

# SCIENTIFIC REPORTS



OPEN

## EGFR gene methylation is not involved in Royalactin controlled phenotypic polymorphism in honey bees

R. Kucharski, S. Foret & R. Maleszka

Received: 18 June 2015  
Accepted: 17 August 2015  
Published: 11 September 2015

The 2011 highly publicised *Nature* paper by Kamakura on honeybee phenotypic dimorphism, (also using *Drosophila* as an experimental surrogate), claims that a single protein in royal jelly, Royalactin, essentially acts as a master “on-off” switch in development via the epidermal growth factor receptor (AmEGFR), to seal the fate of queen or worker. One mechanism proposed in that study as important for the action of Royalactin is differential *amegfr* methylation in alternate organismal outcomes. According to the author differential methylation of *amegfr* was experimentally confirmed and shown in a supportive figure. Here we have conducted an extensive analysis of the honeybee *egfr* locus and show that this gene is never methylated. We discuss several lines of evidence casting serious doubts on the *amegfr* methylation result in the 2011 paper and consider possible origins of the author’s statement. In a broader context, we discuss the implication of our findings for contrasting context-dependent regulation of EGFR in three insect species, *Apis mellifera*, *D. melanogaster* and the carpenter ant, *Camponotus floridanus*, and argue that more adequate methylation data scrutiny measures are needed to avoid unwarranted conclusions.

In the social honey bee *Apis mellifera*, two alternative female phenotypes, long-lived fertile queens and short-lived sterile workers are produced via differential feeding with a diet known as royal jelly (RJ)<sup>1–3</sup>. This complex, still poorly understood nutrition contains various ingredients<sup>4,5</sup> including carbohydrates, vitamins, unusual lipids, antimicrobial agents, epigenomic modifiers such as histone deacetylases inhibitors (HDACs)<sup>6</sup>, as well as many other less characterised compounds<sup>4,5</sup>. The bulk of RJ is formed by several Major Royal Jelly Proteins (MRJPs) that appear to be unique to Hymenoptera<sup>2</sup>. MRJPs evolved from the insect Yellow protein family that has its origins in bacteria<sup>2,7</sup>. No relatives of MRJPs or Yellow proteins have been found in modern vertebrates, but a Yellow-like protein is encoded by the genome of a chordate *Branchiostoma floridae* (GenBank XP\_002607604). The remarkable developmental potency of RJ has been attributed to a synergistic effect of many if not all of its components acting as activators of signalling pathways via threshold based changes in metabolic flux and epigenomic modifications<sup>8–11</sup>. This view has been somewhat obscured by the 2011 study claiming that one of the MRJPs, labelled Royalactin, is capable on its own to drive all the changes needed to make a queen bee<sup>12</sup> effectively reducing the entire process of queen development to the vagaries of one protein. Furthermore, treating *Drosophila* with Royalactin appears to increase body size and ovary development in female flies with the Canton S genetic background<sup>12</sup>. In order to reconcile those findings with prior evidence implicating DNA methylation in queen development<sup>13,14</sup>, Kamakura conducted DNA methylation analysis in the genomic region encoding AmEGFR. In one of the Figures (S34)<sup>12</sup> he shows that “the overall level of methylation of *amegfr* in larvae reared with RJ (5%), which develop into queens, was decreased as compared with that in larvae reared with 40-30d RJ (57%), which emerge as workers. Similar results were observed in queen larvae and

Research School of Biology, The Australian National University, Canberra ACT 0200, Australia. Correspondence and requests for materials should be addressed to R.M. (email: ryszard.maleszka@anu.edu.au)

Dataset	Result	EGFR	NADRIN
Queen and worker larval heads <sup>9</sup>	mean coverage*	34.5	207.6
	methylation	0%	43.6%
Queen and worker adult brains <sup>15</sup>	mean coverage	22.3	16.2
	methylation	0%	40.9%
Haploid and diploid embryos (various stages of development)**	mean coverage	16.1	23.3
	methylation	0%	28.7%
Forager brains <sup>16</sup>	mean coverage	7.0	6.7
	methylation	0%	42.9%
Queen brains <sup>16</sup>	mean coverage	12.7	11.5
	methylation	0%	46.9%
Reverted nurses brains <sup>16</sup>	mean coverage	5.5	4.3
	methylation	0%	40.5%
Worker brains <sup>16</sup>	mean coverage	11.2	11.1
	methylation	0%	43.0%

