq and MRE-seq library preparations are based on previously published protocols (Harris et al., 2010). MeDIP-seq uses monoclonal antibodies against 5-methylcytosine to enrich for methylated DNA independent of DNA sequence. MRE-seq enriches for unmethylated cytosines by using methylation-sensitive enzymes that cut only restriction sites with unmethylated CpGs. Each library was individually indexed. Sequencing was performed on an Illumina HiSeq®2000 instrument (Illumina, Inc.) by the manufacturer's protocol. Multiplexed 100 base paired-read runs were carried out yielding 9390 Mbp for the MeDIP-seq library, 11597 Mbp for the MRE-seq library and 8638 Mbp for the RNA-seq library.

bioRxiv preprint doi: https://doi.org/10.1101/022657; this version posted August 11, 2017. 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 4.0 International license.

Table 1. Bees used in each experiment. K refers to Koppert, A to Agralan and Q to the wild caught Leicester queen.

Experiment	Number	Colony	Tissue
Allele specific expression RNA-seq	1	A1	Whole body
	2	A2	Whole body
	2	A3	Whole body
	1	A4	Whole body
	14	K1	Abdomen
	9	Q1	Abdomen
MeDip/MRE/RNA-seq	1	K2	Whole body
qPCR	2	K3	Head
	1	K4	Head
	1	K5	Head

97 ***Previously published RNA-seq***

98 Full details of the RNA-seq protocols used have been published previously (Harrison et al., 2015;
99 Riddell et al., 2014). Briefly, for the Riddell bees, total RNA was extracted from twenty three individual
100 homogenised abdomens using Tri-reagent (Sigma-Aldrich, UK). TruSeq RNA-seq libraries were made
101 from the 23 samples at NBAF Edinburgh. Multiplexed 50 base single-read runs was performed on an
102 Illumina HiSeq®2000 instrument (Illumina, Inc.) by the manufacturer's protocol. For the Harrison
103 bees, total RNA was extracted from whole bodies using a GenElute Mammalian Total RNA Miniprep kit
104 (Sigma-Aldrich) following the manufacturers' protocol. The six libraries were sequenced as multiplexed
105 50 base single-read runs on an Illumina HiSeq 2500 system in rapid mode at the Edinburgh Genomics
106 facility of the University of Edinburgh.

107 **Monoallelic methylation and expression - Bioinformatic analysis**

108 We searched for genes that were monoallelically methylated (present in both MeDip-seq (the putatively
109 hypermethylated allele) and MRE-seq (the putatively hypomethylated allele) libraries), heterozygous
110 (different alleles in the methylation libraries) and monoallelically expressed (only one allele present in the
111 RNA-seq library).

112 ***Alignment and bam refinement***

113 mRNA reads were aligned to the *Bombus terrestris* genome assembly (AELG00000000) using Tophat
114 (Kim et al., 2013) and converted to bam files with Samtools (Li et al., 2009). Reads were labelled with
115 the AddOrReplaceReadGroups.jar utility in Picard (http://picard.sourceforge.net/). The MRE-seq and
116 MeDIP-seq reads were aligned to the genome using BWA mapper (Li and Durbin, 2009). The resultant
117 sam alignments were soft-clipped with the CleanSam.jar utility in Picard and converted to bam format
118 with Samtools. The Picard utility AddOrReplaceReadGroups.jar was used to label the MRE and MeDIP
119 reads which were then locally re-aligned with GATK (DePristo et al., 2011; McKenna et al., 2010). PCR
120 duplicates for all bams (mRNA, MeDIP and MRE) were marked with the Picard utility Markduplicates.jar.

121 ***Identifying regions of interest and integrating data***

122 Coverage of each data type was calculated using GATK DepthofCoverage (McKenna et al., 2010). Only
123 regions with a read depth of at least six in each of the libraries (RNA-seq, MeDIP-seq and MRE-seq) was
124 used. Heterozygotes were identified using Samtools mpileup and bcftools on each data set separately (Li
125 and Durbin, 2009) and results were merged with vcf tools (Danecek et al., 2011). Regions of mRNA with
126 overlaps of MeDIP, MRE, and monoallelic snps were identified with custom perl scripts.

127 **Allele specific expression - Bioinformatic analysis**

128 We created a pipeline to search for heterozygous loci that show allele specific expression and identify the
129 associated enriched gene ontology (GO) terms in twenty-nine previously published RNA-seq libraries
130 (Harrison et al., 2015; Riddell et al., 2014).

bioRxiv preprint doi: https://doi.org/10.1101/022657; this version posted August 11, 2017. 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 4.0 International license.

131 Each RNA library was mapped to the *Bombus terrestris* reference genome (Bter 1.0, accession 132 AELG00000000.1) (Sadd et al., 2015) using the BWA mapper (Li and Durbin, 2009). The combat method 133 in the R package SVA (version 3.20.0) was used to remove any batch effects and control for original 134 differences in coverage (Leek et al., 2012; Johnson et al., 2007). The success of this control was confirmed 135 by the R package edgeR (version 3.14.0) (McCarthy et al., 2012; Robinson et al., 2010).

136 Bcftools (version 0.1.19-44428cd), bedtools (version 2.17.0), and samtools (version 0.1.19-44428cd) 137 were used to prepare the RNA libraries and call the SNPs, before the SNPs were filtered based on mapping 138 quality score (Quinlan and Hall, 2010; Li and Durbin, 2009). Only SNPs with a mapping quality score of 139 $p < 0.05$ and a read depth of ≥ 6 were included in the analyses.

140 The R package, QuASAR implements a statistical method for: 1) genotyping from next-generation 141 sequencing reads (according to the Hardy-Weinberg equilibrium), and 2) conducting inference on allele 142 specific expression at heterozygous sites (Harvey et al., 2015). One problem with genotyping heterozy-143 gotes is being able to identical true homozygotes that appear heterozygote due to base-calling errors. 144 QuASAR removes snps with extreme differential allele expression from the analyses, thus controlling for 145 any base-calling errors. Despite this inherent conservatism, in benchmark tests, QuaSAR can accurately 146 genotype loci with lower error rates than other methods commonly used for genotyping DNA-seq data 147 (Harvey et al., 2015). The allele specific expression inference step takes into consideration the uncertainty 148 in the genotype calls, base-call errors in sequencing, and allelic over-dispersion. QuASAR is a powerful 149 tool for detecting allele specific expression if, as during most RNA-seq experiments, genotypes are not 150 available (Harvey et al., 2015).

