Data Descriptor: Combinatory annotation of cell membrane receptors and signalling pathways of Bombyx mori prothoracic glands

Panagiotis Moulos1,2, Martina Samiotaki2, George Panayotou2 & Skarlatos G. Dedos3

The cells of prothoracic glands (PG) are the main site of synthesis and secretion of ecdysteroids, the biochemical products of cholesterol conversion to steroids that shape the morphogenic development of insects. Despite the availability of genome sequences from several insect species and the extensive knowledge of certain signalling pathways that underpin ecdysteroidogenesis, the spectrum of signalling molecules and ecdysteroidogenic cascades is still not fully comprehensive. To fill this gap and obtain the complete list of cell membrane receptors expressed in PG cells, we used combinatory bioinformatic, proteomic and transcriptomic analysis and quantitative PCR to annotate and determine the expression profiles of genes identified as putative cell membrane receptors of the model insect species, Bombyx mori, and subsequently enrich the repertoire of signalling pathways that are present in its PG cells. The genome annotation dataset we report here highlights modules and pathways that may be directly involved in ecdysteroidogenesis and aims to disseminate data and assist other researchers in the discovery of the role of such receptors and their ligands.

---

**Design Type(s)** | parallel group design
---|---
**Measurement Type(s)** | transcription profiling assay • protein expression profiling • genome assembly
**Technology Type(s)** | RNA-seq assay • mass spectrometry assay • digital curation
**Factor Type(s)** | life cycle stage • biological replicate role
**Sample Characteristic(s)** | Bombyx mori • prothoracic gland

---

1HybridStat Predictive Analytics, Aiolou 19, Athens 10551, Greece. 2Biomedical Sciences Research Center ‘Alexander Fleming’, Fleming 34, Vari 16672, Greece. 3Department of Biology, National and Kapodistrian University of Athens, Athens 15784, Greece. Correspondence and requests for materials should be addressed to S.G.D. (email: sdedos@biol.uoa.gr).
Background & Summary

Insect cell membrane receptors are excellent drug targets for pest control and management\(^1,2\). As such, an extensive amount of research has been carried out on various cell membrane receptors expressed in model insects, such as *Drosophila melanogaster*\(^3\) and *Bombyx mori*\(^4,5\), or insects that are major pests of crops or agents of animal and human diseases\(^6-5\). Identifying species-specific receptors expressed in cells crucial for the development of insects, such as the prothoracic gland (PG) cells, is even more important as these receptors can become targets for highly specialised, precise and species-specific drug discovery.

The PG cells are the main site of synthesis of ecdysteroids from cholesterol and their subsequent secretion\(^10,11\). Research efforts have shown that the signalling pathways that govern ecdysteroidogenesis are quite complex and involve a very broad array of second messengers and signalling modules with a high degree of overlapping integration and redundancy, albeit a large number of unidentified ligands, receptors and signalling components\(^10,12-16\). Much work on the ligands and receptors that stimulate or inhibit ecdysteroidogenesis has been carried out in model insects such as *Bombyx mori*\(^12,14,17\), whose extensive genome annotation is publicly available with genes mapped to chromosome and scaffold locations and the sequences of their coding protein(s)\(^18,19\).

In this study we further annotated the existing dataset of *Bombyx mori* genes\(^18,19\) (http://sgp.dna.affrc.go.jp/ComprehensiveGeneSet/) to include i) genes expressed in the PG cells identified by our proteomic (using liquid chromatography-tandem mass spectrometry (LC-MS/MS)) and transcriptomic (RNA-seq) datasets (Fig. 1) as described in a supporting paper\(^20\), 2) annotations of signalling pathways that are present in PG cells (Fig. 2), 3) annotations of *Bombyx mori* genes with the LocTree3\(^21\) protein subcellular localisation prediction dataset (Data Citation 1) and 4) annotation of the coding sequences of the proteins (Data Citation 1), by integrating results of our bioinformatics, LC-MS/MS and RNA-seq analyses from PG samples from day 0 (V\(^-0\)) and day 6 (V\(^-6\)) of the 5th instar (Data Citation 1) and further analysis of the expression patterns of these receptors by quantitative PCR (qPCR).

To implement our workflow (Fig. 1a), first we generated experimental data and analysis of LC-MS/MS datasets coupled with experimental data and analysis of RNA-seq datasets (Fig. 1 and Data Citation 1, Data Citation 2 and Data Citation 3). We next created a *Bombyx mori* reference genome dataset (Data Citation 4) to map and visualise the RNA-seq data and, in parallel, we generated literature-based and bioinformatics-assisted annotated lists of cell membrane receptors present in the genome of *Bombyx mori*. Next, we validated the expression of cell membrane receptors in PG cells through qPCR and clarified the presence of false negative and false positive hits in our LC-MS/MS and/or RNA-seq datasets (Fig. 1a and Tables 1 and 2). Critical to our approach was the visualization of RNA-seq data through the University of California, Santa Cruz (UCSC) Genome Browser\(^22\) track data hubs webpage, hosting the *Bombyx mori* reference genome (Data Citation 4), that allowed us to explain false positive and false negative hits (Tables 1 and 2) in our LC-MS/MS, RNA-seq and qPCR data.

All these datasets were integrated back into i) the annotated list of *Bombyx mori* genes that is publicly available (Fig. 1 and Data Citation 1) ii) the publicly available KEGG\(^23\) signalling pathways of *Bombyx mori* that we further annotated, visualized and present in Fig. 2 iii) the UCSC Genome Browser\(^22\) track data hubs webpage where results of the RNA-Seq samples analyses can be visualised by pasting the following link (http://epigenomics.fleming.gr/tracks/his_trackhubs/ekpa_dedos_2/hub.txt) and then pasting the coordinates of a gene of interest.

We provide a substantially increased list of *Bombyx mori* annotated genes (Data Citation 1) integrated with our LC-MS/MS and RNA-seq datasets (Fig. 1b) and other bioinformatic annotations of *Bombyx mori* genes (http://epigenomics.fleming.gr/metaseq_runs/HS/000004/\) and most importantly we identify and visually depict here (Fig. 2) the signalling pathways in PG cells reported in a supporting paper\(^20\). Our datasets can be valuable tools for researchers who want to study the role of signalling pathways in PG cells or conduct comparative studies on the presence of cell membrane receptors in other insect species. Additional information for a comprehensive understanding of the cell membrane receptors and the signalling pathways in PG cells is described in a related research paper\(^20\).