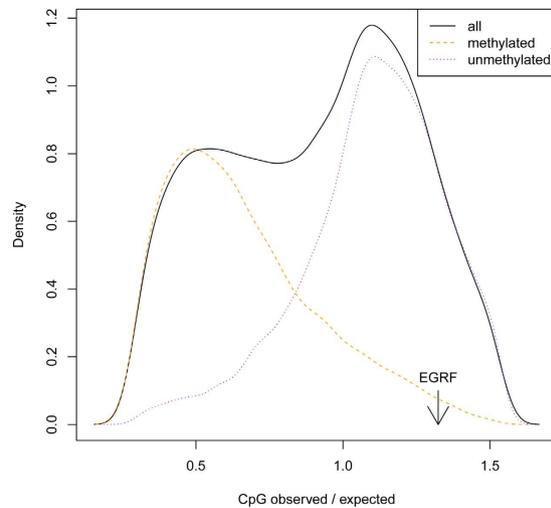
**Table 1. Sequencing coverage and *egfr/nadrin* methylation in the honey bee whole-genome bisulfite sequencing projects.** \*average from queen and worker libraries, \*\*in preparation.

worker larvae reared in hive<sup>12</sup>. Since this result stands in stark contrast to the absence of *amegfr* methylation in published honey bee methylomes<sup>9,15,16</sup>, including larvae of the same age as in the Kamakura paper<sup>12</sup>, we have conducted detailed analyses to re-examine the methylation status of *egfr* in *Apis mellifera*. We show not only that *amegfr* is never methylated but has a high GC content that is consistent with non-methylated genes, and consider the origins of this unfeasible result in the Kamakura study.

## Results

Our desire to meticulously examine the levels of *egfr* methylation in *Apis* was inspired by the absence of any methylated CpGs in this locus in the published methylomes from both queens and workers (Table 1). In particular, the larval methylomes<sup>9</sup> representing queen and workers of the same 96 hrs stage of development as in the Kamakura's paper<sup>12</sup> were indicative of a potential problem. In addition, no methylation of *amegfr* has been found in several brain methylomes generated by two independent labs<sup>15,16</sup>. For comparison, we have contrasted sequencing coverage and methylation levels for *egfr* with a consistently methylated gene *nadrin* using 11 methylomes representing different tissues, cell types and developmental stages (Table 1). We also have noted that the 736 bp region of *amegfr* deemed to be methylated in the Kamakura study<sup>12</sup> is rich in CpG dinucleotides, which is untypical for methylated genes in *Apis* that are CpG-depleted because of the known tendency of 5-methyl cytosines to be converted into thymines<sup>8,15,17,18</sup>. The ratio of observed to expected (o/e) CpG dinucleotides is 1.104 in the 736 bp fragment and 1.324 for the whole gene. Such values are associated with non-methylated genes in *Apis*<sup>8,15</sup> (Fig. 1) that can be easily separated from methylated genes in a bimodal distribution of the o/e frequency of CpGs in all annotated genes. However, given the low sequencing depth of genome-wide methylomes and the possibility that *amegfr* may be methylated in a restricted number of cell types that could have been picked by chance in the 2011 study, we have decided to reproduce the original low-depth plasmid sequencing experiment using ultra-deep sequencing of the same genomic region<sup>12</sup>. Unfortunately, we have failed to amplify the 736 bp DNA fragment following the conditions described in the 2011 paper. This is not entirely surprising as it is nearly impossible to amplify DNA fragments of such length from bisulfite treated DNAs. The DNA fragmentation during the treatment<sup>19</sup> leads to a practical upper size limit of the PCR amplicon ~400–500 bp, and the need for nested primers and a second round of PCR is often essential<sup>20</sup>. Instead, we used two sets of outer, and two sets of nested primers to amplify two overlapping amplicons (A1 = 440 bp and A2 = 441 bp) covering the supposedly methylated region of *amegfr*. To properly design our amplicons we had to re-annotate the old erroneous EGFR gene model used in the 2011 study (see Figure S1 for details). These amplicons have been processed and barcoded for ultra-deep sequencing with the NEB kit for Illumina libraries. Both libraries A1 and A2 have been sequenced together with additional libraries representing confirmed methylated genes on Illumina MiSeq using the Reagent Kit v3 that generates up to 25 million reads (15Gb) of 2 × 300 bp paired end reads.

