bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

1 # Transcriptome analysis reveals the effects of transgenic expression of the
2 Gal4 protein on normal gene expression in silkworm tissues

3 Tao Chen[1], Yan Ma[2], Rongpeng Liu[2], Tingting Tan[2], Lihua Huang[3,*], Hanfu Xu[2,*]

4 [1]College of Biotechnology, Jiangsu University of Science and Technology; The Sericultural Research Institute,
5 Chinese Academy of Agricultural Sciences, Jiangsu212003, China
6 [2]State Key Laboratory of Silkworm Genome Biology, College of Biotechnology, Southwest University, Chongqing
7 400715, China
8 [3]International Bioinformatics Center, BGI Genomics Co., Ltd, Shenzhen, Guangdong518083, China

9 ***Corresponding author**
10 Lihua Huang
11 Email: atomnhuang@gmail.com

12 Hanfu Xu
13 Email: xuhf@swu.edu.cn

14
15
16
17
18
19
20
21
22
23

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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 4.0 International license.

24 ## Abstract

25 The Gal4/upstream activating sequence(UAS) system, a well-known genetic tool, has been widely
26 used to analyze gene function in many organisms, including the silkworm (*Bombyx mori*), a model
27 lepidopteran insect. Several studies have suggested that Gal4 protein activation in tissues can
28 negatively affect transgenic individuals; however, whether and to what extent the Gal4 protein
29 affects normal endogenous gene expression have rarely been studied. Here, we analyzed the
30 transcriptomes of transgenic silkworms expressing the Gal4 protein at high levels in both the wing
31 disc (WD) and epidermis (EP) and investigated gene expression changes in both tissues. Overall,
32 24,593 genes were identified in the WD and EP libraries, and 2,025 and 2,488 were identified as
33 significant differentially expressed genes(DEGs) in the WD and EP between the transgenic and
34 control groups, respectively. These DEGs were further annotated by gene function classification and
35 pathway assessment using public databases. In addition, 506 DEGs were shared (common) between
36 both tissues. Of these, 97 genes were commonly upregulated, and 234 were commonly
37 downregulated; many of them were annotated to be involved in metabolic processes such as "fat
38 digestion and absorption", "glycine, serine and threonine metabolism" and "glutathione metabolism"
39 and in signal transduction pathways such as the "Rap1 signaling pathway", "MAPK signaling
40 pathway" and "Hippo signaling pathway". Overall, this work enhances understanding of the effects
41 of transgenic Gal4 protein expression on normal gene expression in silkworm tissues and suggests
42 that researchers should pay attention to unexpected effects when using the Gal4/UAS system to study
43 gene function.

44 **Keywords** Transgenic silkworm, Gal4 protein, wing disc, epidermis, transcriptome

45 ## Introduction

46 The Gal4/upstream activating sequence(UAS) binary expression system, derived from yeast and
47 originally developed in *Drosophila* [1-3], is a powerful genetic tool that allows manipulation of
48 target gene expression in a spatiotemporally precise fashion. Since its first application in *Drosophila*,
49 the Gal4/UAS system has been widely used to analyze gene function in dozens of organisms,
50 including mice [4], zebrafish [5], *Xenopus* [6], *Bombyx mori* [7], *Arabidopsis thaliana* [8],
51 *Triboliumcastaneum* [9], *Aedes aegypti* [10], *Anopheles stephensi* [11] and *Caenorhabditis elegans*

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. 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 4.0 International license.

52 [12]. The Gal4/UAS system has also been employed to develop novel genetic tools, such as the
53 enhancer/gene trap system and the Q system [13-15], and it has been combined with genome editing
54 tools for conditional manipulation of gene expression *in vivo* [16-17].

55 In recent decades, remarkable progress in gene function analysis has been achieved with the
56 Gal4/UAS system. However, the fact that high protein levels of Gal4 have certain toxicity toward
57 cells coexpressing UAS-linked target genes and Gal4 protein cannot be ignored. Although some
58 researchers have described and developed novel Gal4/UAS systems with smaller sizes but greater
59 transactivation efficiency than the original system [18-20], little attention has been paid to the effects
60 of Gal4 on the normal expression of nontarget genes. To objectively clarify the functions of target
61 genes, it is necessary to determine whether and to what extent Gal4 protein expression affects normal
62 nontarget gene expression.

63 Recently, we established a transgenic silkworm line (named A4G4) that expresses the Gal4
64 mainly in the wing disc (WD) and epidermis (EP) under the control of the promoter of the *B. mori*
65 *Actin4* gene [21]. This line is a good material for evaluation of the effects of Gal4 protein expression
66 on normal gene expression in transgenic tissues. In this study, we conducted a comprehensive
67 transcriptome analysis of WD and EP tissues and identified thousands of differentially expressed
68 genes (DEGs) in both tissues. Our findings provide sufficient evidence, for the first time, that
69 transgenic protein expression of Gal4 in silkworm tissues does affect the normal expression of
70 nontarget genes.

71 # Materials and Methods

72 ## Silkworm and sample collection

73 The wild-type (WT) silkworm strain *Nistari*, the A4G4 transgenic line, and the UtdTomato
74 transgenic line harboring a UAS-linked red fluorescent protein variant (tdTomato) were maintained
75 in our laboratory. The hatched larvae were reared at 24-28°C with fresh mulberry leaves. WT and
76 A4G4 larvae at day5 of the fifth instar were selected, and the WDs and EPs were dissected, washed
77 with precooled phosphate-buffered saline and used for subsequent experiments. The WD samples
78 collected from WT and A4G4 larvae were named W5N_S1/S2/S3 and W5A_S1/S2/S3, respectively.
79 The EP samples collected from WT and A4G4 larvae were named EP5N_S1/S2/S3 and

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. 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 4.0 International license.

80 EP5A_S1/S2/S3, respectively.

81 RNA preparation and sequencing

82 Total RNA was extracted from the WD and EP samples using TRIzol Reagent (Ambion, USA) and
83 examined on a NanoDrop 1000 spectrophotometer (Thermo Scientific, USA) and an Agilent
84 Bioanalyzer 2100 system (Agilent Technologies, USA) for RNA integrity and quality. The qualified
85 RNA samples were purified for poly-A-containing mRNA molecules using poly-T oligo-attached
86 magnetic beads, fragmented into small pieces using divalent cations under elevated temperature and
87 reverse transcribed using random primers. Thesecond-strand cDNA fragments were ligated with
88 index adapters after being purified, end-repaired, and A-tailed. Suitable fragments were used as
89 templates for PCR amplification. After quantification with a Qubit instrument, the PCR products
90 were sequenced on a BGISEQ-500 platform at Beijing Genomics Institute (BGI, China).

91 Alignment and quantification

92 The raw sequencing data were preprocessed using SOAPnuke software
93 (https://github.com/BGI-flexlab/SOAPnuke) to remove reads with adaptors, reads with more than 5%
94 unknown bases, and reads with low sequencing quality. The clean reads were mapped to the *B. mori*
95 genome (http://sgp.dna.affrc.go.jp/KAIKObase/, ver.3.2.2) using HISAT software [22], and the
96 transcripts were reconstructed using StringTie [23]. Subsequently, Cuffcompare (Cufflinks tools,
97 [24]) was utilized to compare the reconstructed transcripts. The novel coding transcripts predicted by
98 the Coding Potential Calculator (CPC) [25] were combined with gene models from KAIKObase to
99 obtain a new reference gene set. In addition, the clean reads were aligned against the new reference
100 gene set using Bowtie [26]. Gene expression levels were quantified by RSEM [27] and normalized
101 using the fragments per kilobase per million mapped reads (FPKM) method [24,27]. The DEGseq
102 method [28] was used to detect DEGs with adjusted *P* values <0.001. Genes were considered
103 significant DEGs if their fold changes were >=2 and their adjusted probability values were <0.001.

104 Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) 105 enrichment analysis

106 The functions of the protein-coding genes were assigned according to the best matches derived from
107 alignments to proteins in the National Center for Biotechnology Information (NCBI) nonredundant

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. 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 4.0 International license.

108 (Nr) protein sequence database using DIAMOND [29]. GO annotation was performed for all
109 identified DEGs, and WEGO software [30] was used to conduct the GO functional classification. GO
110 terms with adjusted *P* values ≤0.05 were defined as significantly enriched GO terms for the DEGs.
111 Pathway enrichment analysis of DEGs was performed based on the KEGG database [31] with the
112 same criteria.