Methods

Animals

The hybrid J106xDAIZO of *Bombyx mori* was used in this study. In this hybrid, the 5th instar period lasts about \(\sim 208\) h, the onset of pupal commitment occurs after \(60\) h (day 3) and the onset of wandering behaviour occurs \(144\) h (day 6) after the final larval ecdysis. This hybrid has a short period of cocoon spinning that lasts \(\sim 38\) h followed by a period of \(\sim 26\) h before pupal metamorphosis. In this study, each day of the 5th (V) instar is designated with its numerical number (i.e., \(V = 0, V = 1\) etc.) while the first day of the pupal stage is designated as \(P = 0\). Larvae were reared on fresh mulberry leaves under a 12:12 L:D photoperiod at \(25 \pm 1\) °C and 60% relative humidity. Larvae were staged after every larval ecdysis, and the day of each ecdysis was designated as day 0. Since larvae mainly moult to the final (5th) instar during the scotophase, all larvae that ecysed during the scotophase were segregated immediately after the onset of photophase. This time was designated as 0 h of the 5th instar and 4 h later samples of prothoracic glands were taken (day 0 samples) while samples of prothoracic glands from day 6 were taken \(144\) h later, at the onset of wandering behaviour.
Figure 1. Schematic workflow of the experimental design. a: The figure shows the iterative approaches we followed to identify the cell membrane receptors expressed in prothoracic gland cells of *Bombyx mori*. False positive and false negative hits based on the liquid chromatography-tandem mass spectrometry (LC-MS/MS) and transcriptomic (RNA-seq) analyses were analysed by qPCR assays and further examined by visualization in the University of California, Santa Cruz (UCSC) Genome Browser. Abbreviations indicate the time of ecdysis (E) to the final larval stage, the time of head critical period (HCP; see text for details), the time of feeding cessation and onset of wandering (W) behaviour, the time of metamorphosis to pupa (P) and the time of initiation of apoptosis (A) by the prothoracic gland cells. b: Venn diagrams showing the results of liquid chromatography-tandem mass spectrometry (LC-MS/MS) and transcriptomic (RNA-seq) analyses of the 3 biological replicates from day 0 (V−0) and day 6 (V−6) from prothoracic gland cells of the 5th instar of *Bombyx mori*. 
Experimental design
The aim of our present study was to provide a thorough map of signalling cascades present in the PG cells of *Bombyx mori* during the crucial final larval stage and the onset of the pupal stage before these cells initiate apoptosis (Fig. 1a). We have chosen to analyse PG samples from day 0 and day 6 of the 5th instar of *Bombyx mori* because there are striking differences in the hormonal milieu in these two developmental time points. On day 0, PG cells secrete very low amounts of ecdysteroids while the juvenile hormone titre...
| No | Receptor Type          | public ID (gene) | chr start position | chr end position | Protein description                         | Identified by qPCR | Identified by RNA-seq | Identified by LC-MS/MS | rpgm Y-0 | rpgm Y-6   | False negative reason | Forward primer | Reverse primer |
|----|----------------------|-----------------|--------------------|-----------------|---------------------------------------------|-------------------|---------------------|----------------------|----------|----------|----------------------|----------------|----------------|
| 1  | Class A GPCR         | BMgn003863      | chr1               | 21644307        | Neuropeptide receptor A19                   | ✓                 | ✓                   | ✓                    | 0.004    | 0.005    | Expressed on days other than Y-0 / Y-6 | ATGAGGAGGGCAGTGGCAGG | CGCACTTGGTGGTCAGACAA |
| 2  | Class A GPCR         | BMgn010317      | chr7               | 498359          | Putative pronatal protein                    | ✓                 | ✓                   | ✓                    | 0.018    | 0.017    | Incorrect annotation | TGGAGTTCATGGAACATGGAC | TGGAGGATGTCAGGACAA |
| 3  | Class A GPCR         | BMgn009929      | chr10              | 11044278        | Serotonin receptor 4 (Ser-4)                | ✓                 | ✓                   | ✓                    | 0.019    | 0.003    | Expressed on days other than Y-0 / Y-6 | AGGAGGCGGAACAGACTGAGC | CTTGAGATGCAAACTGAGAC |
| 4  | Class A GPCR         | BMgn010118      | chr12              | 6248327         | Serotonin receptor 3A                      | ✓                 | ✓                   | ✓                    | 0.006    | 0.021    |                                      | AAACGTTGAGGTCAGGACCA | CTTGAGATGCAAACTGAGAC |
| 5  | Class A GPCR         | BMgn011934      | chr11              | 6698420         | Protein trapped in endolymph-1-like/Trophoblast receptor-like | ✓                 | ✓                   | ✓                    | 0.003    | 0.019    |                                      | CACTGGAGAGAGATGGTCAGG | AAGTATCGAGGAGTTTCAGG |
| 6  | Class A GPCR         | BMgn012114      | chr11              | 1717623         | Neuropeptide receptor A10                  | ✓                 | ✓                   | ✓                    | 0.019    | 0.020    |                                      | GAGAAGAGGGGAACAGATGAC | ATGGTGGAGCAAGTCGAGAC |
| 7  | Class A GPCR         | BMgn010412      | chr12              | 11868448        | Diisopod hormone receptor                   | ✓                 | ✓                   | ✓                    | 0.008    | 0.023    |                                      | TGGAGTTCATGGAACATGGAC | TGGTTGGAGGAGACATGGAC |
| 8  | Class A GPCR         | BMgn012899      | chr24              | 20352346        | RTK                                          | ✓                 | ✓                   | ✓                    | 0.021    | 0.006    |                                      | TATAGGACAGATGGTTCAGG | AATTGACAGGAGACATGGAC |