As shown in Table 2 even with a depth of over 330,000 reads per amplicon virtually no methylation of EGFR in larval samples has been detected. The methylation pattern counts were normalised with MPFE<sup>21</sup> in order to eliminate most of the spurious patterns caused by sequencing errors and incomplete bisulfite conversion. As a result, for both *amegfr* amplicons A1 and A2, over 99.99% of reads were classified as not methylated. The very few methylated patterns observed in *egfr* are most likely technical noise and



**Figure 1. CpG (observed/expected) bias of protein-coding regions in the honey bee genome.** The o/e value of 1.104 for the analysed *egfr* region is indicated with the arrow. The o/e value for the whole gene is even higher (1.324).

Dataset		Amplicon 1	Amplicon 2
96 hrs Queen larvae	Reads per amplicon	1,873	339,596
	methylation	0%	<0.01%
96 hrs Worker larvae	Reads per amplicon	4,168	334,837
	methylation	0%	<0.01%

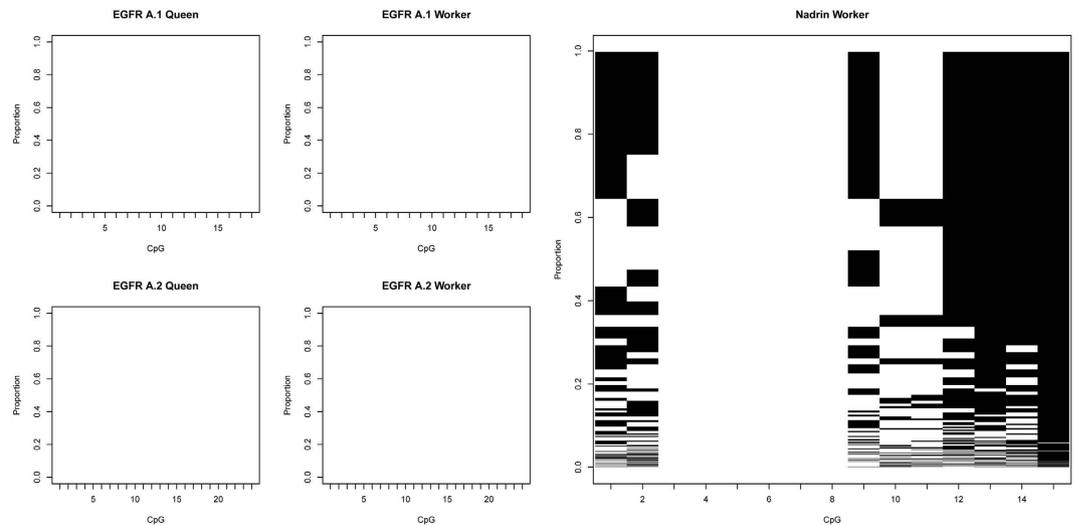
**Table 2. Sequencing coverage and *egfr* methylation level in MiSeq high-throughput amplicon experiment.**

represent less than 0.01% of the total counts. Failure to detect methylation in the *egfr* region is unlikely to be a PCR artefact because methylated DNA tends to be over-represented in bisulfite sequencing data<sup>22</sup>. With this depth of sequencing it is likely the methylation patterns in essentially every cell type can be visualised as illustrated by a genuinely methylated gene, *nadrin* (Fig. 2). In larval samples, 157 distinct methylation patterns were observed in this amplicon after removing the spurious patterns. Whether or not these patterns represent unique epigenetic signatures of 157 cell types in growing larvae remains to be established.