113 Data availability

114 The raw sequence reads are available from the NCBI search database (Bioproject PRJNA601576,
115 run accessions SAMN13870652, SAMN13870653, SAMN13870654, and SAMN13870655). All
116 relevant data are within the paper and its Supporting Information files.

117 Results and Discussion

118 Detection of Gal4 expression in transgenic silkworms

119 Before collecting the WD and EP samples, we first generated A4G4>UtdTomato transgenic
120 silkworms by crossing A4G4 moths with UtdTomato moths (S1 Fig), and we detected the
121 fluorescence distribution in various tissues of A4G4>UtdTomato silkworms to further confirm the
122 locations of Gal4 protein expression. As shown in Fig 1, strong fluorescence intensity was detected
123 in both the WDs and EPs but not in other tissues of the A4G4>UtdTomato individuals. In other
124 words, the expression of UAS-linked tdTomato was activated by the Gal4 protein mainly in these
125 two tissues, which clearly demonstrated that the Gal4 protein was expressed at comparatively high
126 levels in the WDs and EPs of A4G4 transgenic silkworms. Then, WD and EP samples were collected
127 from day-5 fifth-instar larvae of the WT and A4G4 lines and submitted to RNA extraction, library
128 construction and DNA sequencing.

129 Summary of the transcriptome data

130 RNA sequencing (RNA-Seq) generated a total of 540.19 million raw reads for all samples. After
131 processing of these raw data, 478.36 million clean reads were obtained, with an average read depth
132 ranging from 32.26 to 45.38 million. A total of 87.60-89.50% of the reads in each sample reached the
133 Q30 quality score. The majority of reads in each library were mapped to the ver.3.2.2 assembly of
134 the *B. mori* genome, and the average mapping rate of the reads was 77.57% (Table 1). The
135 assembled transcriptome data were used to identify known and predict novel coding transcripts,

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. 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 4.0 International license.

136 which generated 24,593 genes (23,064 known genes and 1,529 novel genes). The gene number in
137 each sample ranged from 16,439 to 18,097. Gene expression levels were calculated with RSEM
138 software using the FPKM normalization method. Approximately 30% of the expressed genes had
139 FPKM values larger than 10.0, and 30% had FPKM values lower than 1.0 (Fig 2).

140 Identification of DEGs between the transgenic and WT groups

141 By comparing transcriptome data between the transgenic and WT groups, a number of genes
142 expressed in the WD and EP were identified as significant DEGs (Fig 3). In WD samples, 2,025
143 genes were identified as DEGs, including 771 upregulated genes and 1,254 downregulated genes
144 (Table S1). In EP samples, a total of 2,488 DEGs were identified, among which 771 genes were
145 upregulated and 1,717 genes were downregulated (Table S2). Thus, approximately 8.23%
146 (2,025/24,593) and 10.12% (2,488/24,593) of the genes in the WD and EP tissues of transgenic
147 silkworms were upregulated and downregulated, respectively, which clearly indicates that transgenic
148 expression of the Gal4 protein in either the WD or EP affects the normal expression of endogenous
149 genes.

150 GO annotation and KEGG pathway enrichment analysis of the DEGs

151 To obtain valuable information for DEG functional prediction, the DEGs were annotated with the
152 GO database. In total, 952 DEGs in WD samples were annotated in 38 functional categories,
153 including 15 biological process categories, 12 cellular component categories and 11 molecular
154 function categories. Among the biological process categories, “cellular process” was the main
155 functional group, followed by “metabolic process” and “response to stimulus”. Among the cellular
156 component categories, “membrane” was the main functional group, followed by “membrane part”
157 and “cell”. Among the molecular function categories, “binding” and “catalytic activity” were the two
158 main functional groups (Fig 4A). In EP samples, 1,621 DEGs were functionally annotated with 15
159 biological process categories, 14 cellular component categories and 10 molecular function categories.
160 The most enriched GO terms in the biological process category were “cellular process”, “metabolic
161 process” and “biological regulation”. The terms “membrane”, “membrane part” and “cell” were
162 significantly enriched at the cellular component level, and the terms “binding” and “catalytic activity”
163 were significantly enriched at the molecular function level (Fig 4B). The top 10 up- and
164 downregulated annotated DEGs as well as the significantly enriched GO terms for the DEGs in WD

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. 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 4.0 International license.

165 and EP samples are listed in Table S3 ~ Table S6.

166 To better interpret the pathways in which the DEGs were involved and enriched, we annotated
167 the DEGs against the KEGG database. Briefly, the DEGs in WD samples were mainly enriched for
168 the "phototransduction - fly", "Influenza A", "Hippo signaling pathway", "fat digestion and
169 absorption", "viral myocarditis", and "oxytocin signaling pathway" terms (Fig 5A). In EP samples,
170 the DEGs were mainly annotated with the "complement and coagulation cascades", "amoebiasis",
171 "tyrosine metabolism", "ECM-receptor interaction", "insect hormone biosynthesis", "Hippo
172 signaling pathway", "axon guidance", and "fat digestion and absorption" pathway terms, among
173 others (Fig 5B).

174 **Comparative analysis of the DEGs between the WD and EP tissues**

175 Considering that the WD and EP are known to be involved in the regulation of wing development in
176 *B. mori*, the DEGs in both tissues were further analyzed to identify commonalities and differences.
177 As shown in the Venn diagram in Fig 6A, 506 genes were identified as common DEGs in both
178 tissues. Of these, 331 DEGs were common up- or downregulated genes (97 upregulated and 234
179 downregulated). Moreover, 111 DEGs were upregulated in WD tissues and downregulated in EP
180 tissues, while 64 DEGs were upregulated in EP tissues and downregulated in WD tissues (Table S7).
181 We further focused on the 331 common DEGs, the patterns of which might be influenced by the
182 Gal4 protein in a similar way in these two tissues.

183 First, the 331 DEGs were annotated in the NR NCBI database to a total of 208 Nr terms that
184 encompassed 124 Nr functions. Genes annotated with more than 2 functions are listed in Table 2.
185 The downregulated genes were annotated with the "actin-5C-like", "actin-4", "actin, cytoplasmic 2",
186 "glucose dehydrogenase", "atlastin", "E3 ubiquitin-protein ligase", "26S proteasome non-ATPase
187 regulatory subunit", and "histidine-rich glycoprotein-like" terms. Similarly, some upregulated genes
188 were also annotated with the "actin-5C-like", "actin-4", "actin, cytoplasmic 2", "glucose
189 dehydrogenase", "E3 ubiquitin-protein ligase", and "26S proteasome non-ATPase regulatory subunit"
190 terms. These findings imply that the Gal4 protein in either the WD or EP affects the expression of
191 actin genes as well as genes involved in metabolic processes.

192 Next, we mapped all of the genes to terms in the GO database to look for significantly enriched
193 GO terms. Among the 331 DEGs, 72 genes were annotated with 444 GO terms. Among the 60 most

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. 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 4.0 International license.

194 enriched GO terms were "multi-organism process", "multicellular organismal process" and
195 "biological regulation" for the biological process category; "organelle part", "cell" and
196 "macromolecular complex" for the cellular component category; and "binding", "catalytic activity"
197 and "signal transducer activity" for the molecular function category (Fig 6B). Overall, 39 common
198 genes were annotated with these GO terms, including 16 upregulated genes and 23 downregulated
199 genes.

200 Finally, the 331 DEGs were annotated against the KEGG database to better understand the
201 biochemical pathways in which they were involved. Among the 331 DEGs,119 genes were annotated
202 in 5 main categories. The 20 most enriched KEGG terms were "fat digestion and absorption",
203 "phagosome", "antigen processing and presentation", "phototransduction - fly", "Rap1 signaling
204 pathway", "platelet activation", "oxytocin signaling pathway", "MAPK signaling pathway - fly",
205 "Hippo signaling pathway", "pentose phosphate pathway", "apoptosis", "arachidonic acid
206 metabolism", "axon guidance", "glycine, serine and threonine metabolism", "glutathione
207 metabolism", "MAPK signaling pathway", "Hippo signaling pathway - fly", "phototransduction",
208 "thyroid hormone signaling pathway", and "leukocyte transendothelial migration" (Fig 6C). Overall,
209 59 genes were annotated with these 20 KEGG terms, 36 of which were downregulated and 23 of
210 which were upregulated. Taken together, these functional annotations suggest that the expression of
211 genes involved in a series of metabolic processes and signal transduction pathways is influenced by
212 the Gal4 protein in either the WDs or the EPs of transgenic silkworms.