151 Sequence regions (the snp position +/- 2900bp), encompassing the loci identified as showing ASE 152 in at least three of the thirty libraries, were compared to *Drosophila melanogaster* proteins (non-153 redundant (nr) database) with Blastx (Altschul et al., 1997). The blast results were annotated us-154 ing Blast2Go (Gotz et al., 2008). We carried out an enrichment analysis (Fisher exact test) using a 155 custom R script (https://dx.doi.org/10.6084/m9.figshare.3201355.v1) on this list of GO terms. This 156 identified GO terms that are overrepresented ($p < 0.05$) relative to the entire bumblebee transcriptome 157 (https://dx.doi.org/10.6084/m9.figshare.3458828.v1). We then used REVIGO to summarize and visualise 158 these terms (Supek et al., 2011). REVIGO summarizes lists of GO terms using a clustering algorithm 159 based on semantic similarity measures. To identify which bumblebee genes the snps were located in, the 160 snp position +/- 25 bp was compared against the *Bombus terrestris* genome (Sadd et al., 2015) using 161 Blastn.

162 Candidate gene allele specific qPCR

163 DNA was extracted from four bees from three Koppert colonies using the Qiagen DNA Micro kit according 164 to manufacturer's instructions. RNA was extracted from samples of the heads of the same worker bees 165 with the QIAGEN RNeasy Mini Kit according to manufacturer's instructions. cDNA was synthesized from 166 a 8 µl sample of RNA using the Tetro cDNA synthesis Kit (Bioline) as per manufacturer's instructions.

167 We amplified numerous fragments of the 19 candidate genes. Sanger sequencing results were analyzed 168 using the heterozygote analysis module in Geneious version 7.3.0 to identify heterozygotic nucleotide 169 positions. It was difficult to identify snps in exonic regions of the 19 loci, which could be amplified 170 with primers of suitable efficiency. We managed to identify a suitable region in *toll-like receptor Tollo* 171 (AELG01000623.1 exonic region 1838-2420).

172 The locus was run for 3 different reactions; T allele, G allele and reference. Reference primers were de-173 signed according to Gineikiene et al. (2009). A common reverse primer (CTGGTTCCCGTCCAATCTAA) 174 was used for all three reactions. A reference forward primer (CGTGTCCAGAATCGACAATG) was 175 designed to the same target heterozygote sequence, upstream of the heterozygote nucleotide position. The 176 reference primers measure the total expression of the gene, whereas the allele specific primers (T allele: 177 CCAGAATCGACAATGACTCGT, G allele: CAGAATCGACAATGACTCGG) measure the amount of 178 expression due to the allele. Thus the ratio between the allele specific expression and reference locus 179 expression would be the relative expression due to the allele.

180 Three replicate samples were run for each reaction. All reactions were prepared by the Corbett 181 robotics machine, in 96 well qPCR plates (Thermo Scientific, UK). The qPCR reaction mix (20 µl) was 182 composed of 1 µl of diluted cDNA (50ng/ µl), 1 µl of forward and reverse primer (5 µM/ µl each), 10 µl 183 2X SYBR Green JumpStart Taq ReadyMix (Sigma Aldrich, UK) and 7 µl ddH_20. Samples were run in a 184 PTC-200 MJ thermocycler. The qPCR profile was; 4 minutes at 95°C denaturation followed by 40 cycles

bioRxiv preprint doi: https://doi.org/10.1101/022657; this version posted August 11, 2017. 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 4.0 International license.

of 30s at 95°C, 30s at 59°C and 30s at 72°C and a final extension of 5 minutes at 72°C.

Forward primers are different, both in their terminal base (to match the snp) and in their length. It is entirely possible that they may amplify more or less efficiently even if there was no difference in amount of template (Pfaffl, 2001). To test for this we repeated all qPCRs with genomic DNA (1 µl of diluted DNA (20ng/ µl) from the same bees as the template. We would expect equal amounts of each allele in the genomic DNA. We also measured efficiency of each reaction as per Liu and Saint (2002).

Median C_t was calculated for each set of three technical replicates. A measure of relative expression (ratio) was calculated for each allele in each worker bee as follows:

$$ratio_{allele} = \frac{E_{allele}^{-Ct_{allele}}}{E_{reference}^{-Ct_{reference}}} \quad (1)$$

E is the median efficiency of each primer set (Liu and Saint, 2002; Pfaffl, 2001). All statistical analysis was carried out using R (3.3.1) (core Team, 2016).

Data Availability

All sequence data for this study are archived at the NCBI Sequence Read Archive (SRA) Accession no. PRJEB9366 and PRJNA391408. GO-analysis results and lists of differentially expressed transcripts are available as Supporting Information.

RESULTS

Discovery of monoallelically methylated and expressed genes

In total, we found nineteen genes that were both monoallelically methylated (present in both MeDIP and MRE-seq libraries) and monoallelically expressed (only one allele present in the RNA-seq library). Figure 1 and Figure 2 show the coverage of the three libraries for two examples of these genes (*ras GTPase-activating protein nGAP-like* and *bicaudal-D*). Of the nineteen genes, fourteen had the hypermethylated (MeDIP) allele expressed, while five had the hypomethylated (MRE-seq) allele expressed (see supplementary table 1). The nineteen genes were compared to the nr/nt database using Blastn. Six returned noninformative hits (Table 2).

Confirmation of monoallelic expression

Monoallelic expression was confirmed in one of these nineteen (*toll-like receptor Tollo* (LOC100644648)) by allele specific qPCR (Amarasinghe et al., 2015). The allele with a guanine at the snp position had a mean expression of 6.04 $\pm$8.28 (standard deviation) in four bees from three different colonies. The thymine allele was not expressed at all in these bees. This was not due to the efficiency of the primers as the DNA controls of both alleles showed similar amplification (G mean = 422.70 $\pm$507.36, T mean = 1575.17 $\pm$503.02). In the three other loci tested (*Ras GTPase-activating protein 1*, LOC107964816, *Elbow*) we found apparent monoallelic expression, but could not dismiss primer efficiency as the cause.

We then looked at these nineteen genes in twenty-nine previously published RNA-seq libraries. Fifteen of these nineteen genes expressed a single allele in all twenty nine RNA-seq libraries, see supplementary table 2. The remaining four genes were inconsistent; they showed expression of one allele in some *B. terrestris* workers, and expression of two alleles in other workers.

Removing batch effects

The twenty nine RNA-seq libraries do not come from the same experiment (Table 1). This gives rise to the possibility of batch effects, sources of variation due to samples not being from the same source or not being run together. We must remove these before any other analysis.

The mean GC content of the 29 libraries was 42.34%, with individual libraries having a similar GC content ranging from 40-46%. GC content differed with run (Nested ANOVA: F = 20.302, df = 1, $p < 0.001$), but not by colony (Nested ANOVA: F = 1.763, df = 4, p = 0.171). The mean coverage of the 29 libraries was 13.29, with mean library coverage ranging from 9.84 to 17.61. Run had an effect on coverage (Nested ANOVA: F = 7.554, df = 1, p = 0.011), as did colony (Nested ANOVA: F = 6.962, df = 4, $p < 0.001$).

bioRxiv preprint doi: https://doi.org/10.1101/022657; this version posted August 11, 2017. 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 4.0 International license.