## Discussion

Our findings showing that the honey bee *egfr* gene belongs to the non-methylated category make the result shown in the Kamakura paper<sup>12</sup> impossible to reproduce. There are three lines of evidence questioning the correctness of the original claim. First, on the basis of its high CG content, *egfr* is bioinformatically predicted to be non-methylated. Second, genome-wide methylation profiles in larvae, brains and embryos from both queens and workers show no evidence of methylation. Third, our new data using ultra-deep amplicon sequencing failed to detect any sign of methylated cytosines in larval samples. At this stage the source of this questionable result is unclear. One possibility is that the original DNA samples were not properly converted with bisulfite thus leading to a false impression that some cytosines in *egfr* were protected by methyl tag. This problem could have been exaggerated by omitting nested primers that typically are used to improve the recovery of AT-rich methylated amplicons from BS-treated DNA. However, the increase of *amegfr* methylation only in worker larvae in two situations shown in the Kamakura paper (Figure S34) suggests that such an experimental clarification is unlikely. Whatever the reason for this doubtful result is, our findings have important consequences for understanding how conditional phenotypes are implemented in various lineages by engaging cell surface signalling via EGFR and its growth factor ligands.

While it is not surprising that this important cell surface receptor has been implicated in the queen crafting process in honey bees and growth regulation in *Drosophila*, it is evident that its regulation is not contingent on DNA methylation. Although a queen bee can be experimentally induced by silencing *de novo* DNA methylation in newly hatched larvae by means of RNAi approach<sup>13</sup>, such treatment



**Figure 2. Methylation patterns in *egfr* and *nadrin* revealed by deep amplicon sequencing.** Each row represents a methylation pattern (black: methylated CpGs, white: not methylated CpGs), the height of each pattern is proportional to the pattern's abundance. Two EGFR amplicons (A1 and A2) were amplified from 96hrs old queen and worker larvae. One *nadrin* amplicon was amplified from 96hrs worker larvae that have been shown in the Kamakura paper to have elevated methylation levels. After normalising pattern frequencies using MPFE<sup>21</sup>, no methylation was detected in *egfr*, whereas several distinct and highly abundant methylation patterns are detected in the *nadrin* amplicon. The pattern proportions are sorted from the most abundant at the top to the least abundant at the bottom.

is pleiotropic and affects hundreds of genes and pathways relevant to nutritional sensing, such as for example, the TOR/insulin network, and importantly, Anaplastic Lymphoma Kinase that has the capacity to directly induce downstream events by bypassing insulin signalling<sup>9,23</sup>. A queen phenotype can also be induced by interfering with the expression of genes belonging to the TOR/insulin network<sup>24</sup> and possibly other manipulations affecting nutritional signalling. For example, royal jelly is exceedingly rich in both methionine and sources of methyl groups (choline) and some of its MRJPs are unusually rich in this essential amino acid, providing substrates for methylation activities. In this context, the finding by Grandison *et al.*<sup>25</sup> that in *Drosophila*, methionine alone was necessary and sufficient to increase fecundity as much as did full feeding, but without reducing lifespan, is striking. Although the effects of RJ on honey bee larval growth are still not fully appreciated, there is little doubt that methylation changes and diet are clearly linked. It may be prudent to more closely examine the relationships between caloric restriction, methylation and foods rich in methionine, acetylcholine with longevity and fecundity in both *Apis* and *Drosophila*. Royalactin is simply one of very many components that contribute to network flux. Obviously, it has a defined role in this process, but until all components of RJ are better understood its exclusive role *in vivo* should be considered with caution.

It is most interesting that *egfr* methylation has recently been implicated in generating quantitative variation in size of the carpenter ant<sup>26</sup> allowing for comparative analyses of a highly conserved EGFR in three insect species in the epigenetic context. *Drosophila* lost its DNA methyltransferases and has to use other epigenetic mechanisms for EGFR regulation, whereas in two Hymenopterans with the complete DNA methylation toolkit, ants and honey bees, only one recruits methylation to regulate *egfr*. One possible explanation for such a contrast between ants and honey bees is that methylation in the carpenter ant is utilised as an indirect modulator of *egfr* for continuous size variations, whereas in *Apis* such a flexible epigenomic modification of *egfr* would not be desirable for the proper development of the focal individual that is critical for the colony. These comparative analyses underscore the inherent pitfalls in data transferability between different species at particular levels. In spite of a high level of EGFR conservation, both structural and functional, its context-dependent regulation appears to be driven by distinct mechanisms in insect species. An organism can utilise many different cellular systems to accomplish the same end result as long as it is the desired phenotypic outcome. Indeed, to solve the same problem many different designs can be constructed even from heteromorphic, but functionally similar elements because of the high level of degeneracy in biology<sup>27</sup>.