| 9  | Class A GPCR         | BMgn013547      | chr1               | 908702          | Neuropeptide receptor B4                    | ✓                 | ✓                   | ✓                    | 0.009    | 0.007    |                                      | TGGTGCTATATGTCAGGACAA | CGCTTGAGCGAGACATGGAC |
| 10 | Class A GPCR         | BMgn013928      | chr8               | 18707217        | Neuropeptide receptor B1                    | ✓                 | ✓                   | ✓                    | 0.006    | 0.009    |                                      | AACAGGATTGCAACAGGACCA | TGGTTGAGGAGACATGGAC |
| 11 | Class A GPCR         | BMgn012443      | chr9               | 9527811         | Neuropeptide receptor B5                    | ✓                 | ✓                   | ✓                    | 0.005    | 0.010    |                                      | AGGAGGACACAGAAACATGGAC | TCTACAGGAAGGAGACATGGAC |
| 12 | Other cell membrane receptors | BMgn010141 | chr12              | 3182218         | Pregp1 and adipokine receptor family member 3-like | ✓                 | ✓                   | ✓                    | 0.028    | 0.023    |                                      | GTGTGGCAATCAATGGATTCATGGAC | ACACTCAGGAGAGAATGGATTCATGGAC |
| 13 | RTK                  | BMgn009797      | chr3               | 5949013         | Receptor tyrosine-protein kinase B-like     | ✓                 | ✓                   | ✓                    | 0.005    | 0.002    |                                      | TGCTCGTGGGGTTGTTC | TTTATGGGGATAGAGAGATGG |
| 14 | RTK                  | BMgn015800      | chr4               | 20345640        | Vomus kinase receptor 2                      | ✓                 | ✓                   | ✓                    | 0.016    | 0.029    |                                      | AAAGGAAGTCCTGGAGGGAGAGGC | TGGAGGATGGGAGGACACAGAG |
| 15 | RTK                  | BMgn012597      | chr6               | 5719821         | Receptor tyrosine kinase (RTK) Drl-like 2   | ✓                 | ✓                   | ✓                    | 0.016    | 0.046    |                                      | GTAGATGCTGACGAAATGGATTCATGGAC | ACACTCAGGAGAGAATGGATTCATGGAC |
| 16 | RTK                  | BMgn009036      | chr9               | 64319           | Receptor tyrosine kinase Cad96Ca-like       | ✓                 | ✓                   | ✓                    | 0.008    | 0.006    |                                      | ACCTATGGGAGAGACATGGAC | TGGTGGAGGAGACATGGAC |
| 17 | RTK                  | BMgn013124      | chr22              | 10332260        | Neuregulin receptor kinase (NeuR/KcNp)      | ✓                 | ✓                   | ✓                    | 0.001    | 0.003    |                                      | TGGTGGACGAGTGTCAGGAC | TGGTCACAGGAGGATGGAGAGG |
| 18 | RTK                  | BMgn010169      | chr22              | 23167192        | Receptor tyrosine kinase Cad96Ca-like 2     | ✓                 | ✓                   | ✓                    | 0.005    | 0.009    | Expressed on days other than Y-0 / Y-6 | TGGTGGACGAGTGTCAGGAC | TGGTGGACGAGGATGGAGAGG |
| 19 | RTK                  | BMgn010489      | chr22              | 20340909        | Protein-encoding tyrosine kinase ROS-like    | ✓                 | ✓                   | ✓                    | 0.002    | 0.006    |                                      | TGTTGTGACGAGGAGACATGGAC | TGGTGGACGAGGATGGAGAGG |
| 20 | RTK                  | BMgn011533      | chr22              | 30100900        | Protein-encoding tyrosine kinase ROS-like    | ✓                 | ✓                   | ✓                    | 0.003    | 0.007    |                                      | TGTTGTGACGAGGAGACATGGAC | TGGTGGACGAGGATGGAGAGG |
| 21 | RTK                  | BMgn006801      | chr10              | 10388675        | Receptor type gastrin/cyclase S2-like isoform | ✓                 | ✓                   | ✓                    | 0.014    | 0.020    |                                      | AGGAGGAGGGAACTGTCGTTG | TGGAGTTCAGGACACATGGAGG |
| 22 | Other cell membrane receptors | BMgn012138 | chr1               | 3790955         | Fusedro homologue                           | ✓                 | ✓                   | ✓                    | 0.006    | 0.007    |                                      | GTGGACGACAGCGAGCGAGG | TTGGAGACATGCAGAGGAGG |
| 23 | Other cell membrane receptors | BMgn004245 | chr10              | 14374056        | Similar to lipocalin-1 interacting membrane receptor (lim) | ✓                 | ✓                   | ✓                    | 0.025    | 0.024    |                                      | AGGAGAGAATGGCAGTTGAGG | AGGAGCTTGAGGACATGGAG |
| 24 | Other cell membrane receptors | BMgn004372 | chr10              | 5639803         | Patched domain-containing protein 5,like     | ✓                 | ✓                   | ✓                    | 0.003    | 0.017    |                                      | AGGAGGAGAGAAATGGCAGTTG | ATGGAGTTCAGGACACATGGAC |
| 25 | Other cell membrane receptors | BMgn010723 | chr15              | 2701530         | Similar to Notch                           | ✓                 | ✓                   | ✓                    | 0.003    | 0.023    |                                      | ACTCATGGGAGATGTCAGGAC | ATGGAGTTCAGGACACATGGAC |
| 26 | Other cell membrane receptors | BMgn008052 | chr18              | 11886109        | Interference hedgehog                      | ✓                 | ✓                   | ✓                    | 0.017    | 0.013    |                                      | GAGAAGGGGAGACAGACATGGAC | TATGGGGAGAGAGAGACAGACATGGAC |