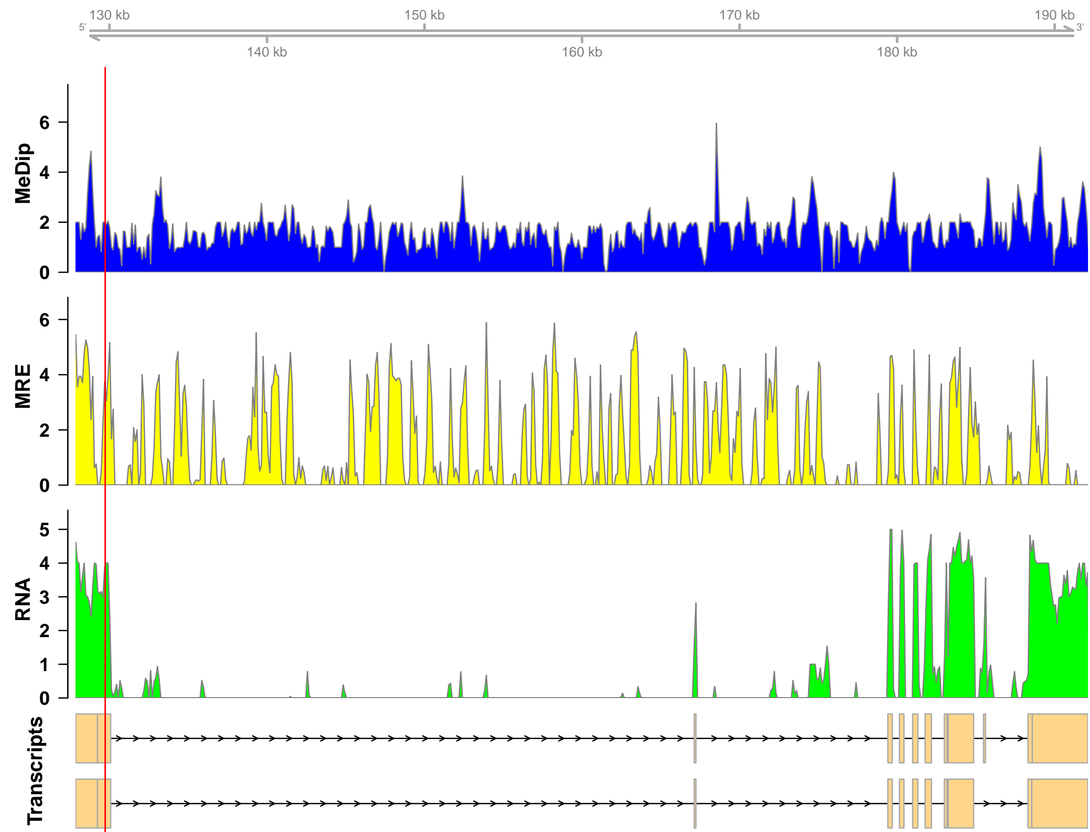

Figure 1. Coverage of the three libraries for *ras GTPase-activating protein nGAP-like* (LOC100652225). The transcript models come from GCF_000214255.1_Bter_1.0. The y-axis in the coverage plots is log (1 + coverage). The red vertical line represents the heterozygote position. The MeDip allele was expressed in this locus, see Table 2).

bioRxiv preprint doi: https://doi.org/10.1101/022657; this version posted August 11, 2017. 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 4.0 International license.

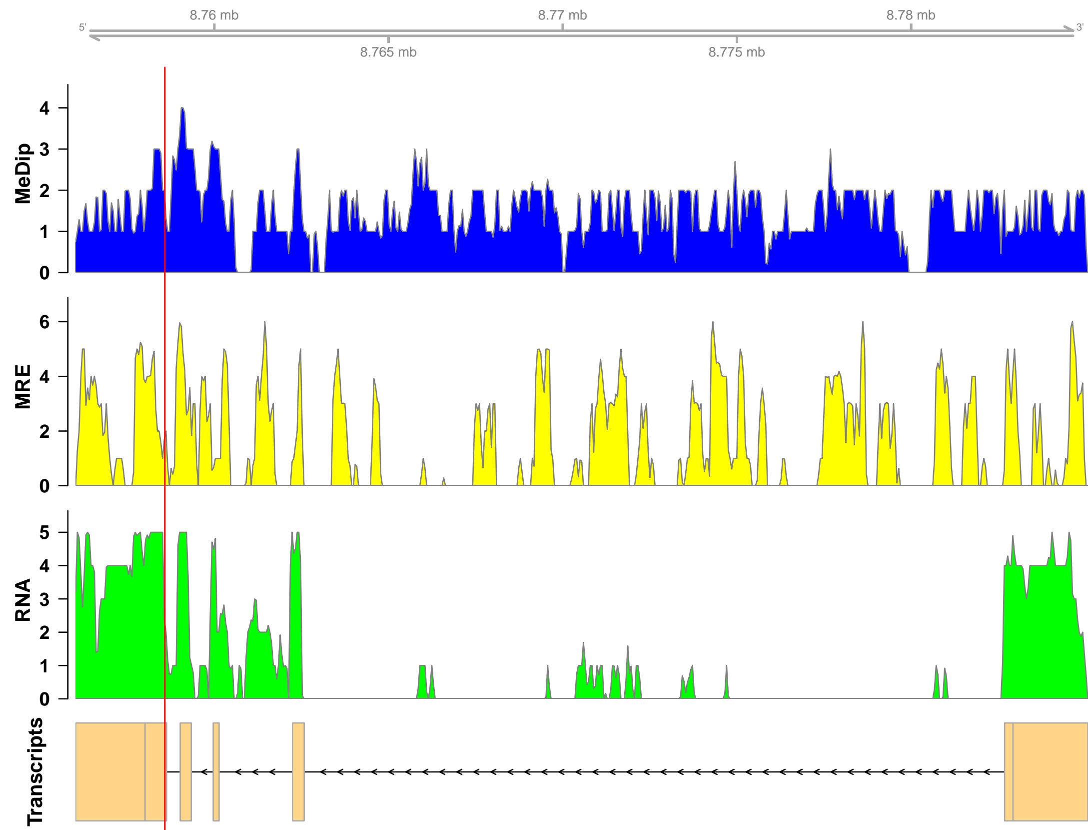

Figure 2. Coverage of the three libraries for *bicaudal D-related protein homolog* (LOC100650109). The transcript model come from GCF_000214255.1_Bter_1.0. The y-axis in the coverage plots is log (1 + coverage). The red vertical line represents the heterozygote position. The MeDip allele was expressed in this locus, see Table 2).

bioRxiv preprint doi: https://doi.org/10.1101/022657; this version posted August 11, 2017. 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 4.0 International license.