In conclusion, we argue that methylation data claimed to be relevant to specific biological processes need to be supported by in-depth analyses using newer, high resolution methods. However, whatever platform one utilizes for methylomic insights, all of them have to be properly scrutinised to answer the critical question: how relevant is the observed differential methylation to the phenomenon under investigation?

## Materials and Methods

**DNA bisulfite conversion and amplicon preparation.** 1.5 µg of larval genomic DNA<sup>28</sup> was bisulfite converted using the QIAGEN Epitect<sup>®</sup> Bisulfite Kit, as per the manufacturer's protocol<sup>29</sup>. We routinely perform two consecutive treatments to avoid incomplete conversion especially in GC-rich regions. The converted DNA was amplified via a nested PCR reaction with *amegfr* specific primers (see Table S2 for BS-seq primers). The PCR products were purified utilising Agencourt<sup>®</sup> AMPure<sup>®</sup> XP PCR Purification system (Beckman Coulter).

**NGS library preparation.** Libraries were prepared from 500–600 ng of each amplicon utilising the NEBNext<sup>®</sup> DNA Library Prep Master Mix for Illumina<sup>®</sup>, and NEBNext<sup>®</sup> Multiplex Oligos for Illumina<sup>®</sup> Index Primers Set 1 and Set2 (New England Biolabs). Size selection of adaptor ligated DNA was performed using Agencourt AMPure XP beads (Beckman Coulter), with the bead:DNA ratio of the first bead selection 0.9X, followed by a second bead selection with bead:DNA ratio at 0.2X (not sure about the ratios, I do not think this is necessary). Each library was eluted in 30 µL of 0.1X TE, library size confirmed via agarose gel electrophoresis (we used Caliper LabChip GXII and HT DNA High Sensitivity Assay), and diluted to a final concentration of 4 nM.

**NGS MiSeq sequencing.** Next generation sequencing was performed on Illumina MiSeq instrument using MiSeq Reagent Kit v3 (Illumina) and 600 cycles. PhiX spike was added at 5% concentration as recommended by Illumina for low-diversity libraries.

**Analysis of bs-seq results.** For each individual analysed the frequency at which a mCpG occurred was calculated across all reads using custom Python scripts and open-source software. The process comprised of two steps. In the first, pairs of reads with the 30 nucleotide sequence starting at position 4 matching exactly the last 30 nucleotides of the primers used for nested amplicon PCR were extracted from FASTQ files, aligned with *in silico* bisulfite-converted genomic template using MUSCLE<sup>29</sup>, overlapping regions (if any) were proportionally truncated and, after removing all aligner-introduced gaps, both reads were combined into one continuous sequence and appended to a separate file ("extract file") for each amplicon and each library/sample. In addition, a quality filter was applied, rejecting all sequences shorter than 90% of the length of the template or containing in excess of 5% gaps. In the second step, batches of sequences from the "extract" files were re-aligned with the template using MUSCLE (to eliminate any potential positional errors introduced by read indels), the aligned template sequence was used to calculate positional information of all the expected CpGs and SNPs, and the positional data were used to score methylation status [ie. 0 for T and 1 for C occurring at a CpG position] for each combined read pair. The data were next appended to a separate table for each amplicon and each library/sample. The final tables were used to calculate and graph amplicon methylation data using MPFE<sup>21</sup>.