**Table 1.** False negative results of cell membrane receptors expressed in prothoracic glands cells of *Bombbyx mori*. The 29 genes presented in this table were identified by both our transcriptomic (RNA-seq) and proteomic (LC-MS/MS) analysis and conclusive evidence of their expression came from the quantitative PCR (qPCR) assays using the indicated primers. Abbreviations: GPCR: G protein-coupled receptor; RTK: receptor tyrosine kinase; RTSK: receptor serine/threonine kinase; RTGC: receptor type guanylate cyclase; Chr: Chromosome; rpgm: reads per gene model; V − 0: Day 0 of the 5th instar; V − 6: Day 6 of the 5th instar.
Table 2. False positive results of cell membrane receptors that are not expressed in prothoracic gland cells of *Bombyx mori*. The 19 genes presented in this table were identified by either our transcriptomic (RNA-seq) or proteomic (LC-MS/MS) analysis but the evidence of their expression were not supported by the quantitative PCR (qPCR) assays using the indicated primers. Critical to the assessment of these hits as false positives was the visualisation of the reads in the University of California, Santa Cruz (UCSC) Genome Browser.

| No | Receptor Type | public ID (gene) | chr ID |chr. start position | chr. end position | Protein description | Identified by qPCR | Identified by RNA-seq | rpmsg V−− | rpmsg V=− | False positive reason | Forward primer | Reverse primer |
|----|---------------|------------------|--------|-------------------|------------------|----------------------|-------------------|---------------------|-------------|-------------|-------------------|----------------|----------------|
| 1  | Class A GPCR  | BMgn012112       | chr2   | 17472801          | 17480580         | Neuropeptide receptor A2 | —                 | —                   | V\(\rightarrow\)V | 0.005       | 0            | Incorrect entry in Uniprot | AGCGAGAGAA ACAYTAAGAG | CTGCTGACTACG |
| 2  | Class A GPCR  | BMgn010694       | chr14  | 21820159          | 21873854         | Neuropeptide receptor A30 | —                 | —                   | V\(\rightarrow\)V | 0.001       | 0            | Incorrect entry in Uniprot | GTCGCGACAC ACAAGTAGACG | CTGCTGACTACG |
| 3  | Class A GPCR  | BMgn012244       | chr15  | 5602489           | 5602489          | Neuropeptide receptor A8-A | —                 | —                   | V\(\rightarrow\)V | 0.003       | 0.005        | Possible LC-MS/MS artifact | AGCAAGAGAA CGCATAGAAGA | GAAGGAGCG |
| 4  | Class A GPCR  | BMgn010142       | chr17  | 12086933          | 12126114         | Neuropeptide receptor A28ACP receptor | —                 | —                   | V\(\rightarrow\)V | 0.006       | 0.007        | Possible LC-MS/MS artifact | TGTCGGCTTG CTTTGTGTGC | TGGTCAAGAGG |
| 5  | Class A GPCR  | BMgn012559       | chr19  | 7510918           | 7518997          | Opsin | —                 | —                   | V\(\rightarrow\)V | 0.001       | 0.0006      | Possibly expressed in the 4th instar or LC-MS/MS artifact | GAGCAGCTTTA CGGAAACAGTG | AGCTGCGAGG |
| 6  | Class A GPCR  | BMgn017930       | chr21  | 2534865           | 2543865          | Serotonin receptor A3 (D1-like) | —                 | √                   | —           | 0.006       | 0.005        | Reads aligned in intron | ITCTGCGGG CGGCAATTTCG | GCAGCGACAT |
| 7  | Class A GPCR  | BMgn014420       | chr22  | 2103777           | 21152042         | Neuropeptide receptor A5 | —                 | √                   | —           | 0.177       | 0.180        | Reads aligned in intron | AGCCAGUCGG GGAATAATCG | CGAGAAGACG |
| 8  | Class B GPCR  | BMgn014256       | chr8   | 72619              | 12058            | Metabotropic glutamate receptor 2-like | —                 | —                   | V\(\rightarrow\)V | 0.0002      | 0.0004      | Possibly expressed in the 4th instar or LC-MS/MS artifact | YCTGCTGCTG TGAGATACAG | ACGTGGTAGC |
| 9  | Olfactory receptor | BMgn014699       | chr1   | 12377435          | 12385061         | Olfactory receptor (BmOr-1) | —                 | √                   | —           | 0.001       | 0.0024      | Reads aligned in intron | AGTCATCAG AAGCAGAACG | GCAAGGTAAGG |
| 10 | Olfactory receptor | BMgn017182       | chr15  | 11127835          | 1113624         | Olfactory receptor (BmOr-40) | —                 | —                   | V\(\rightarrow\)V | 0.01        | 0.01         | Possible LC-MS/MS artifact | GTTGGAACGC CGGAACTGTC | CCGGAACTAC |
| 11 | RTGC.       | BMgn005120       | chr2    | 19480001          | 19490229         | Glutamate cyclase | —                 | —                   | V\(\rightarrow\)V | 0.002       | 0.0007      | Possibly expressed in the 4th instar or LC-MS/MS artifact | ATCGTCGACG CGTATACGGT | CATGATTCGG |
| 12 | Insignin       | BMgn018454       | chr12  | 2561262           | 2582187         | Intergenic alpha 3 | —                 | —                   | V\(\rightarrow\)V | 0.001       | 0.007       | Incorrect entry in Uniprot | AGAAAGGCGC AGGAGCGTGG | TAAAGGCGTA CAGAAGAG |
| 13 | Insignin       | BMgn009002       | chr4   | 4502110           | 4509742         | Intergenic beta 2 | —                 | —                   | V\(\rightarrow\)V | 0.001       | 0.0003      | Possible LC-MS/MS artifact | AAAGTGTGG CGGACATGTGC | ACCGATTCTGC |
| 14 | Immunity-related receptor | BMgn008244 | chr6  | 11216171          | 11233988         | BmToll 12 | —                 | —                   | V\(\rightarrow\)V | 0.002       | 0.002       | Possible LC-MS/MS artifact | CGACAAACAG GGGAGTACGG | TTGTGGCTTG GGGTTG |
| 15 | Immunity-related receptor | BMgn010675 | chr1  | 9419495           | 9419495         | BmGRP-L6 | —                 | √                   | —           | 0.177       | 0.185        | Possibly incorrect annotation in genome | CAGACGACATG HUTTGTGCAG | TGGGAGAAGC FCAATAG |
| 16 | Immunity-related receptor | BMgn011085 | chr23  | 7692223           | 7702059         | BmToll 8 | —                 | —                   | V\(\rightarrow\)V | 0.035       | 0.042       | Possible LC-MS/MS artifact | AGCTGTTTGC TGACGATGTCG | CAGGCATTCTGC |
| 17 | Immunity-related receptor | BMgn011210 | chr23  | 18935856          | 18947357         | BmToll 9 | —                 | —                   | V\(\rightarrow\)V | 0.0003      | 0.0003      | Possibly expressed in the 4th instar | AGCTGCACTAC AGGCTTGCACG | CAGGCATTCTG |
| 18 | Other cell membrane receptors | BMgn012926 | chr17  | 6549039           | 6550918         | Low density lipoprotein receptor-related protein 6 (LRPL) | —                 | —                   | V\(\rightarrow\)V | 0.01        | 0.01        | Possible LC-MS/MS artifact | TGTCGCTGCG CGGCAATGACG | GCAAGAAATCT |
| 19 | Other cell membrane receptors | BMgn012138 | chr1   | 36700195          | 3684462         | Furin/Endo/Scabina/ Sushi von Willebrand factor type A, EGF and pentraxin domain-containing protein 1 | —                 | —                   | V\(\rightarrow\)V | 0.008       | 0.007       | Possible LC-MS/MS artifact | GTTGCGAATAA GCGGGACAGC | GCAGGATTCCC CAGGATC |

is high24,25, whereas on day 6 PG cells secrete high amounts of ecdysteroids while the juvenile hormone titre is low24,25 while in both days the PG cells are not fully stimulated by prothoracicotropic hormone25. In addition, the onset of wandering behaviour on day 6 is the safest benchmark that the animals have been developing orderly and there has been extensive research carried out on the signalling pathways of PG cells of *Bombyx mori* on day 6. We identified the receptors that participate in these cascades, determined their expression profile during this developmental stage and then illustrated and linked the expression of these receptors to KEGG database signalling pathways24,25 present in PG cells (Fig. 2). The datasets presented in this study were generated from PG cells from day 0 and day 6 (onset of wandering stage) of the final larval instar, the 5th instar (Fig. 1a,b). A set of 3 biological replicates of PG samples from day 0, and 3 biological replicates of PG samples from day 6 were subjected to proteomic analysis by liquid chromatography-tandem mass spectrometry (LC-MS/MS) and an identical set was subjected to transcriptomic (RNA-seq) analysis (Fig. 1).