213 Conclusion

214 In this study, RNA-Seq, de novo assembly and functional annotation were performed to characterize
215 the transcriptome profiles of two types of tissues, the WD and EP, that were collected from day-5
216 fifth-instar larvae of silkworms of the WT *Nistari* line and the transgenic A4G4 line expressing the
217 Gal4 protein mainly in the WD and EP. We conducted comparative transcriptome analyses to
218 identify the DEGs and potentially related biological pathways in the silkworms. A total of 2,025 and
219 2,488 genes were identified as significant DEGs in the WD and EP, respectively, among which 506
220 DEGs were common to both tissues (97 were commonly upregulated, and 234 were commonly
221 downregulated). Many of the common genes were annotated to be involved in metabolic processes
222 ("fat digestion and absorption", "glycine, serine and threonine metabolism", "glutathione

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. 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 4.0 International license.

223 metabolism", etc.), and signal transduction pathways ("Rap1 signaling pathway", "MAPK signaling
224 pathway", "Hippo signaling pathway", etc.). Overall, our results present a comprehensive view of
225 gene expression profiles in the WDs and EPs of WT and A4G4 silkworms and reveal that transgenic
226 expression of the Gal4 protein affects normal gene expression in these tissues. Our findings also
227 provide timely and valuable information for future studies on gene function using the Gal4/UAS
228 binary system.

229 Acknowledgments

230 This work was supported by grants (31872291) from the National Natural Science Foundation of
231 China and grants (cstc2017jcyjBX0041) from the Chongqing Research Program of Basic Research
232 and Frontier Technology. American Journal Experts performed English language editing in this
233 manuscript.

234 Author Contributions

235 Conceived and designed the experiments: HFX. Performed the experiments: RPL TTT. Analyzed the
236 data: TC YM LHH HFX. Contributed reagents/materials/analysis tools: YM RPL. Wrote the paper:
237 HFX LHH.

238 References

239 1. Laughon A, Gesteland RF. Primary structure of the Saccharomyces cerevisiae GAL4 gene.
240 Molecular and Cellular Biology. 1984;4: 260–267. doi:10.1128/mcb.4.2.260
241 2. Giniger E, Varnum SM, Ptashne M. Specific DNA binding of GAL4, a positive regulatory
242 protein of yeast. Cell. 1985;40: 767–774. doi:10.1016/0092-8674(85)90336-8
243 3. Fischer JA, Giniger E, Maniatis T, Ptashne M. GAL4 activates transcription in Drosophila.
244 Nature. 1988;332: 853–856. doi:10.1038/332853a0
245 4. Ornitz DM, Moreadith RW, Leder P. Binary system for regulating transgene expression in mice:
246 targeting int-2 gene expression with yeast GAL4/UAS control elements. Proceedings of the
247 National Academy of Sciences. 1991;88: 698–702. doi:10.1073/pnas.88.3.698
248 5. Scheer N, Campos-Ortega JA. Use of the Gal4-UAS technique for targeted gene expression in
249 the zebrafish. Mechanisms of Development. 1999;80: 153–158.
250 doi:10.1016/s0925-4773(98)00209-3
251 6. Hartley KO, Nutt SL, Amaya E. Targeted gene expression in transgenic Xenopus using the
252 binary Gal4-UAS system. Proceedings of the National Academy of Sciences. 2002;99:

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. 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 4.0 International license.

253 1377–1382. doi:10.1073/pnas.022646899
254 7. Imamura M, Nakai J, Inoue S, Quan GX, Kanda T, Tamura T. Targeted gene expression using
255 the GAL4/UAS system in the silkworm Bombyx mori. Genetics. 2003;16 : 1329–1340.
256 doi:10.1023/B:GENE.0000003842.72339.df
257 8. Laplaze L, Parizot B, Baker A, Ricaud L, Martinière A, Auguy F, et al. GAL4-GFP enhancer
258 trap lines for genetic manipulation of lateral root development in Arabidopsis thaliana. Journal of
259 Experimental Botany. 2005;56: 2433–2442. doi:10.1093/jxb/eri236
260 9. Schinko JB, Weber M, Viktorinova I, Kiupakis A, Averof M, Klingler M, et al. Functionality of
261 the GAL4/UAS system in Tribolium requires the use of endogenous core promoters. BMC
262 Developmental Biology. 2010;10: 53. doi:10.1186/1471-213x-10-53
263 10. Kokoza VA, Raikhel AS. Targeted gene expression in the transgenic Aedes aegypti using the
264 binary Gal4-UAS system. Insect Biochemistry and Molecular Biology. 2011;41: 637–644.
265 doi:10.1016/j.ibmb.2011.04.004
266 11. Obrochta DA, Alford RT, Pilitt KL, Aluvihare CU, Harrell RA. piggyBac transposon
267 remobilization and enhancer detection in Anopheles mosquitoes. Proceedings of the National
268 Academy of Sciences. 2011;108: 16339–16344. doi:10.1073/pnas.1110628108
269 12. Wang H, Liu J, Gharib S, Chai CM, Schwarz EM, Pokala N, et al. cGAL, a temperature-robust
270 GAL4–UAS system for Caenorhabditis elegans. Nature Methods. 2016;14: 145–148.
271 doi:10.1038/nmeth.4109
272 13. Yang MY, Armstrong J, Vilinsky I, Strausfeld NJ, Kaiser K. Subdivision of the drosophila
273 mushroom bodies by enhancer-trap expression patterns. Neuron. 1995;15: 45–54.
274 doi:10.1016/0896-6273(95)90063-2
275 14. Asakawa K. The Tol2-mediated Gal4-UAS method for gene and enhancer trapping in zebrafish.
276 Methods. 2009;49: 275–281. doi:10.1016/j.ymeth.2009.01.004
277 15. Potter CJ, Tasic B, Russler EV, Liang L, Luo L. The Q System: A Repressible Binary System for
278 Transgene Expression, Lineage Tracing, and Mosaic Analysis. Cell. 2010;141: 536–548.
279 doi:10.1016/j.cell.2010.02.025
280 16. Xue Z, Wu M, Wen K, Ren M, Long L, Zhang X, et al. CRISPR/Cas9 Mediates Efficient
281 Conditional Mutagenesis in Drosophila. G3 (Bethesda). 2014;4: 2167–2173.
282 doi:10.1534/g3.114.014159
283 17. Santis FD, Donato VD, Bene FD. Clonal analysis of gene loss of function and tissue-specific
284 gene deletion in zebrafish via CRISPR/Cas9 technology. Methods in Cell Biology The Zebrafish
285 - Genetics, Genomics, and Transcriptomics. 2016;135: 171–188.
286 doi:10.1016/bs.mcb.2016.03.006
287 18. Köster RW, Fraser SE. Tracing Transgene Expression in Living Zebrafish Embryos.
288 Developmental Biology. 2001;233: 329–346. doi:10.1006/dbio.2001.0242
289 19. Kobayashi I, Kojima K, Uchino K, Sezutsu H, Iizuka T, Tatematsu K-I, et al. An efficient binary
290 system for gene expression in the silkworm, Bombyx mori, using GAL4 variants. Archives of
291 Insect Biochemistry and Physiology. 2011;76: 195–210. doi:10.1002/arch.20402

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. 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 4.0 International license.