Gene	Accession	Expressed allele	Function
yippee-like 1	LOC100642754	MeDIP	Yippee is an intracellular protein with a zinc-finger like domain. DNA methylation of a CpG island near the *yippie-like 3* promoter in humans represents a possible epigenetic mechanism leading to decreased gene expression in tumours (Kelley et al., 2010).
toll-like receptor Tollo	LOC100644648	MeDIP	Tollo regulates antimicrobial response in the insect respiratory epithelium (Akhouayri et al., 2011).
zinc finger protein Elbow	LOC100650465	MeDIP	The elbow (elB) gene is involved in the formation of the insect tracheal system (Dorfman et al., 2002).
heterogeneous nuclear ribonucleoprotein A3	LOC100651168	MeDIP	Heterogeneous nuclear ribonucleoproteins associated with precursors of functional, protein coding mRNAs (Dreyfuss et al., 1993).
calmodulin-lysine N-methyltransferase-like	LOC100749522	MRE	Calmodulin-lysine N-methyltransferase catalyses the trimethylation of a lysine residue of calmodulin. Calmodulin is a ubiquitous, calcium-dependent, eukaryotic signalling protein with a large number of interactors. The methylation state of calmodulin causes phenotypic changes in growth and developmental processes (Magnani et al., 2010).
Na/K/Ca exchanger CG1090	LOC107998466	MRE	CG1090 functions in the maintenance of calcium homeostasis.
Shaker	LOC100648438	MeDIP	Shaker is involved in the operation of potassium ion channel. *Shaker* expression was upregulated in sterile versus reproductive honeybee workers (Cardoen et al., 2011) .
Centrosomal and chromosomal factor-like	LOC105665737	MeDIP	Essential protein required for proper condensation of mitotic chromosomes and progression through mitosis. Expressed during oogenesis in *Drosophila* (Kodjabachian et al., 1998).
excitatory amino acid transporter 1	LOC100744217	MRE	Excitatory amino acid transporters are neurotransmitter transporters. *Excitatory amino acid transporter 3* expression was upregulated in sterile honeybee workers (Cardoen et al., 2011). *Excitatory amino acid transporter 1* expression differences were also associated with worker - queen differentiation in the paper wasp *Polistes metricus* (Toth et al., 2014).
aminopeptidase M1-like	LOC105666993	MeDIP	M1 aminopeptidases are zinc-dependent enzymes that catalyze the removal of amino acids from the N terminus of polypeptides (Drinkwater et al., 2017).
ras GTPase-activating protein nGAP-like	LOC100652225	MeDIP	*Ras GTPase-activating protein 1* was found to be upregulated in reproductive honeybee workers (Cardoen et al., 2011). It is involved in oocyte meiosis.
neuromedin-B receptor-like	LOC100745453	MeDIP	In humans, this G protein-coupled receptor binds neuromedin B, a peptide that stimulates mitosis in gastrointestinal epithelial tissue.
bicaudal D-related protein homolog	LOC100650109	MeDIP	Bicaudal is involved in embryonic pattern formation in *Drosophila* (Markesich et al., 2000). It is thought to be involved in the differentiation between soldiers and workers in the termite *Reticulitermes flavipes* (Scharf et al., 2003). *Bicaudal protein D* has been shown to be methylated more in eggs than sperm in honeybees (Drewell et al., 2014).

Table 2. The thirteen of the nineteen monoallelically methylated and expressed genes that returned informative blast hits.

bioRxiv preprint doi: https://doi.org/10.1101/022657; this version posted August 11, 2017. 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 4.0 International license.

Therefore, the combat method in the R package SVA (version 3.20.0) was used to remove any batch effects and control for original differences in coverage (Leek et al., 2012; Johnson et al., 2007). The success of this control was confirmed by the R package edgeR (version 3.14.0) (McCarthy et al., 2012; Robinson et al., 2010). The SVA adjustment reduced the edgeR dispersion value from 3.9994 (BCV=2) to 0 (BCV=0.0003) (see Figure 3). That is we successfully removed the batch effects due to the separate runs.

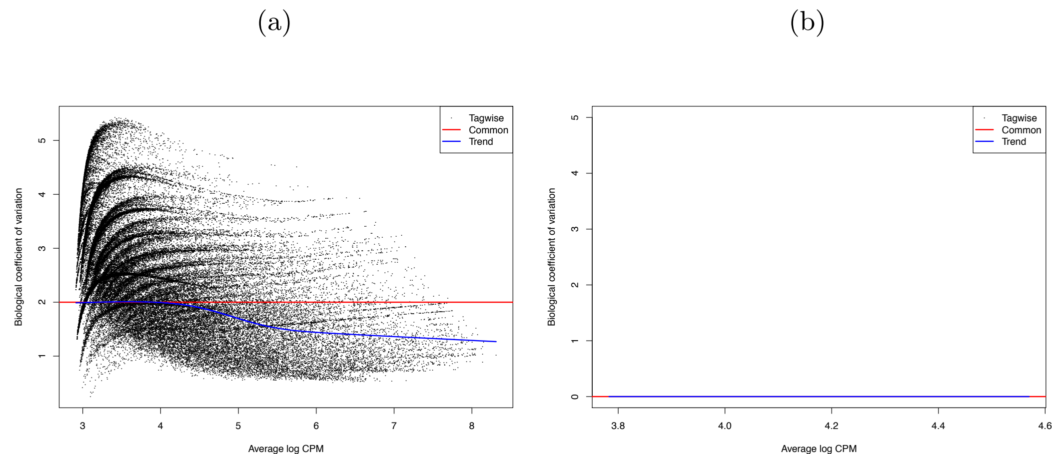

Figure 3. Biological coefficient of variation (BCV) of a) raw data, and b) SVA-adjusted data for the 29 RNA-seq *Bombus terrestris* libraries. The black dots represent the BCV if it were calculated individually for each gene (tagwise). The blue line is the trend of this data. The red line represents the BCV of the samples if a common dispersion value, over all genes, were used. In (b) tagwise values are exactly the same as common values so no black dots are visible.

Allele specific expression - RNA-seq

We then searched more generally for allele specific expression in the twenty-nine RNA-seq libraries. 555 loci showed allele-specific expression in ≥ 3 of the 29 RNA-seq libraries (supplementary table 3). Comparing these loci against the *Bombus terrestris* genome using Blastn returned 211 hits. To search for gene ontology terms, we compared them against *Drosophila melanogaster* proteins, using Blastx, which returned 329 hits. We tested for enriched gene ontology (GO) terms against their background value in the bumblebee transcriptome. One hundred and fifty-one Gene Ontology(GO) terms were enriched in the 555 regions showing allele specific expression (Fisher's exact test $p > 0.05$), however none were significant at the more stringent FDR > 0.05. Figure 4 shows the large number of biological functions associated with these 555 genes.