To isolate the PGs, larvae were anesthetized by submersion in water and the two PGs of each larva were dissected rapidly (~2 min/animal) in sterile saline (0.85% NaCl). For LC-MS/MS analysis, following pre-incubation for 15–30 min in Grace’s medium (Invitrogen), glands were meticulously cleared of any associated tissue or debris, pooled and successively transferred to gradually diminishing volumes of
Fisher Scientific peptides were analysed by HPLC MS/MS coupled to an LTQ Orbitrap XL Mass spectrometer (Thermo resolution set to 60,000 at m/z and transferred into glass sample vials. Peptide concentration was determined by nanodrop absorbance (Promega) and incubated overnight at 37 °C. Peptides were eluted by centrifugation and upon speed-vac.

The binary pumps of the HPLC (RSLCnano, Thermo Scientific) consisted of solution A (2% (v/v) acetonitrile in 0.1% (v/v) formic acid) and solution B (80% acetonitrile in 0.1% formic acid). The peptides were separated using a C18 trap column (Acclaim PepMap100, 100 μm × 2 cm, Thermo Scientific) and then loaded onto a 50 cm C18 column (75 μm ID, particle size 2 μm, 100 Å, Acclaim PepMap RSLC, Thermo Scientific). The Orbitrap raw data (Data Citation 3), with peak Xcorr specificity and with maximum two missed cleavages and variable modifications of methionine oxidation, deamidation of glutamine and asparagine residues and acetylation of the N-terminus. Carbamidomethylation of cysteines was set as static modification. The identified peptides were filtered based on their Xcorr values versus peptide charge states (XCorr >2 for charge state +2 and XCorr >2.5 for charge state +3). The raw results without any further processing to remove false positives and duplicate entries are presented in Data Citation 5.

The Bombyx mori reference genome

For our analyses, a Bombyx mori reference genome was assembled20 (Data Citation 4) by incorporating assembled scaffolds anchored to chromosomes (http://sgp.dna.affrc.go.jp/pubdata/genomicsequences.html) and genome contigs assembled to scaffolds, but not anchored to chromosomes. Scaffolds that were less than 20 kb in length were excluded. The remaining sequences (chromosomes and scaffolds) were merged to a final FASTA file which was used to construct a Bowtie229 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) index for subsequent use with TopHat220 and Bowtie229 aligners. In addition, a comprehensive gene set was constructed from genes anchored to chromosomes and additional genes that
were inferred in the scaffold sequences and are available in KAIKObase (http://sgp.dna.affrc.go.jp/KAIKObase/). This comprehensive gene set was used to construct a gene file in GTF format to supply it to the TopHat2\textsuperscript{30} aligner. The source code and all files pertinent to this analysis are included in Data Citation 4.

**Sequencing and alignment of short reads**

Image analysis, base calling, and quality check were performed with the Illumina data analysis pipeline RTA v1.18.64 and Bcl2fastq v1.8.4. Reads were on average 13.26 Gb for day 0 (V − 0) samples (n = 3) and 13.69 Gb for day 6 (V − 6) samples (n = 3) (Data Citation 2), 90.6% of clusters passed Illumina filters\textsuperscript{27} and percentage of bases with Q-score ≥ 30 were 89.36%. The resulting FASTQ files containing pair-end 125 bp sequence reads were subjected to quality control using the FastQC\textsuperscript{31} package and mapped on the reference genome using TopHat2\textsuperscript{30} with the standard parameters for reads obtained with Illumina platforms apart from the following:

- the --GTF parameter was supplied with additional transcript annotation data for the *Bombyx mori* reference genome as described above.
- the --mate-inner-dist and --mate-std-dev parameters which are crucial for paired-end reads were estimated from the Bioanalyzer reports provided from the sequencing of each sample.
- --read-gap-length and --read-edit-dist were set to 3 (default is 2) to allow some more freedom in the strictness of the overall alignment procedure since the *Bombyx mori* genome is not fully annotated yet.

After completing a first round of spliced alignment with TopHat2, reads which failed to map to the reference genome were supplied to the bedtools bamtofastq command from the BEDTools suite (https://github.com/arq5x/bedtools2) to create a FASTQ subset of the original raw short reads. These short reads subset was subjected to a second round of unspliced alignment with Bowtie2 in sensitive mode (options applied: --local --very-sensitive-local --maxins 1,000 --dovetail) to allow mapping of part of this subset back to the reference genome. Mapping of this subset back to the reference genome occurred when a sufficient and continuous proportion of each read (first 50 bases) was successfully aligned. This procedure allowed for the alignment of paired-end reads located quite further than the average pair distance. These two rounds of alignment procedure led to increased alignment rates.

**Statistical analysis**

Statistical analysis was performed using the Bioconductor package metaseqR\textsuperscript{32}. Specifically, the BAM files, one for each RNA-seq sample, were summarized to a gene read counts table, using the Bioconductor package GenomicRanges\textsuperscript{33}. In the final read counts table, each row represented one gene, each column one RNA-Seq sample and each cell the corresponding read counts associated with each row and column. The gene counts table was normalized for inherent systematic or experimental biases (e.g., sequencing depth, gene length) using the Bioconductor package edgeR\textsuperscript{34} after removing genes that had zero counts over all the RNA-Seq samples (891 genes). The output of the normalization algorithm was a table with normalized counts. Prior to the statistical testing procedure, the gene read counts were filtered for possible artefacts that could affect the subsequent statistical testing procedures. Genes presenting read counts below the median read counts of the total normalized count distribution (7,139 genes with cut-off value 199 normalized read counts) were excluded from further analysis. The total number of genes excluded due to the application of gene filters was 8,051. The resulting gene counts table was subjected to differential expression analysis for the contrasts day 0, 5th instar versus day 6, 5th instar using the Bioconductor packages DESeq\textsuperscript{35}, edgeR\textsuperscript{34}, limma\textsuperscript{36}, NBSeq\textsuperscript{37}, NOISeq\textsuperscript{38} and baySeq\textsuperscript{39}. To combine the statistical significance from multiple algorithms so as to optimize the trade-off between true positives and false hits, we applied the PANDORA\textsuperscript{32} weighted p-value method across all results. Setting the p-value (FDR or adjusted p-value) threshold to 0.05, we identified 5,539 differentially expressed genes with statistically significant p-value and of these 928 were up-regulated, 1,077 were down-regulated and 3,534 were not differentially expressed according to an absolute fold change cut-off value of 1 in log2 scale.