292 20. Zhang Y, Ouyang J, Qie J, Zhang G, Liu L, Yang P. Optimization of the Gal4/UAS transgenic
293 tools in zebrafish. Applied Microbiology and Biotechnology. 2019;103: 1789–1799.
294 doi:10.1007/s00253-018-09591-0
295 21. Tan T, Liu R, Luo Q, Ma J, Ou Y, Zeng W, et al. The intronic promoter of Actin4 mediates
296 high-level transgene expression mainly in the wing and epidermis of silkworms. Transgenic
297 Research. 2020; doi:10.1007/s11248-020-00192-0
298 22. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements.
299 Nature Methods. 2015;12: 357–360. doi:10.1038/nmeth.3317
300 23. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables
301 improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology. 2015;33:
302 290–295. doi:10.1038/nbt.3122
303 24. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and
304 transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature
305 Protocols. 2012;7: 562–578. doi:10.1038/nprot.2012.016
306 25. Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, et al. CPC: assess the protein-coding
307 potential of transcripts using sequence features and support vector machine. Nucleic Acids
308 Research. 2007;35. doi:10.1093/nar/gkm391
309 26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012;9:
310 357–359. doi:10.1038/nmeth.1923
311 27. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a
312 reference genome. BMC Bioinformatics. 2011;12. doi:10.1186/1471-2105-12-323
313 28. Wang L, Feng Z, Wang X, Wang X, Zhang X. DEGseq: an R package for identifying
314 differentially expressed genes from RNA-seq data. Bioinformatics. 2009;26: 136–138.
315 doi:10.1093/bioinformatics/btp612
316 29. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nature
317 Methods. 2014;12: 59–60. doi:10.1038/nmeth.3176
318 30. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, et al. WEGO: a web tool for plotting GO
319 annotations. Nucleic Acids Research. 2006;34. doi:10.1093/nar/gkl031
320 31. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes
321 to life and the environment. Nucleic Acids Research. 2007;36. doi:10.1093/nar/gkm882

322 ## Supporting Information

323 **S1 Fig. Generation of A4G4>UtdTomato transgenic silkworms.** Adults positive for UtdTomato
324 (carrying 3×P3-EGFP) and A4G4 (carrying 3×P3-DsRed) showed only GFP and RFP fluorescence,
325 respectively, in the compound eye. A4G4>UtdTomato adults showed both GFP and RFP
326 fluorescence in the compound eye. 3×P3 is an artificial promoter that can drive EGFP or DsRed
327 expression mainly in the ocelli and nervous tissues of *B. mori*.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. 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 4.0 International license.

328 **S1 Table. List of DEGs in the WD between the control and transgenic groups.**

329 **S2 Table. List of DEGs in the EP between the control and transgenic groups.**

330 **S3 Table. Top 10 annotated DEGs in the WD between the control and transgenic groups.**

331 **S4 Table. Top 10 annotated DEGs in the EP between the control and transgenic groups.**

332 **S5 Table. Top enriched GO terms for the DEGs in the WD between the control and transgenic**
333 **groups.**

334 **S6 Table. Top enriched GO terms for the DEGs in the EP between the control and transgenic**
335 **groups.**

336 **S7 Table. List of common DEGs identified in both WD and EP samples.**

337 Figure legends

338 **Fig 1. Detection of tdTomato expression in A4G4>UtdTomato transgenic silkworms.** (A)
339 Expression of tdTomato in A4G4>UtdTomato silkworms at different developmental stages. (B)
340 Expression of tdTomato in tissues of day-5 A4G4>UtdTomato fifth-instar larvae. E9D, day-9
341 embryo stage; 1L2D, day-2 first-instar larval stage; 2L2D, day-2 second-instar larval stage; 3L3D,
342 day-3 third-instar larval stage; 4L3D, day-3 fourth-instar larval stage; 4LM, fourth molting stage;
343 5L1D ~ 5L6D, day-1 to day-6 fifth-instar larval stages; W1D, day-1 wandering stage; P1D ~ P7D,
344 day-1 to day-7 pupal stages; A1D, day-1 adult stage. The numbers on the photos denote the different
345 microscope magnifications.

346 **Fig 2. Gene expression level distribution in different FPKM value ranges.** The X-axis represents
347 the number of genes. FPKM <= 1, genes with very low expression levels; FPKM = 1-10, genes with
348 relatively low expression levels; FPKM >= 10, genes with medium to high expression levels.

349 **Fig 3. DEGs identified in the WD and EP samples.** (A) Number of DEGs between the transgenic
350 group and the WT group. (B) Volcano map of DEGs between the transgenic group and the WT
351 group. Upregulated genes, downregulated genes, and non-significantly altered genes are indicated
352 with red, blue, and gray points, respectively. The X-axis represents the fold change in each gene, and
353 the Y-axis represents the significance level.

354 **Fig 4. GO analysis of the DEGs between the transgenic group and the WT group.** (A) GO
355 classification of the DEGs in the WD. (B) GO classification of the DEGs in the EP.

356 **Fig 5. KEGG analysis of DEGs between the transgenic group and WT group.** (A) KEGG

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. 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 4.0 International license.

357 classification of DEGs in the WD. (B) KEGG classification of DEGs in the EP.

358 **Fig 6. Comparison of the DEGs between WD and EP tissues.** (A) Venn plot comparing DEGs in

359 WD and EP tissues. (B) Bubble chart of the enriched GO terms for the DEGs. (C) Bubble chart of

360 the enriched KEGG pathways for the DEGs.

361 # Table

362 **Table 1. Characteristics of the RNA-Seq reads of the WD and EP samples.**

Sample	Total Raw Reads (Mb)	Total Clean Reads (Mb)	Clean Reads Q30 (%)	Clean Reads Percentage (%)	Total Mapped (%)	Uniquely Mapped (%)
W5N_S1	47.33	42.34	88.28	89.47	73.15	61.30
W5N_S2	37.42	33.16	88.66	88.61	68.49	55.94
W5N_S3	49.08	42.96	88.37	87.52	70.17	58.66
W5A_S1	37.46	32.26	89.06	86.14	73.13	61.09
W5A_S2	45.25	40.10	88.56	88.62	70.76	59.23
W5A_S3	45.68	40.70	88.34	89.09	70.80	59.01
EP5N_S1	41.18	36.47	87.70	88.56	73.26	60.73
EP5N_S2	47.43	42.69	89.23	90.00	75.15	62.65
EP5N_S3	49.19	43.92	89.50	89.28	76.73	64.15
EP5A_S1	50.94	45.38	89.47	89.07	77.01	63.81
EP5A_S2	49.19	43.07	89.22	87.57	76.59	63.26
EP5A_S3	40.04	35.31	87.60	88.20	78.23	59.64

363 **Table 2. Genes annotated with more than 2 functions.**

Gene ID	**Log2 (EP5A/EP5N)**	**Log2 (W5A/W5N)**	**Nr Annotation**
BMgn003598	-2.48	-1.15253	XP_004930247.2|0.0e+00|atlastin [*Bombyx mori*]
BMgn005742	-1.17889	-1.09857	XP_004923914.2|0.0e+00|atlastin [*Bombyx mori*]
BMgn012863	-2.8081	-1.65447	XP_012548093.1|0.0e+00|glucose dehydrogenase [FAD, quinone] isoform X1 [*Bombyx mori*]
BMgn012872	-3.46208	-2.74461	XP_012548096.1|0.0e+00|glucose dehydrogenase [FAD, quinone] [*Bombyxmori*]
BMgn013005	-4.07261	-1.21234	XP_021205532.1|0.0e+00|glucose dehydrogenase [FAD, quinone] [*Bombyxmori*]
BMgn013008	-2.9674	-4.17712	XP_021205528.1|8.0e-163|glucose dehydrogenase [FAD, quinone]-like [*Bombyx mori*]
BMgn015000	-1.11199	-1.18581	XP_011564474.1|2.2e-68|PREDICTED: probable 26S proteasome non-ATPase regulatory subunit 3 [*Plutella xylostella*]
E_FL_bmmt_01L08_R_0	-4.79443	-1.64395	XP_019758493.1|7.9e-101|PREDICTED: actin-5C-like

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. 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 4.0 International license.