DISCUSSION

An important caveat about the integrative analysis of monoallelic methylation and expression carried out here is that all three libraries were from a single bee. It is certain that there is variation in methylation and allele specific expression between bees just as there is in other species (Pignatta et al., 2014). We attempted to confirm this monoallelic expression in other bees using RNA-seq and qPCR but with limited success. This analysis is only a first step in understanding the link between monoallelic methylation and expression.

Of the nineteen genes displaying monoallelic methylation and monoallelic expression, fourteen had the hypermethylated (MeDIP) allele expressed, while five had the hypomethylated (MRE-seq) allele expressed (see supplementary table 1). In ant genes with allele specific methylation, the hypermethylated allele showed more expression than the hypomethylated allele (Bonasio et al., 2012). This fits with genome wide analysis that shows exonic methylation in insects associated with increased gene expression

bioRxiv preprint doi: https://doi.org/10.1101/022657; this version posted August 11, 2017. 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 4.0 International license.

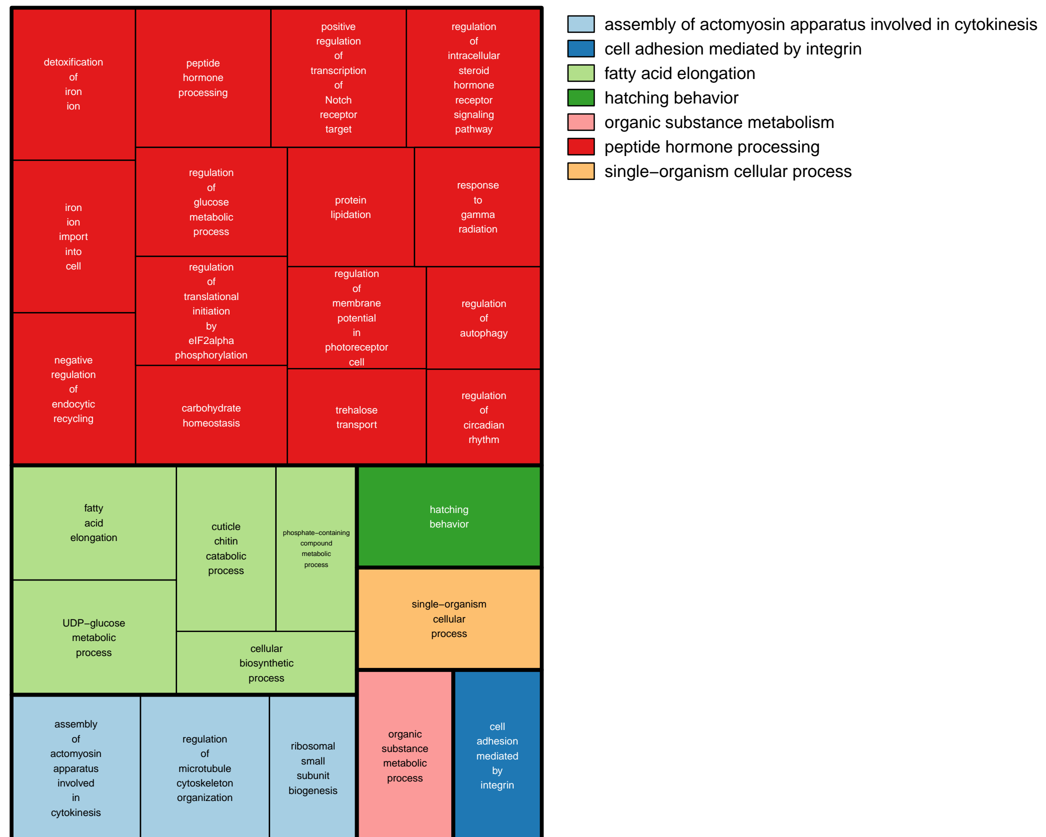

Figure 4. GO terms associated with allele specific expression. A summary of the enriched GO terms ($p < 0.05$, based on Blast2Go annotation) found for genes displaying allele specific expression. This figure was produced using Revigo (Supek et al., 2011). Each rectangle represents a single cluster of closely related GO terms. These rectangles are joined into different coloured 'superclusters' of loosely related terms. The area of the rectangles represents the p-value associated with that cluster's enrichment.

bioRxiv preprint doi: https://doi.org/10.1101/022657; this version posted August 11, 2017. 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 4.0 International license.

(Glastad et al., 2014; Yan et al., 2015). Our fourteen genes with the hypermethylated allele expressed agree with this pattern. But how to explain the five genes where the hypomethylated allele was expressed? Firstly, the role of methylation in insect gene expression is not clear cut, with the relationship between exonic methylation and expression often disappearing at the gene level (Yan et al., 2015). For example, *EGFR* expression is lower in ant workers that exhibit higher DNA methylation of *EGFR* (Alvarado et al., 2015). Secondly, even in the canonical mammalian methylation system, the "wrong" allele has been shown to be expressed occasionally due to lineage specific effects (Dean et al., 1998; Pardo-Manuel de Villena et al., 2000; Onyango et al., 2002; Sapienza, 2002; Zhang et al., 1993).

We analysed RNA-seq libraries from different published sources. This lead to two confounding problems. The first is that as the samples were run at different times, using different machines this could lead to a batch effect. We were able to successfully remove this. The second, that the libraries were made from abdomens in some cases and whole bodies in others, is still a confounding effect. Allele specific expression is known to vary between tissues (Chamberlain et al., 2015). Any variation in which allele is expressed could be due to these tissue effects.

We looked at the expression of the nineteen genes in all twenty-nine RNA-seq libraries. If they are monoallelically expressed in these bees, we would find only one allele in a given RNA-seq library. Fifteen of these nineteen genes were confirmed to show a single allele in all twenty-nine RNA-seq libraries. We would also find only one allele if that bee was homozygous. We cannot rule out that these fifteen genes just happen to be homozygous in all twenty-nine bees from five different colonies from multiple sources.

The remaining four genes showed inconsistent expression with one allele being expressed in some *B. terrestris* workers, and expression of two alleles in other workers. Natural intraspecific variation in allele specific expression has been found in other species (Pignatta et al., 2014). The tissue variation mentioned above is also a possibility. Another explanation is that these loci are not epigenetically controlled but rather their allele specific expression is derived from genetic effects (Remnant et al., 2016).

There are three main genetic, as opposed to epigenetic, affectors of allele specific expression (Edsgard et al., 2016). Allele specific expression can be caused by differences in the alleles' sequence within the translated part resulting in a modified protein. A change at the alleles' cis regulatory sites, could cause differential binding of transcription factors. Transcript processing can be affected by a change in the alleles' sequence a splice site or untranslated region. This large number of possible causes of allele specific expression could explain why we see so many functions associated with the 555 genes showing allele specific expression (Table 4).