**RNA-seq data visualisation**

To create the UCSC Genome Browser visualization track data hub (http://epigenomics.fleming.gr/tracks/hs_trackhubs/ekpa_dedos_2/hub.txt), BAM files resulting from the alignment procedure were converted to BED format (https://genome.ucsc.edu/FAQ/FAQformat.html#format1) using the bedtools bamtobed command from the BEDTools suite (http://bedtools.readthedocs.org/en/latest/index.html)\textsuperscript{40} with the -split option to report RNA-seq reads split by the TopHat2 algorithm as separate alignments, referred hereafter as ‘tags’. The RNA signal from these files was extracted by reformatting them in BedGraph (http://genome.ucsc.edu/goldenPath/help/bedgraph.html) format using the bedtools genbedcov command from the BEDTools suite with the -bg option and then to bigwig (https://genome.ucsc.edu/goldenpath/help/bigWig.html) format using the bedGraphToBigWig program supplied by UCSC. The bigWig tracks were visualized in a custom UCSC Genome Browser track data hub, hosting the *Bombyx mori* (bmori2) reference genome and the normalized (total signal of 10\textsuperscript{10}) RNA-seq samples. The track data hub is available at http://epigenomics.fleming.gr/tracks/hs_trackhubs/ekpa_dedos_2/hub.txt and this link must be pasted in the My Hubs tab of the UCSC Genome Browser application.
Signalling pathway analysis and annotation through KEGG database

KEGG pathway maps (http://www.genome.ad.jp/kegg/pathway.html) where available for the silkworm Bombyx mori\textsuperscript{23,26} or the fruit fly Drosophila melanogaster\textsuperscript{23} were used to identify proteins that are components of each pathway and examine the presence of each protein in our LC-MS/MS datasets (Data Citation 1, Data Citation 3 and Data Citation 5) and its coding gene expression in our RNA-seq datasets (Data Citation 1 and Data Citation 2). Each component of the map was subjected to blastp at the KAIKObase\textsuperscript{18} website to identify the corresponding gene in our dataset. Signalling pathway maps were manually curated to highlight the different features we identified (Fig. 2). Each of the illustrated pathways in Fig. 2 met the criterion that it had at least three of its components identified in our LC-MS/MS datasets and was not previously reported to be comprehensively present in PG cells of Bombyx mori.

Code availability

The code used to analyse the RNA-seq data, described at various levels herein, is available in an archive within Data Citation 4 with further explanations in this link (http://epigenomics.fleming.gr/metaseq_runs/HS/000004/). Briefly, the code consists of: i) Custom Perl and Linux shell scripts used to 1) create the Bombyx mori reference genome assembly for RNA-seq short read alignment, 2) create the GTF gene file and 3) generate files suitable for custom visualization of the UCSC Genome Browser. ii) Custom Linux shell scripts that can be used to reproduce the two rounds of alignment procedure described in ‘Methods’. iii) Custom R script that can be used to reproduce the differential expression analysis.

Data Records

The annotated list of Bombyx mori genes file (Data Citation 1) is available on Figshare (doi:10.6084/m9.figshare.3420235).

The raw reads of our transcriptomic (RNA-seq) data have been deposited to the NCBI Short Read Archive (SRA, http://www.ncbi.nlm.nih.gov/sra/) under accession number SRP062258 (Data Citation 2). This record combines the 3 biological replicates from day 0 (http://www.ncbi.nlm.nih.gov/biosample/SAMN03978782) and the 3 biological replicates from day 6 (http://www.ncbi.nlm.nih.gov/biosample/SAMN03978783) presented in this study.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE\textsuperscript{40} partner repository with the dataset identifier PXD004265 and 10.6019/PXD004265 (Data Citation 3).

The Bombyx mori reference gene file in GTF format together with a series of files centred around the Bombyx mori genome (Data Citation 4) is available on Figshare (doi:10.6084/m9.figshare.3420412).

The Proteome Discoverer data analysis files (Data Citation 5) are available on Figshare (doi:10.6084/m9.figshare.342031).

Technical Validation

Assessment of false negative and false positive hits of Bombyx mori cell membrane receptors in RNA-seq and LC-MS/MS datasets

Based on the initial bioinformatic identification of 369 cell membrane receptors present in the genome of Bombyx mori, we analysed by qPCR the expression of 339 transcripts out of the initial 369 cell membrane receptor genes. Due to high sequence similarity between some transcripts a total of 30 genes could not be analysed by qPCR. Of those 339 genes that were analysed by qPCR, 29 genes were found to be false negative, i.e., expressed in the PG cells (Table 1), mainly because their expression showed peaks on days other than day 0 (V = 0) or day 6 (V = 6) from which the RNA-seq and LC-MS/MS samples were derived. In 2 of the 29 instances, the RNA-seq data returned negative results due to incorrect annotation of the Bombyx mori genome and in 1 instance (i.e., putative prostanooid receptor; Table 1), gene BMgn010037 is annotated as only a fragment of the full open reading frame.

The RNA-seq and LC-MS/MS data was more complicated for the false positive results, i.e., receptors that were not found to be expressed in the PG cells by qPCR assays (Table 2). These cases were 19 in total (Table 2). In 3 cases the reason for being false positives could be attributed to no longer being expressed in the 5th instar but were expressed in the 4th instar (Table 2). In 3 cases the RNA-seq reads were aligning with sequences found within introns while in 3 other cases the LC-MS/MS data returned hits incorrectly annotated in the UniProt Bombyx mori sequence database (14,788 sequences). In the remaining 10 cases the reason for getting negative results in qPCR assays and the RNA-seq data is probably due to a combination of LC-MS/MS analyses artefacts and incorrect annotation of the Bombyx mori genome.