			[*Dendroctonusponderosae*]
E_FL_bmmt_07O06_F_0	-6.0958	-3.09502	XP_013136486.1|3.2e-56|PREDICTED: E3 ubiquitin-protein ligase ZNRF2 isoform X2 [*Papilio polytes*]
E_FL_dpe-_02E19_F_0	-2.66122	-2.27044	XP_004924489.1|1.2e-50|probable G-protein coupled receptor Mth-like 2 [*Bombyx mori*]
E_FL_fmgV_51I15_R_0	-1.18466	-4.27579	XP_019758493.1|7.2e-102|PREDICTED: actin-5C-like [*Dendroctonusponderosae*]
E_FL_fner_10A05_R_0	-1.21217	-1.26425	XP_019758493.1|9.0e-94|PREDICTED: actin-5C-like [*Dendroctonusponderosae*]
E_FL_fner_47N09_R_0	-2.21741	-1.94716	KOB77672.1|1.3e-85|26S proteasome non-ATPase regulatory subunit [*Operophtera brumata*]
E_FL_fner_52N05_R_0	-3.27478	-1.48226	XP_011501341.1|7.5e-09|PREDICTED: histidine-rich glycoprotein-like [*Ceratosolen solmsi marchali*]
E_FL_ftes_09L13_R_0	-2.56488	-2.28381	XP_004924489.1|1.5e-30|probable G-protein coupled receptor Mth-like 2 [*Bombyx mori*]
E_FL_ftes_43K14_R_0	-2.92985	-3.10448	XP_004924489.1|4.4e-24|probable G-protein coupled receptor Mth-like 2 [*Bombyx mori*]
E_FL_fwgP_51N21_F_0	-8.37787	-9.919	AGR44787.1|5.4e-151|actin-4 [*Bombyx mori*]
E_FL_wd--_21P11_R_0	-6.90228	-5.30772	XP_009896276.1|1.2e-30|PREDICTED: LOW QUALITY PROTEIN: actin, cytoplasmic 2 [*Picoides pubescens*]
E_FL_wd--_28G10_R_0	-2.55679	-5.206	XP_009896276.1|3.4e-30|PREDICTED: LOW QUALITY PROTEIN: actin, cytoplasmic 2 [*Picoides pubescens*]
brP-0880	-1.68186	-1.23617	XP_011501341.1|9.5e-09|PREDICTED: histidine-rich glycoprotein-like [*Ceratosolen solmsi marchali*]
brP-1638	-1.43669	-2.065	AGR44854.1|6.2e-119|truncated actin-4 [*Bombyx mori*]
fe8d-P11_F_I20	-8.65711	-1.18149	XP_014366484.1|2.7e-48|PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like isoform X5 [*Papilio machaon*]
ovS311F08f	-1.17925	-1.20354	XP_009896276.1|3.5e-30|PREDICTED: LOW QUALITY PROTEIN: actin, cytoplasmic 2 [*Picoides pubescens*]
BMgn012997	6.761191	1.726472	XP_004928920.1|0.0e+00|glucose dehydrogenase [FAD, quinone]-like isoform X1 [*Bombyx mori*]
E_FL_bmmt_10M22_F_0	7.550044	6.623201	XP_013136486.1|3.9e-56|PREDICTED: E3 ubiquitin-protein ligase ZNRF2 isoform X2 [*Papilio polytes*]
E_FL_bmmt_14H21_R_0	1.800765	5.493634	XP_019758493.1|1.5e-81|PREDICTED: actin-5C-like [*Dendroctonusponderosae*]
E_FL_fmgV_33M18_F_0	8.646377	6.611614	KOB77672.1|1.7e-86|26S proteasome non-ATPase regulatory subunit [*Operophtera brumata*]
E_FL_fner_08N04_R_0	7.881922	5.108628	XP_019758493.1|3.2e-82|PREDICTED: actin-5C-like [*Dendroctonusponderosae*]
E_FL_fner_10D19_R_0	5.097958	8.602168	AGR44881.1|1.8e-88|actin-4 [*Bombyx mori*]
E_FL_fner_47G22_R_0	8.34665	8.575025	XP_019758493.1|7.7e-107|PREDICTED: actin-5C-like [*Dendroctonusponderosae*]
E_FL_wd--_27A01_R_0	7.145351	8.021206	XP_009896276.1|1.2e-30|PREDICTED: LOW QUALITY PROTEIN: actin, cytoplasmic 2 [*Picoides pubescens*]

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. 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 4.0 International license.

fepMP03_F_B14	1.731906	7.428733	XP_014366484.1|4.5e-51|PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like isoform X5 [*Papilio machaon*]