But it is not just allele specific expression that may have genetic as well as epigenetic effects. It has been shown in humans that some allele specific methylation is determined by DNA sequence in cis and therefore shows Mendelian inheritance patterns (Meaburn et al., 2010). An extreme example of genetically controlled allele specific methylation is found in *Nasonia* wasps, where there is no evidence for methylation driven allele specific expression but inheritable cis-mediated allele specific methylation has been found (Wang et al., 2016). This cis-mediated methylation has recently been suggested as being important in social insect biology (Remnant et al., 2016; Wedd et al., 2016).

We have found that allele specific expression is widespread in the bumblebee. We have also found that the extreme version of allele specific expression, monoalleic expression is associated with monoallelic methylation. Genomic imprinting in mammals usually involves monoallelic methylation and expression. Although tempting to associate our results with genomic imprinting, this current work is unable to identify genomic imprinting. In any case, caution should be applied due to the lack of understanding of the functional role of methylation in gene expression in insects and in the, as yet unquantified, role of genetic cis effects in insect allele specific methylation and expression.

ACKNOWLEDGEMENTS

This work was financially supported by NERC grant no. NE/H010408/1 and NE/N010019/1 and NERC Biomolecular Analysis Facility research grants (NBAF 606 and 829) to EBM. Illumina library preparation, sequencing and bioinformatics were carried out by Edinburgh Genomics, The University of Edinburgh. Edinburgh Genomics is partly supported through core grants from NERC (R8/H10/56), MRC (MR/K001744/1) and BBSRC (BB/J004243/1). ZNL would like to thank UK BBSRC for its financial support via MIBTP. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/022657; this version posted August 11, 2017. 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 4.0 International license.

SUPPORTING INFORMATION LEGENDS

Table S1. Nineteen genes showing both monoallelic methylation and monoallelic expression. Blast results and genomic coordinates of the reads from the RNA-seq, MRE-seq and MeDip-seq libraries.

Table S2. Confirmation of single allele expression of nineteen monoallelically expressed genes in twenty-nine previously published transcriptomes. For each of the 19 contigs are the previously published RNA-seq libraries with associated read counts.

Table S3. 555 genes showing allele specific expression in at least three of the 29 previously published RNA-seq libraries. This table details the blast results from both the bumblebee and drosophila genomes and the GO terms associated with the drosophila hits.

ETHICAL DECLARATION

The protocol reported here conforms to the regulatory requirements for animal experimentation in the United Kingdom.

REFERENCES

Akhouayri, I., Turc, C., Royet, J., and Charroux, B. (2011). Toll-8/Tollo negatively regulates antimicrobial response in the Drosophila respiratory epithelium. *PLoS pathogens*, 7(10):e1002319.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25:3389–3402. 17.

Alvarado, S., Rajakumar, R., Abouheif, E., and Szyf, M. (2015). Epigenetic variation in the Egfr gene generates quantitative variation in a complex trait in ants. *Nat Commun*, 6.

Amarasinghe, H., Toghill, B., Nathanael, D., and Mallon, E. B. (2015). Allele specific expression in worker reproduction genes in the bumblebee Bombus terrestris. *PeerJ*, 3:e1079.

Amarasinghe, H. E., Clayton, C. I., and Mallon, E. B. (2014). Methylation and worker reproduction in the bumble-bee (Bombus terrestris). *Proceedings of the Royal Society B: Biological Sciences*, 281(1780):20132502.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes & Development*, 16(1):6–21.

Bonasio, R., Li, Q., Lian, J., Mutti, N. S., Jin, L., Zhao, H., Zhang, P., Wen, P., Xiang, H., Ding, Y., Jin, Z., Shen, S. S., Wang, Z., Wang, W., Wang, J., Berger, S. L., Liebig, J., Zhang, G., and Reinberg, D. (2012). Genome-wide and Caste-Specific DNA Methylomes of the Ants Camponotus floridanus and Harpegnathos saltator. *Current Biology*, 22(19):1755–1764.

Cardoen, D., Wenseleers, T., Ernst, U. R., Danneels, E. L., Laget, D., DE Graaf, D. C., Schoofs, L., and Verleyen, P. (2011). Genome-wide analysis of alternative reproductive phenotypes in honeybee workers. *Molecular Ecology*, 20(19):4070–4084.

Chamberlain, A. J., Vander Jagt, C. J., Hayes, B. J., Khansefid, M., Marett, L. C., Millen, C. A., Nguyen, T. T. T., and Goddard, M. E. (2015). Extensive variation between tissues in allele specific expression in an outbred mammal. *BMC Genomics*, 16.

Chittka, A., Wurm, Y., and Chittka, L. (2012). Epigenetics: The making of ant castes. *Current Biology*, 22(19):R835–R838.

core Team, R. (2016). R: A language and environment for statistical computing.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., and Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158.

Dean, W., Bowden, L., Aitchison, A., Klose, J., Moore, T., Meneses, J. J., Reik, W., and Feil, R. (1998). Altered imprinted gene methylation and expression in completely ES cell-derived mouse fetuses: association with aberrant phenotypes. *Development (Cambridge, England)*, 125(12):2273–2282.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., and Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5):491–498.

bioRxiv preprint doi: https://doi.org/10.1101/022657; this version posted August 11, 2017. 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 4.0 International license.

360 Dorfman, R., Glazer, L., Weihe, U., Wernet, M. F., and Shilo, B.-Z. (2002). Elbow and Noc define a
361 family of zinc finger proteins controlling morphogenesis of specific tracheal branches. *Development*
362 *(Cambridge, England)*, 129(15):3585–3596.

363 Drewell, R. A., Bush, E. C., Remnant, E. J., Wong, G. T., Beeler, S. M., Stringham, J. L., Lim, J., and
364 Oldroyd, B. P. (2014). The dynamic DNA methylation cycle from egg to sperm in the honey bee Apis
365 mellifera. *Development*, 141(13):2702–2711.

366 Dreyfuss, G., Matunis, M. J., Piñol-Roma, S., and Burd, C. G. (1993). hnRNP proteins and the biogenesis
367 of mRNA. *Annual Review of Biochemistry*, 62:289–321.

368 Drinkwater, N., Lee, J., Yang, W., Malcolm, T. R., and McGowan, S. (2017). M1 aminopeptidases as
369 drug targets: broad applications or therapeutic niche? *The FEBS journal*, 284(10):1473–1488.

370 Edsgard, D., Iglesias, M. J., Reilly, S.-J., Hamsten, A., Tornvall, P., Odeberg, J., and Emanuelsson,
371 O. (2016). GeneiASE: Detection of condition-dependent and static allele-specific expression from
372 RNA-seq data without haplotype information. *Scientific Reports*, 6:21134.

373 Farh, K. K.-H., Grimson, A., Jan, C., Lewis, B. P., Johnston, W. K., Lim, L. P., Burge, C. B., and Bartel,
374 D. P. (2005). The widespread impact of mammalian MicroRNAs on mRNA repression and evolution.
375 *Science (New York, N.Y.)*, 310(5755):1817–1821.