## References

- 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).
- Drapeau, M. D., Albert, S., Kucharski, R., Prusko, C. & Maleszka, R. Evolution of the Yellow/Major Royal Jelly Protein family and the emergence of social behavior in honey bees. *Genome Res* **16**, 1385–1394 (2006).
- Weaver, N. Rearing of Honeybee Larvae on Royal Jelly in the Laboratory. *Science* **121**, 509–510 (1955).
- Shuel, R. W. & Dixon, S. E. Studies in the mode of action of royal jelly in honeybee development II. Respiration of newly emerged larvae on various substrates. *Can J Zool* **37**, 803–813 (1959).
- Rembold, H. & Dietz, A. Biologically active substances in royal jelly. *Vitamins & Hormones* **23**, 359–382 (1965).
- Spannhoff, A. *et al.* Histone deacetylase inhibitor activity in royal jelly might facilitate caste switching in bees. *Embo Reports* **12**, 238–243 (2011).
- Maleszka, R. & Kucharski, R. Analysis of Drosophila yellow-B cDNA reveals a new family of proteins related to the royal jelly proteins in the honeybee and to an orphan protein in an unusual bacterium *Deinococcus radiodurans*. *Biochem Biophys Res Commun* **270**, 773–776 (2000).
- 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).
- Foret, S. *et al.* DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *P Natl Acad Sci USA* **109**, 4968–4973 (2012).
- Metallo, C. M. & Vander Heiden, M. G. Metabolism strikes back: metabolic flux regulates cell signaling. *Gene Dev* **24**, 2717–2722 (2010).
- Wellen, K. E. *et al.* ATP-Citrate Lyase Links Cellular Metabolism to Histone Acetylation. *Science* **324**, 1076–1080 (2009).
- Kamakura, M. Royalactin induces queen differentiation in honeybees. *Nature* **473**, 478–483 (2011).
- Kucharski, R., Maleszka, J., Foret, S. & Maleszka, R. Nutritional control of reproductive status in honeybees via DNA methylation. *Science* **319**, 1827–1830 (2008).
- Lyko, F. & Maleszka, R. Insects as innovative models for functional studies of DNA methylation. *Trends Genet* **27**, 127–131 (2011).
- Lyko, F. *et al.* The Honey Bee Epigenomes: Differential Methylation of Brain DNA in Queens and Workers. *Plos Biol* **8**, e1000506 (2010).
- Herb, B. R. *et al.* Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nature Neuroscience* **15**, 1371–1373 (2012).
- Miklos, G. L. G. & Maleszka, R. Epigenomic communication systems in humans and honey bees: From molecules to behavior. *Horm Behav* **59**, 399–406 (2011).

18. Wang, Y. & Leung, F. C. C. In Silico Prediction of Two Classes of Honeybee Genes with CpG Deficiency or CpG Enrichment and Sorting According to Gene Ontology Classes. *J Mol Evol* **68**, 700–705 (2009).
19. Tusnady, G. E., Simon, I., Varadi, A. & Aranyi, T. BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. *Nucleic Acids Research* **33**, e9 (2005).
20. Warnecke, P. M. *et al.* Identification and resolution of artifacts in bisulfite sequencing. *Methods* **27**, 101–107 (2002).
21. Lin, P., Foret, S., Wilson, S. R. & Burden, C. J. Estimation of the methylation pattern distribution from deep sequencing data. *BMC Bioinformatics* **16**, 145 (2015).
22. Ji, L. *et al.* Methylated DNA is over-represented in whole-genome bisulfite sequencing data. *Front Genet* **5**, 341 (2014).
23. Teleman, A. A. Privileged Signaling for Brain Growth. *Cell* **146**, 346–347 (2011).
24. Patel, A. *et al.* The Making of a Queen: TOR Pathway Is a Key Player in Diphenic Caste Development. *Plos One* **2**, e509 (2007).
25. Grandison, R. C., Piper, M. D. W. & Partridge, L. Amino-acid imbalance explains extension of lifespan by dietary restriction in *Drosophila*. *Nature* **462**, 1061–1064 (2009).
26. Alvarado, S., Rajakumar, R., Abouheif, E. & Szyf, M. Epigenetic variation in the *Egfr* gene generates quantitative variation in a complex trait in ants. *Nat Commun* **6**, 6513 (2015).
27. Maleszka, R., Mason, P. H. & Barron, A. B. Epigenomics and the concept of degeneracy in biological systems. *Brief Funct Genomics* **13**, 191–202 (2014).
28. Wojciechowski, M. *et al.* Insights into DNA hydroxymethylation in the honeybee from in-depth analyses of TET dioxygenase. *Open Biol* **4**, 140110 (2014).
29. Evans, J. D. *et al.* Standard methods for molecular research in *Apis mellifera*. *J Apicult Res* **52**, 1–30 (2013).

## Acknowledgements

This work was supported by the National Health and Medical Research Council project grant APP1050593 awarded to RM.

## 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:** Kucharski, R. *et al.* EGFR gene methylation is not involved in Royalactin controlled phenotypic polymorphism in honey bees. *Sci. Rep.* **5**, 14070; doi: 10.1038/srep14070 (2015).



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/>