364

Figure 1

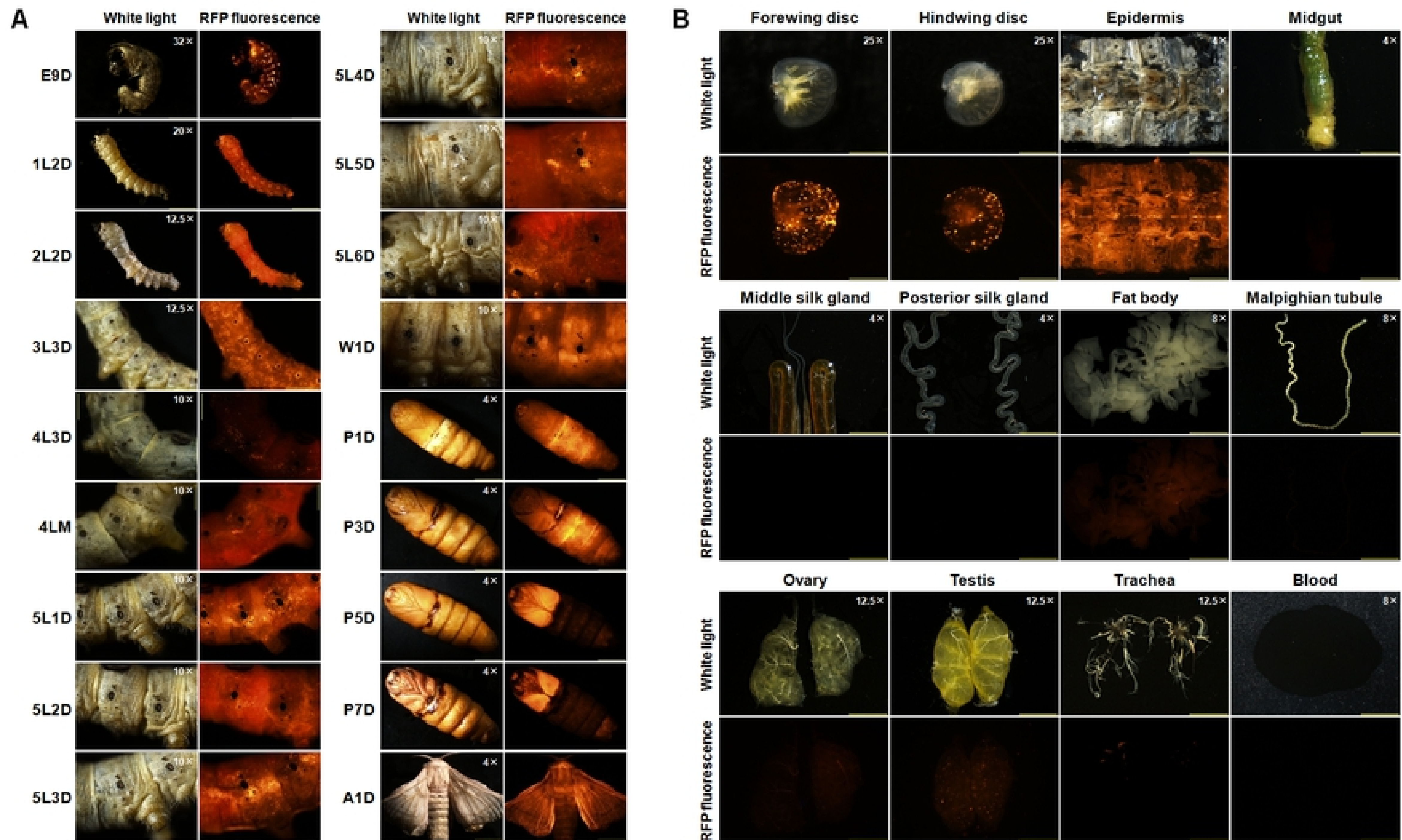


Fig 1

Figure 2

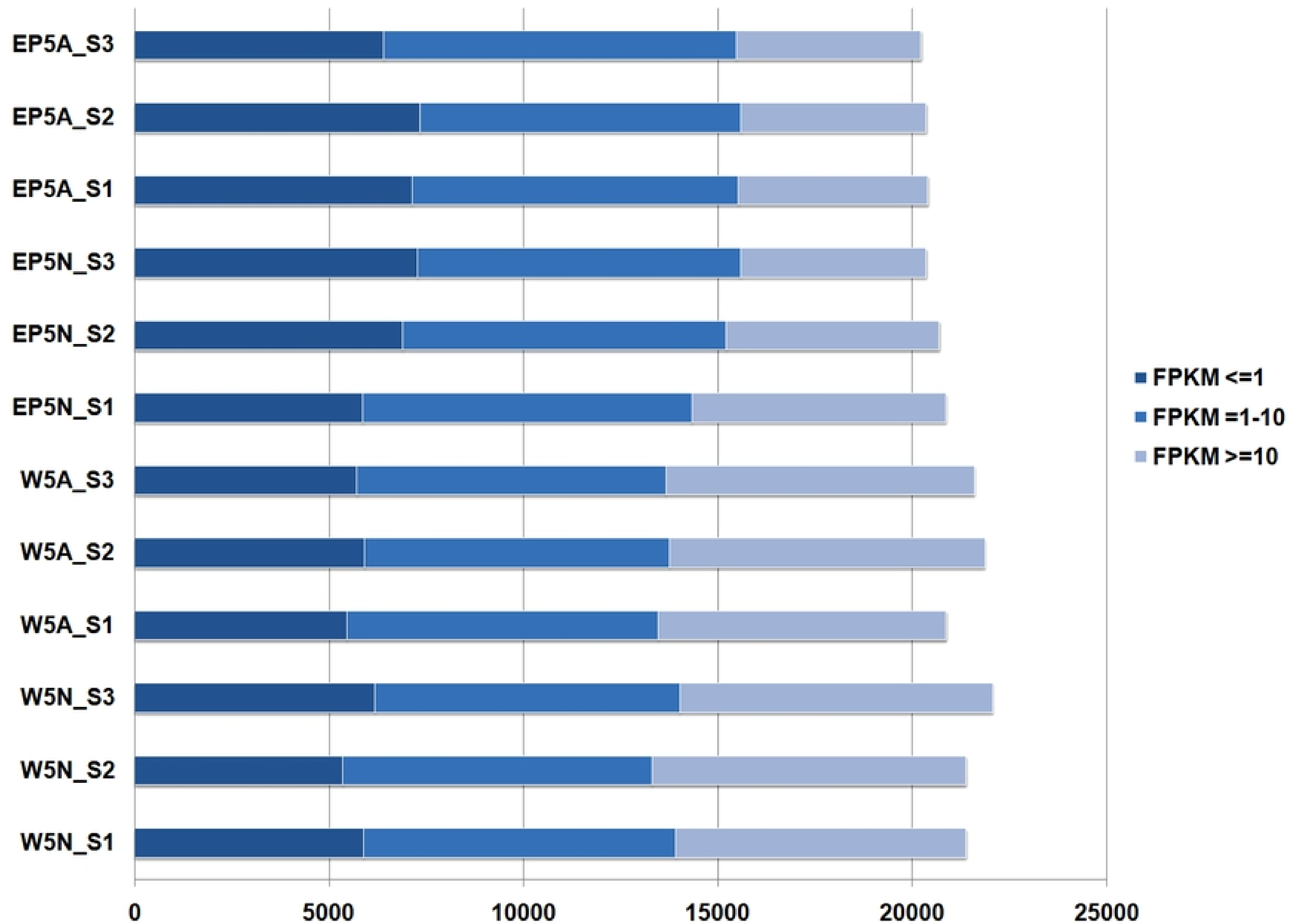
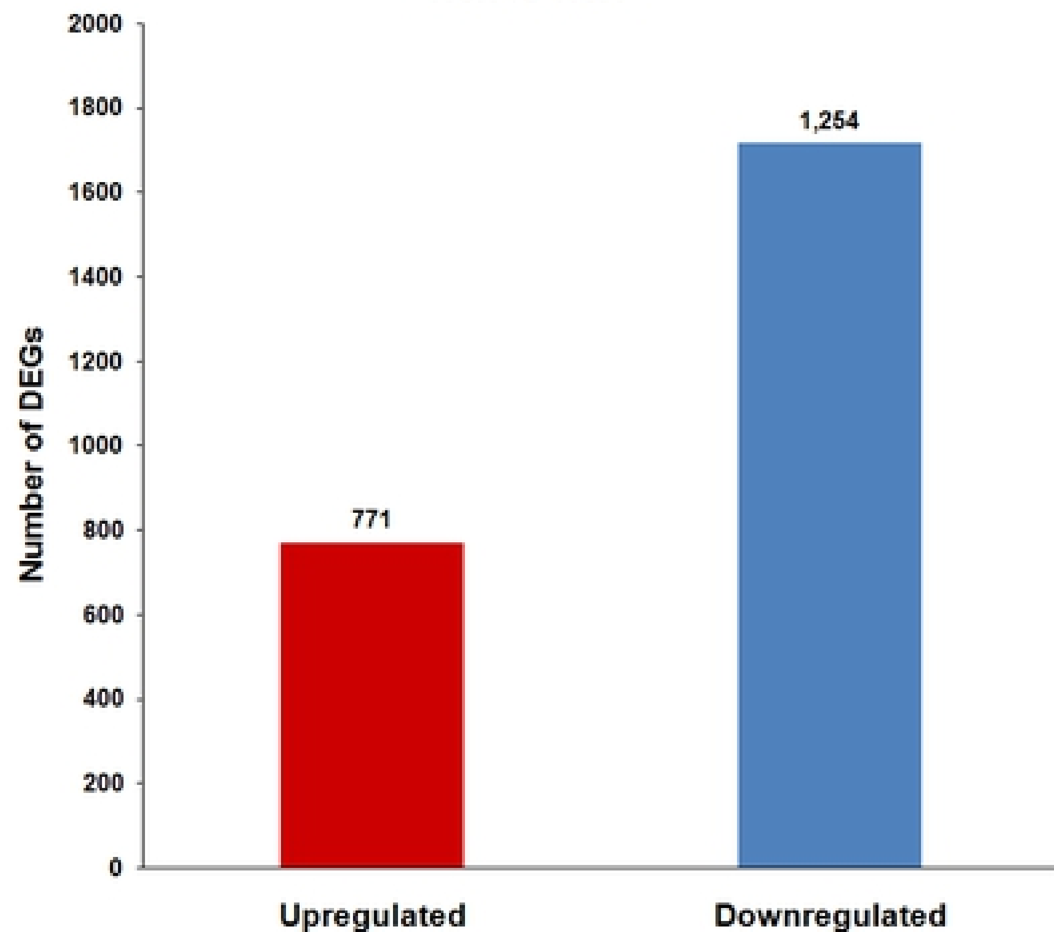