Quantitative PCR analysis of false negative and false positive hits of Bombyx mori cell membrane receptors in RNA-seq and LC-MS/MS datasets

The MIQE guidelines-adopted\textsuperscript{27} complete protocols of the qPCR assays are fully described in our supporting paper\textsuperscript{20}. All primers, including those shown in Tables 1 and 2, were designed with an online tool (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi) using the following custom settings: 1) amplicons size range 230–270 bp, 2) primer size minimum 18 bases, optimum 20 bases, maximum 23 bases, 3) primer Tm minimum 59 °C, optimum 60 °C, maximum 62 °C, maximum difference 1 °C, 4)
Primer GC% minimum 30, optimal 50, maximum 80, and the other settings were left to default values. Among the various combinations of primers returned by the online tool, we selected the pair that combined the following criteria: 1) minimal penalty value, 2) amplicon at the 3' end of the cDNA sequence, 3) unique hit in the NCBI Primer-BLAST webpage (http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome) using the RefSeqmRNA database and Bombyx mori as the organism, 4) unique hit in the KAIKObase website using Blastn.

Usage Notes
The annotation of KEGG pathway maps we provide in Fig. 2 can serve as a reference tool for other researchers who want to analyse the contribution of each of these signalling pathways to ecdysteroidogenesis. However, it is critical to note that all these signalling modules may not simply serve ecdysteroidogenesis, i.e., ecdysteroids secretion can not be the only readout of the activation or inhibition of components of each pathway in PG cells. The perplexing array of these pathways and the scarcity of data regarding other functions that PG cells may carry out in the insect body makes it difficult at present to assign discrete roles to each one of them and maybe there can not be a single role to be assigned to each one of them. For example, the Hedgehog signalling pathway has been linked to ecdysteroidogenesis in Drosophila melanogaster but does this link mean that the Wnt and the TGF signalling pathways (Fig. 2) are also involved in ecdysteroidogenesis?

Data Citation 2 contains the raw RNA-seq data from Bombyx mori PGs. The files have been analysed with up-to-date software packages such as TopHat230 and Bowtie229 aligners for short read mapping to the Bombyx mori reference genome, and the Bioconductor package metaseqR with the PANDORA32 method for normalization and statistical analyses. Our data can be used to examine the performance of other short read aligners or it can be a suitable resource in studies that map differential gene expression in various tissues or organs of this or other insect species.

Data Citation 3 contains the raw proteomics data from Bombyx mori PGs. The files have been analysed with the software package Proteome Discoverer 1.4. We believe that extending the data analysis to obtain quantitative protein profiles from the existing raw files using available proteomic analysis platforms would be quite interesting because it will provide a dimension of protein abundance in PG cells. Furthermore, identification and differential quantitation of post-translational modifications will be of major importance. Being datasets derived from control, untreated whole tissue extracts, our data can be a suitable resource when combined with future studies on organelle-enriched proteomic analysis or phosphoproteomic analysis of ecdysteroidogenesis in insects.

We believe that the publicly available custom track data hub resource on UCSC server (http://epigenomics.fleming.gr/tracks/hg_trackhubs/ekpa_dedos_2/hub.txt) hosting our RNA-seq data can be the appropriate visual tool in deciphering gene and exon/intron boundaries of yet uncharacterized Bombyx mori genes and identifying previously unidentified genes through the extensive visual inspection features of the tracks.

We encourage other research teams to compare our Proteome Discoverer version 1.4. data analysis files (Data Citation 5) with different or more up-to-date software and analyse our raw data (Data Citation 3) to yield further insight than we already report.

The annotated gene list of Bombyx mori that we provide here (Data Citation 1) can be integrated with the publicly available datasets on KAIKObase and can serve as a template for further annotation of Bombyx mori genes.

The Bombyx mori reference gene file in GTF format (contained within Data Citation 4) can be the appropriate resource to use in follow-up research on the expression profiles of other genes that participate in ecdysteroidogenesis. Data Citation 4 contains i) the assembled Bombyx mori genome as described in the ‘Methods’ section, ii) the assembled Bombyx mori transcriptome assembled like the Bombyx mori genome and used for optimizing the short read alignment procedure, iii) the assembled Bombyx mori genes in GFF, GTF and text tab-delimited format, iv) bowtie2 indexes for the Bombyx mori genome, v) bowtie2 indexes for the Bombyx mori transcriptome and vi) custom code used for the analysis.

With the full report produced by the Bioconductor package metaseqR (http://epigenomics.fleming. gt/metaseqR_runs/HG/0000004/), researchers may explore a rich set of quality controls regarding the raw data (Data Citation 2), as well as several analytics regarding sample quality, normalization effects and differential expression analysis. In addition, researchers may download the full list of Bombyx mori genes coupled with RNA abundance values and other quality metrics as well as all the necessary parameters that will help them reproduce the differential expression analysis.

Samples, subjects, and data outputs
Detailed information accounting for each sample, the data-generating assays applied to each sample and the resulting data outputs are provided in Data Citation 1.

References
1. Pandey, U. B. & Nichols, C. D. Human Disease Models in Drosophila melanogaster and the Role of the Fly in Therapeutic Drug Discovery. Pharmacological Reviews 63, 411–436 (2011).
2. Strachan, R. T., Ferrara, G. & Roth, B. L. Screening the receptorome: an efficient approach for drug discovery and target validation. Drug Discovery Today 11, 708–716 (2006).
3. Brody, T. & Cravchik, A. Drosophila melanogaster G Protein–Coupled Receptors. The Journal of Cell Biology 150, F83–F88 (2000).

4. Li, C. et al. Comparative genomic analysis and evolution of family-B G protein-coupled receptors from six model insect species. Gene 519, 1–12 (2013).

5. Fan, Y. et al. The G protein-coupled receptors in the silkworm, Bombyx mori. Insect Biochemistry and Molecular Biology 40, 581–591 (2010).

6. Li, C. et al. Identification of G protein-coupled receptors in the pea aphid, Acyrthosiphon pisum. Genomics 102, 345–354 (2013).

7. Nowling, R. et al. Identification of novel arthropod vector G protein-coupled receptors. Parasites & Vectors 6, 150 (2013).

8. Vogel, K. J., Brown, M. R. & Strand, M. R. Phylogenetic investigation of peptide hormone and growth factor receptors in five dipteran genomes. Frontiers in Endocrinology 4, 193 (2013).