376 Galbraith, D. A., Kocher, S. D., Glenn, T., Albert, I., Hunt, G. J., Strassmann, J. E., Queller, D. C., and
377 Grozinger, C. M. (2016). Testing the kinship theory of intragenomic conflict in honey bees (Apis
378 mellifera). *Proceedings of the National Academy of Sciences*, page 201516636.

379 Gineikiene, E., Stoskus, M., and Griskevicius, L. (2009). Single Nucleotide Polymorphism-Based System
380 Improves the Applicability of Quantitative PCR for Chimerism Monitoring. *The Journal of Molecular*
381 *Diagnostics : JMD*, 11(1):66–74.

382 Glastad, K. M., Hunt, B. G., and Goodisman, M. A. (2014). Evolutionary insights into DNA methylation
383 in insects. *Current Opinion in Insect Science*, 1:25–30.

384 Glastad, K. M., Hunt, B. G., Yi, S. V., and Goodisman, M. a. D. (2011). DNA methylation in insects: on
385 the brink of the epigenomic era. *Insect Molecular Biology*, 20(5):553–565.

386 Goldberg, A., Allis, C., and Bernstein, E. (2007). Epigenetics: A Landscape Takes Shape. *Cell*,
387 128(4):635–638.

388 Gotz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón,
389 M., Dopazo, J., and Conesa, A. (2008). High-throughput functional annotation and data mining with
390 the Blast2go suite. *Nucleic Acids Research*, 36(10):3420–3435.

391 Harris, R. A., Wang, T., Coarfa, C., Nagarajan, R. P., Hong, C., Downey, S. L., Johnson, B. E., Fouse,
392 S. D., Delaney, A., Zhao, Y., Olshen, A., Ballinger, T., Zhou, X., Forsberg, K. J., Gu, J., Echipare, L.,
393 O'Geen, H., Lister, R., Pelizzola, M., Xi, Y., Epstein, C. B., Bernstein, B. E., Hawkins, R. D., Ren, B.,
394 Chung, W.-Y., Gu, H., Bock, C., Gnirke, A., Zhang, M. Q., Haussler, D., Ecker, J. R., Li, W., Farnham,
395 P. J., Waterland, R. A., Meissner, A., Marra, M. A., Hirst, M., Milosavljevic, A., and Costello, J. F.
396 (2010). Comparison of sequencing-based methods to profile DNA methylation and identification of
397 monoallelic epigenetic modifications. *Nature biotechnology*, 28(10):1097–1105.

398 Harrison, M. C., Hammond, R. L., and Mallon, E. B. (2015). Reproductive workers show queen-like
399 gene expression in an intermediately eusocial insect, the buff-tailed bumble bee Bombus terrestris.
400 *Molecular Ecology*, 24:121–129.

401 Harvey, C. T., Moyerbrailean, G. A., Davis, G. O., Wen, X., Luca, F., and Pique-Regi, R. (2015). QuASAR:
402 quantitative allele-specific analysis of reads. *Bioinformatics*, 31(8):1235–1242.

403 Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data
404 using empirical Bayes methods. *Biostatistics*, 8(1):118–127.

405 Kelley, K., Miller, K. R., Todd, A., Kelley, A., Tuttle, R., and Berberich, S. J. (2010). YPEL3, a
406 p53-regulated gene that induces cellular senescence. *Cancer research*, 70(9):3566–3575.

407 Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate
408 alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*,
409 14(4):R36.

410 Kocher, S. D., Tsuruda, J. M., Gibson, J. D., Emore, C. M., Arechavaleta-Velasco, M. E., Queller, D. C.,
411 Strassmann, J. E., Grozinger, C. M., Gribskov, M. R., San Miguel, P., Westerman, R., and Hunt, G. J.
412 (2015). A Search for Parent-of-Origin Effects on Honey Bee Gene Expression. *G3 (Bethesda, Md.)*.

413 Kodjabachian, L., Delaage, M., Maurel, C., Miassod, R., Jacq, B., and Rosset, R. (1998). Mutations in
414 ccf, a novel Drosophila gene encoding a chromosomal factor, affect progression through mitosis and

bioRxiv preprint doi: https://doi.org/10.1101/022657; this version posted August 11, 2017. 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 4.0 International license.

415 interact with Pc-G mutations. *The EMBO Journal*, 17(4):1063–1075.
416 Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The sva package for
417 removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*,
418 28(6):882–883.
419 Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform.
420 *Bioinformatics (Oxford, England)*, 25(14):1754–1760.
421 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.,
422 and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format
423 and SAMtools. *Bioinformatics (Oxford, England)*, 25(16):2078–2079.
424 Liu, W. and Saint, D. A. (2002). A New Quantitative Method of Real Time Reverse Transcription
425 Polymerase Chain Reaction Assay Based on Simulation of Polymerase Chain Reaction Kinetics.
426 *Analytical Biochemistry*, 302(1):52–59.
427 Magnani, R., Dirk, L. M. A., Trievel, R. C., and Houtz, R. L. (2010). Calmodulin methyltransferase is an
428 evolutionarily conserved enzyme that trimethylates Lys-115 in calmodulin. *Nature Communications*,
429 1:43.
430 Markesich, D. C., Gajewski, K. M., Nazimiec, M. E., and Beckingham, K. (2000). bicaudal encodes the
431 Drosophila beta NAC homolog, a component of the ribosomal translational machinery*. *Development*
432 *(Cambridge, England)*, 127(3):559–572.
433 McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012). Differential expression analysis of multifactor
434 RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 40(10):4288–4297.
435 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K.,
436 Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010). The Genome Analysis Toolkit:
437 A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*,
438 20(9):1297–1303.
439 Meaburn, E. L., Schalkwyk, L. C., and Mill, J. (2010). Allele-specific methylation in the human genome:
440 implications for genetic studies of complex disease. *Epigenetics*, 5(7):578–582.
441 Onyango, P., Jiang, S., Uejima, H., Shamblott, M. J., Gearhart, J. D., Cui, H., and Feinberg, A. P. (2002).
442 Monoallelic expression and methylation of imprinted genes in human and mouse embryonic germ
443 cell lineages. *Proceedings of the National Academy of Sciences of the United States of America*,
444 99(16):10599–10604.
445 Palacios, R., Gazave, E., Goñi, J., Piedrafita, G., Fernando, O., Navarro, A., and Villoslada, P. (2009).
446 Allele-Specific Gene Expression Is Widespread Across the Genome and Biological Processes. *PLoS*
447 *ONE*, 4(1):e4150.
448 Pardo-Manuel de Villena, F., de la Casa-Esperón, E., and Sapienza, C. (2000). Natural selection and the
449 function of genome imprinting: beyond the silenced minority. *Trends in genetics: TIG*, 16(12):573–579.
450 Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic*
451 *Acids Research*, 29(9):e45–e45.
452 Pignatta, D., Erdmann, R. M., Scheer, E., Picard, C. L., Bell, G. W., and Gehring, M. (2014). Natu-
453 ral epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *eLife*,
454 3:e03198.
455 Queller, D. C. (2003). Theory of genomic imprinting conflict in social insects. *Bmc Evolutionary Biology*,
456 3:art. no.–15.
457 Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic
458 features. *Bioinformatics (Oxford, England)*, 26(6):841–842.
459 Reik, W. and Walter, J. (2001). Genomic imprinting: Parental influence on the genome. *Nature Reviews*
460 *Genetics*, 2:21–32. 1.
461 Remnant, E. J., Ashe, A., Young, P. E., Buchmann, G., Beekman, M., Allsopp, M. H., Suter, C. M.,
462 Drewell, R. A., and Oldroyd, B. P. (2016). Parent-of-origin effects on genome-wide DNA methylation
463 in the Cape honey bee (Apis mellifera capensis) may be confounded by allele-specific methylation.
464 *BMC Genomics*, 17(1).
465 Riddell, C. E., Garces, J. D. L., Adams, S., Barribeau, S. M., Twell, D., and Mallon, E. B. (2014).
466 Differential gene expression and alternative splicing in insect immune specificity. *BMC Genomics*,
467 15(1):1031.
468 Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for
469 differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140.