Fig 2

Figure 3A



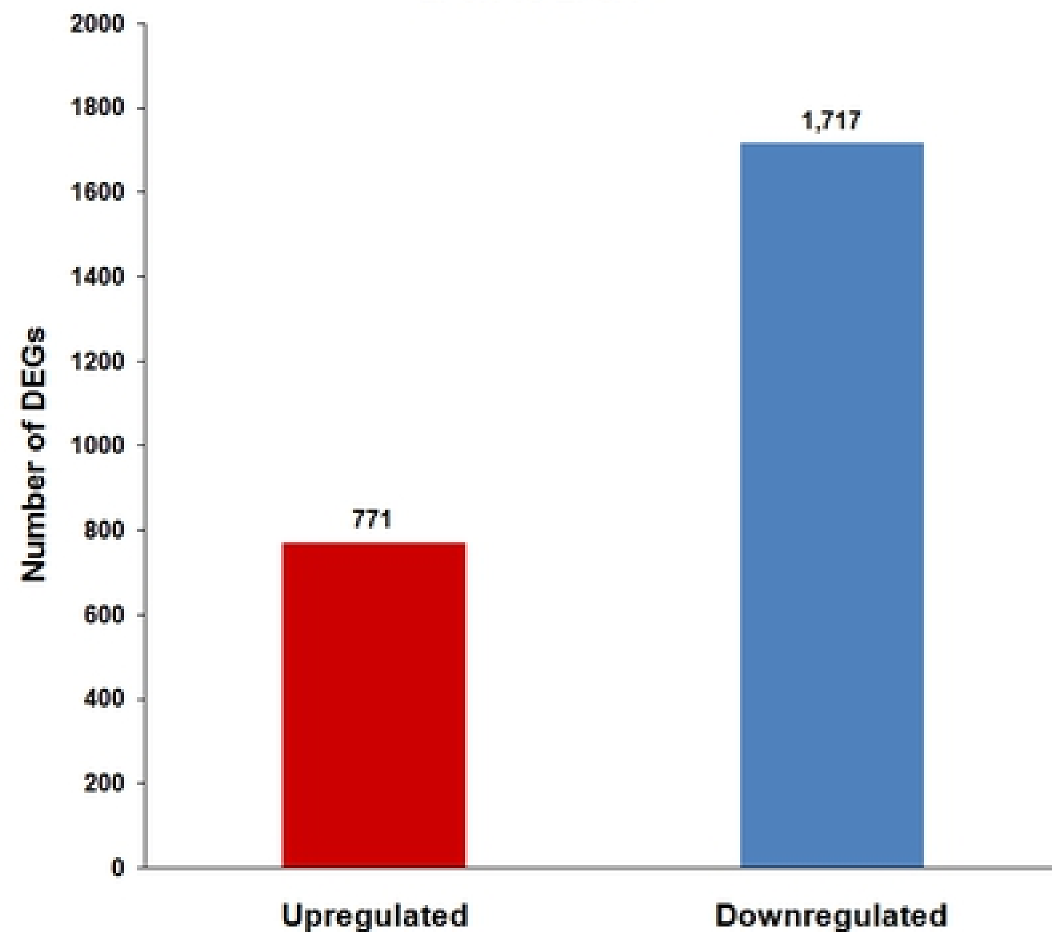


Fig 3A

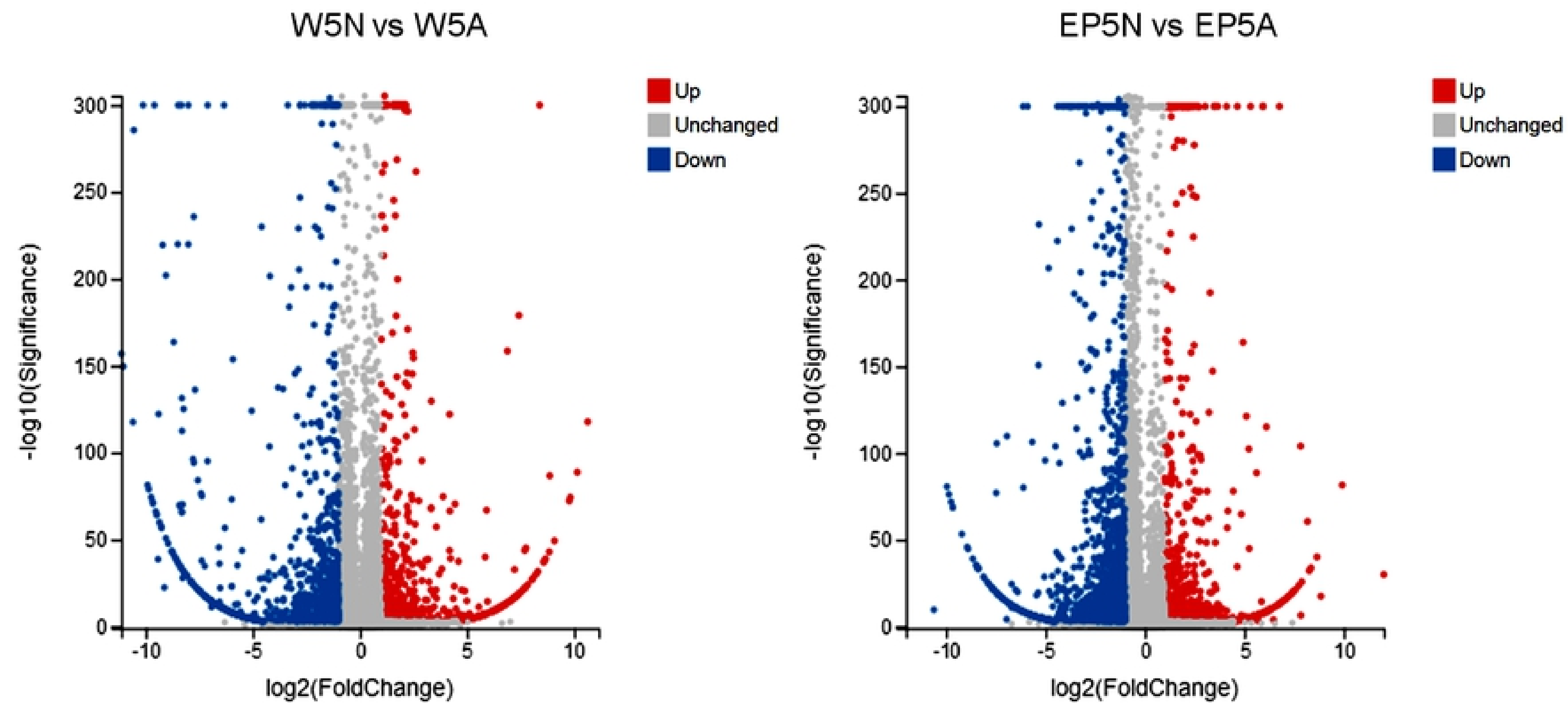


Fig 3B

Figure 4

A

Cellular process
Metabolic process
Immune system process
Multi-organism process
Response to stimulus
Biological regulation
Signaling
Cellular component organization or biogenesis
Localization
Biological adhesion
Developmental process
Multicelluar organismal process
Behavior
Reproduction
Reproductive process
Membrane
Membrane part
Extracellular region
Cell
Organelle
Macromolecular complex
Organelle part
Cell part
Virion
Virion part
Membrane-enclosed lumen
Supramolecular complex
Binding
Molecular transducer activity
Signal transducer activity
Structural molecule activity
Catalytic activity
Molecular function regulator
Transporter activity
Molecular carrier activity
Transcription regulator activity
Protein tag
Antioxidant activity
0
50
100
150
Number of genes
Biological process
Cellular component
Molecular function

B

Response to stimulus
Immune system process
Multi-organism process
Cellular component organization or biogenesis
Cellular process
Metabolic process
Biological regulation
Signaling
Localization
Behavior
Multicelluar organismal process
Reproduction
Reproductive process
Developmental process
Biological adhesion
Membrane
Membrane part
Extracellular region
Cell
Membrane-enclosed lumen
Organelle
Organelle part
Cell junction
Synapse
Macromolecular complex
Synapse part
Cell part
Supramolecular complex
Virion
Virion part
Binding
Catalytic activity
Structural molecule activity
Molecular function regulator
Molecular transducer activity
Signal transducer activity
Transporter activity
Antioxidant activity
Transcription regulator activity
0
50
100
150
200
250
Number of genes
Biological process
Cellular component
Molecular function