9. Hauser, F. et al. A genome-wide inventory of neurohormone GPCRs in the red flour beetle Tribolium castaneum. Frontiers in Neuroendocrinology 29, 142–165 (2008).

10. Smith, W., Rybczynski, R. Insect Endocrinology (ed. Gilbert L. J.) 62 (Academic Press, 2012).

11. Gilbert, L. I., Rybczynski, R. & Warren, J. T. Control and biochemical nature of the ecdysteroidogenic pathway. Annual Review of Entomology 47, 883–916 (2002).

12. Marchal, E. et al. Control of ecdysteroidogenesis in prothoracic glands of insects: A review. Peptides 31, 506–519 (2010).

13. Rodenfels, J. et al. Production of systemically circulating Hedgehog by the intestine couples nutrition to growth and development. Nature Reviews Genetics 14, 263–265 (2011).

14. Iga, M., Nakaoka, T., Suzuki, Y. & Kataoka, H. Pigment Dispersing Factor Regulates Ecdysone Biosynthesis via Bombyx Neuropeptide G Protein Coupled Receptor-B2 in the Prothoracic Glands of Bombyx mori. PLoS ONE 9, e103239 (2014).

15. Gibbens, Y. Y., Warren, J. T., Gilbert, L. I. & O’Connor, M. B. Neuroendocrine regulation of Drosophila metamorphosis requires TGFβ/Activin signaling. Development 138, 2693–2703 (2011).

16. Yamanaka, N. et al. Bombyx prothoracostatic peptides activate the sex peptide receptor to regulate ecdysteroid biosynthesis. Proceedings of the National Academy of Sciences 107, 2060–2065 (2010).

17. Yamanaka, N. et al. Neuropeptide Receptor Transcriptome Reveals Unidentified Neuroendocrine Pathways. PLoS ONE 3, e3048 (2008).

18. The International Silkworm Genome Consortium. The genome of a lepidopteran model insect, the silkworm Bombyx mori. Insect Biochemistry and Molecular Biology 38, 1036–1045 (2008).

19. Duan, J. et al. SilkDB v2.0: a platform for silkworm (Bombyx mori) genome biology. Nucleic Acids Research 38, D453–D456 (2010).

20. Alexandratos, A., Moulos, P., Nellas, I., Mavridis, K. & Dedos, S. G. Reassessing ecdysteroidogenic cells from the cell membrane receptors’ perspective. Scientific Reports 6, 20229 (2016).

21. Goldberg, T. et al. LocTree3 prediction of localization. Nucleic Acids Research 42, W350–W355 (2014).

22. Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools. Briefings in Bioinformatics 14, 144–161 (2013).

23. Kaneshia, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research 44, D457–D462 (2015).

24. Nimi, S. & Sakurai, S. Development changes in juvenile hormone and juvenile hormone acid titer in the hemolymph and in-vitro juvenile hormone synthesis by corpora allata of the silkworm, Bombyx mori. Journal of Insect Physiology 43, 875–884 (1997).

25. Mizoguchi, A., Dedos, S. G., Fugo, H. & Kataoka, H. Basic pattern of fluctuation in hemolymph PTTH titers during larval-pupal and pupal-adult development of the silkworm, Bombyx mori. General and Comparative Endocrinology 127, 181–189 (2002).

26. Kaneshia, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 28, 27–30 (2000).

27. Bustin, S. A. et al. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clinical Chemistry 55, 611–622 (2009).

28. Wiśniewski, J., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. Nature Methods 6, 359–362 (2009).

29. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359 (2012).

30. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14, R36 (2013).

31. FastQC A quality control tool for high throughput sequence data. Babraham Bioinformatics Web site http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

32. Moulos, P. & Hatzis, P. Systematic integration of RNA-Seq statistical algorithms for accurate detection of differential gene expression patterns. Nucleic Acids Research 43, e25 (2015).

33. Lawrence, M. et al. Software for Computing and Annotating Genomic Ranges. PLoS Comp Biol 9, e1003118 (2013).

34. 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).

35. Anders, S. & Huber, W. Differential expression analysis for sequence count data. Genome Biology 11, R106 (2010).

36. Smyth, G. K. in Bioinformatics and Computational Biology Solutions Using R and Bioconductor Statistics for Biology and Health, (eds Gentleman R. et al.) Ch. 23, 397–420 (Springer, 2005).

37. Di, Y., Schafer Daniel, W., Cumbie Jason, S. & Chang Jeff, H. in Statistical Applications in Genetics and Molecular Biology Vol. 10 (2011).

38. Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. & Conesa, A. Differential expression in RNA-seq: A matter of depth. Genome Research 21, 2213–2223 (2011).

39. Hardcastle, T. J. & Kelly, K. A. baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11, 1–10 (2010).

40. Viscaínos, J. A. et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat Biotech 32, 223–226 (2014).

Data Citations
1. Moulos, P., Samiotaki, M., Panayotou, G. & Dedos, S. G. figshare http://dx.doi.org/10.6084/m9.figshare.3420235 (2016).

2. NCBI Sequence Read Archive, SRP062258 (2016).

3. ProteomeXchange Consortium, PXD004280 (2016).

4. Moulos, P., Samiotaki, M., Panayotou, G. & Dedos, S. G. figshare http://dx.doi.org/10.6084/m9.figshare.3420412 (2016).

5. Moulos, P., Samiotaki, M., Panayotou, G. & Dedos, S. G. figshare http://dx.doi.org/10.6084/m9.figshare.3420319 (2016).
Acknowledgements
Research reported in this publication was supported from funds by the National and Kapodistian University of Athens under award number Kapodistrias: 11240.

Author Contributions
P.M. conceived and designed the experiments, performed the experiments, provided software and resources, analysed the data and wrote the paper. M.S. provided software and resources, analysed the data and wrote the paper. G.P. provided software and resources and analysed the data. S.G.D. conceived and designed the experiments, performed the experiments, provided software and resources, analysed the data and wrote the paper.

Additional Information
Competing financial interests: Dr Panagiotis Moulos is a co-founder of HybridStat Predictive Analytics G.P. and a member of its scientific advisory board.

How to cite this article: Moulos, P. et al. Combinatory annotation of cell membrane receptors and signalling pathways of Bombyx mori prothoracic glands. Sci. Data 3:160073 doi: 10.1038/sdata.2016.73 (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

Metadata associated with this Data Descriptor is available at http://www.nature.com/sdata/ and is released under the CC0 waiver to maximize reuse.

© The Author(s) 2016