bioRxiv preprint doi: https://doi.org/10.1101/022657; this version posted August 11, 2017. 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 4.0 International license.

470 Sadd, B. M., Barribeau, S. M., Bloch, G., Graaf, D. C. d., Dearden, P., Elsik, C. G., Gadau, J., Grimmelikhuijzen, C. J., Hasselmann, M., Lozier, J. D., Robertson, H. M., Smagghe, G., Stolle, E., Vaerenbergh, M. V., Waterhouse, R. M., Bornberg-Bauer, E., Klasberg, S., Bennett, A. K., Câmara, F., Guigó, R., Hoff, K., Mariotti, M., Munoz-Torres, M., Murphy, T., Santesmasses, D., Amdam, G. V., Beckers, M., Beye, M., Biewer, M., Bitondi, M. M., Blaxter, M. L., Bourke, A. F., Brown, M. J., Buechel, S. D., Cameron, R., Cappelle, K., Carolan, J. C., Christiaens, O., Ciborowski, K. L., Clarke, D. F., Colgan, T. J., Collins, D. H., Cridge, A. G., Dalmay, T., Dreier, S., Plessis, L. d., Duncan, E., Erler, S., Evans, J., Falcon, T., Flores, K., Freitas, F. C., Fuchikawa, T., Gempe, T., Hartfelder, K., Hauser, F., Helbing, S., Humann, F. C., Irvine, F., Jermiin, L. S., Johnson, C. E., Johnson, R. M., Jones, A. K., Kadowaki, T., Kidner, J. H., Koch, V., Köhler, A., Kraus, F. B., Lattorff, H. M., Leask, M., Lockett, G. A., Mallon, E. B., Antonio, D. S. M., Marxer, M., Meeus, I., Moritz, R. F., Nair, A., Näpflin, K., Nissen, I., Niu, J., Nunes, F. M., Oakeshott, J. G., Osborne, A., Otte, M., Pinheiro, D. G., Rossié, N., Rueppell, O., Santos, C. G., Schmid-Hempel, R., Schmitt, B. D., Schulte, C., Simões, Z. L., Soares, M. P., Swevers, L., Winnebeck, E. C., Wolschin, F., Yu, N., Zdobnov, E. M., Aqrawi, P. K., Blankenburg, K. P., Coyle, M., Francisco, L., Hernandez, A. G., Holder, M., Hudson, M. E., Jackson, L., Jayaseelan, J., Joshi, V., Kovar, C., Lee, S. L., Mata, R., Mathew, T., Newsham, I. F., Ngo, R., Okwuonu, G., Pham, C., Pu, L.-L., Saada, N., Santibanez, J., Simmons, D., Thornton, R., Venkat, A., Walden, K. K., Wu, Y.-Q., Debyser, G., Devreese, B., Asher, C., Blommaert, J., Chipman, A. D., Chittka, L., Fouks, B., Liu, J., O'Neill, M. P., Sumner, S., Puiu, D., Qu, J., Salzberg, S. L., Scherer, S. E., Muzny, D. M., Richards, S., Robinson, G. E., Gibbs, R. A., Schmid-Hempel, P., and Worley, K. C. (2015). The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biology*, 16(1):76.

491 Sapienza, C. (2002). Imprinted gene expression, transplantation medicine, and the "other" human embryonic stem cell. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16):10243–10245.

494 Scharf, M. E., Wu-Scharf, D., Pittendrigh, B. R., and Bennett, G. W. (2003). Caste and development-associated gene expression in a lower termite. *Genome Biology*, 4(10):R62.

496 Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE*, 6(7):e21800.

498 Toth, A. L., Tooker, J. F., Radhakrishnan, S., Minard, R., Henshaw, M. T., and Grozinger, C. M. (2014). Shared genes related to aggression, rather than chemical communication, are associated with reproductive dominance in paper wasps (Polistes metricus). *BMC Genomics*, 15(1):75.

501 Wang, X., Werren, J. H., and Clark, A. G. (2016). Allele-Specific Transcriptome and Methylome Analysis Reveals Stable Inheritance and Cis -Regulation of DNA Methylation in Nasonia. *PLOS Biol*, 14(7):e1002500.

504 Wedd, L., Kucharski, R., and Maleszka, R. (2016). Differentially methylated obligatory epialleles modulate context-dependent LAM gene expression in the honeybee Apis mellifera. *Epigenetics*, 11(1):1–10.

507 Weiner, S. A. and Toth, A. L. (2012). Epigenetics in social insects: a new direction for understanding the evolution of castes. *Genetics research international*, 2012:609810.

509 Welch, M. and Lister, R. (2014). Epigenomics and the control of fate, form and function in social insects. *Current Opinion in Insect Science*, 1:31–38.

511 Yan, H., Bonasio, R., Simola, D. F., Liebig, J., Berger, S. L., and Reinberg, D. (2015). DNA Methylation in Social Insects: How Epigenetics Can Control Behavior and Longevity. *Annual Review of Entomology*, 60(1):435–452.

514 Yan, H., Simola, D. F., Bonasio, R., Liebig, J., Berger, S. L., and Reinberg, D. (2014). Eusocial insects as emerging models for behavioural epigenetics. *Nature Reviews Genetics*, 15(10):677–688.

516 Zhang, Y., Shields, T., Crenshaw, T., Hao, Y., Moulton, T., and Tycko, B. (1993). Imprinting of human H19: allele-specific CpG methylation, loss of the active allele in Wilms tumor, and potential for somatic allele switching. *American Journal of Human Genetics*, 53(1):113–124.