Fig 4

Figure 5

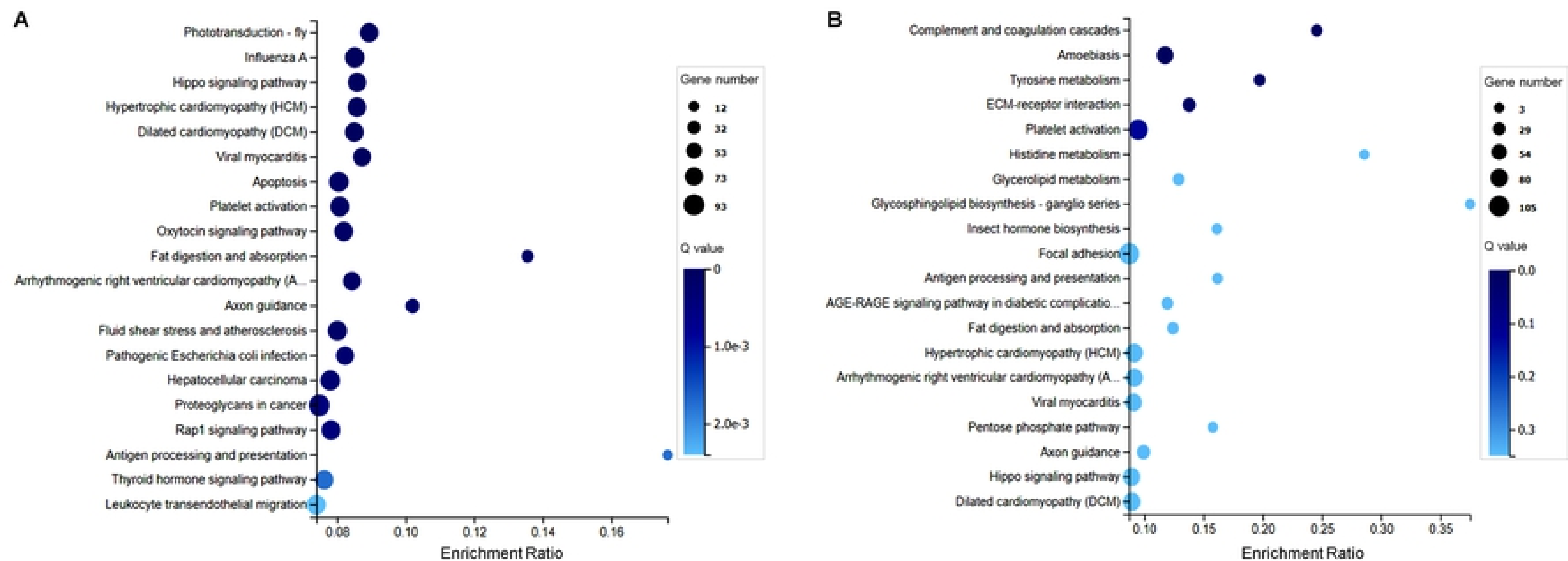


Fig 5

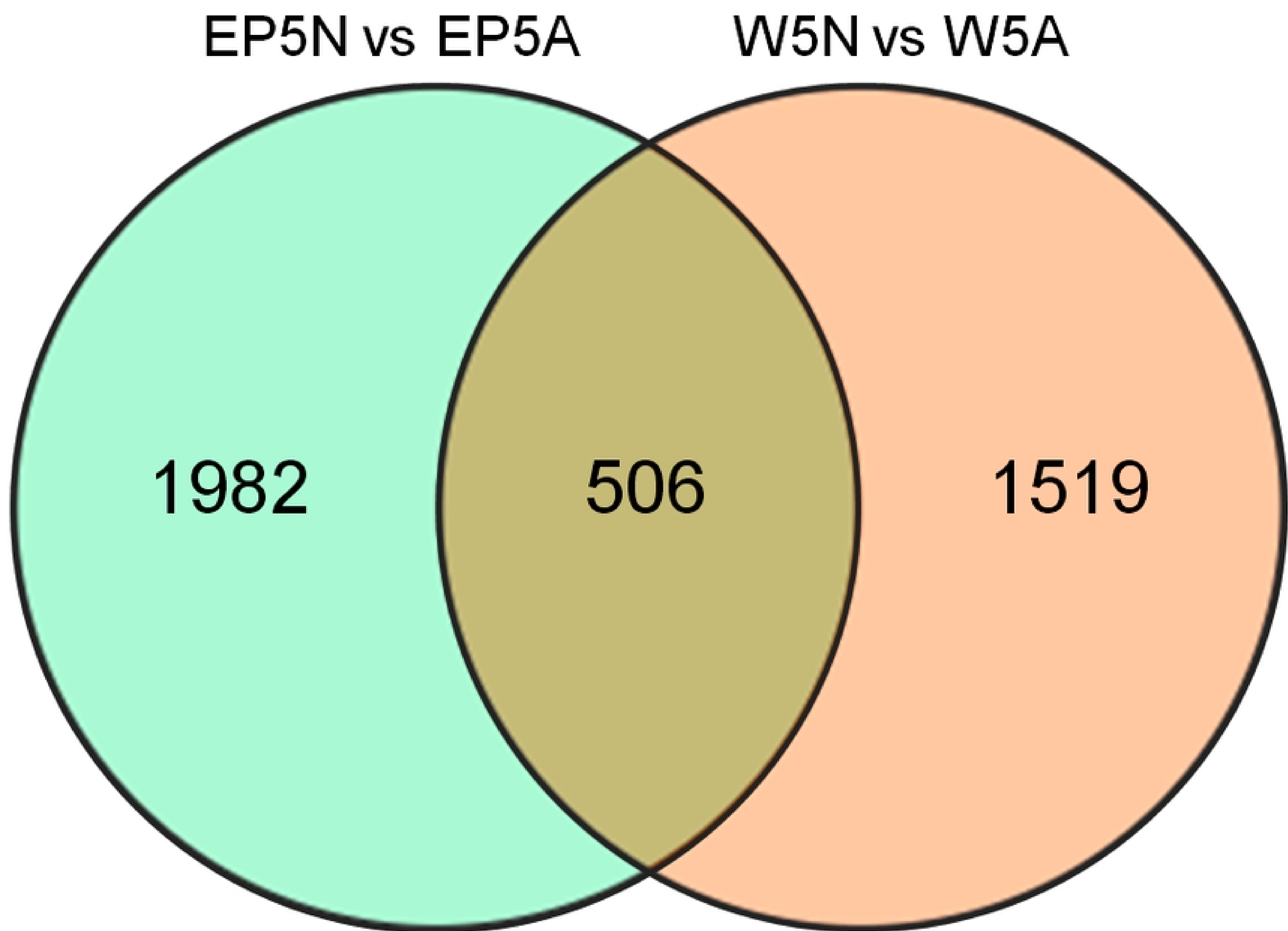
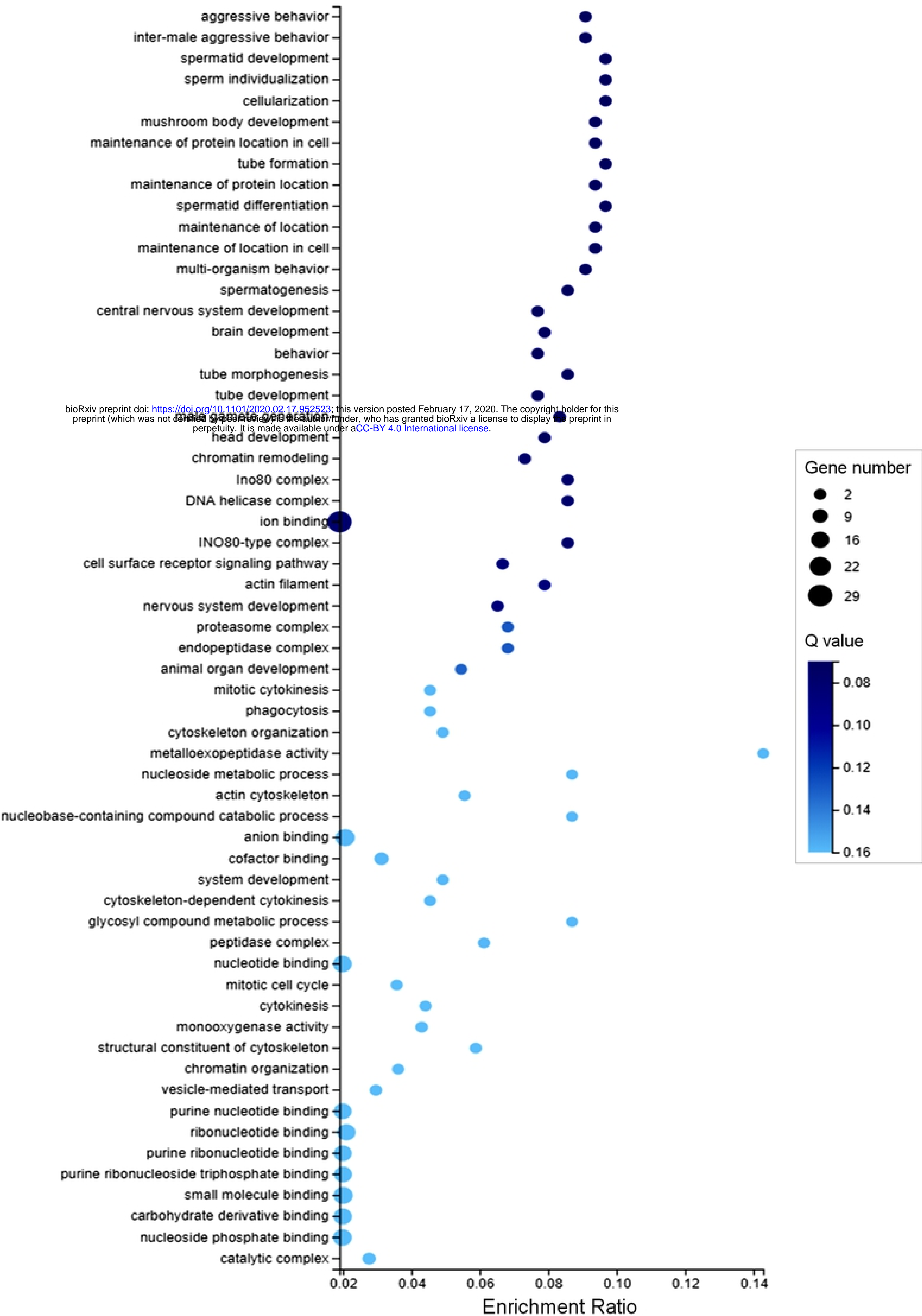


Fig 6A

Figure 6B



bioRxiv preprint doi: https://doi.org/10.1101/2020.02.17.952523; this version posted February 17, 2020. 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 4.0 International license.

Fig 6B

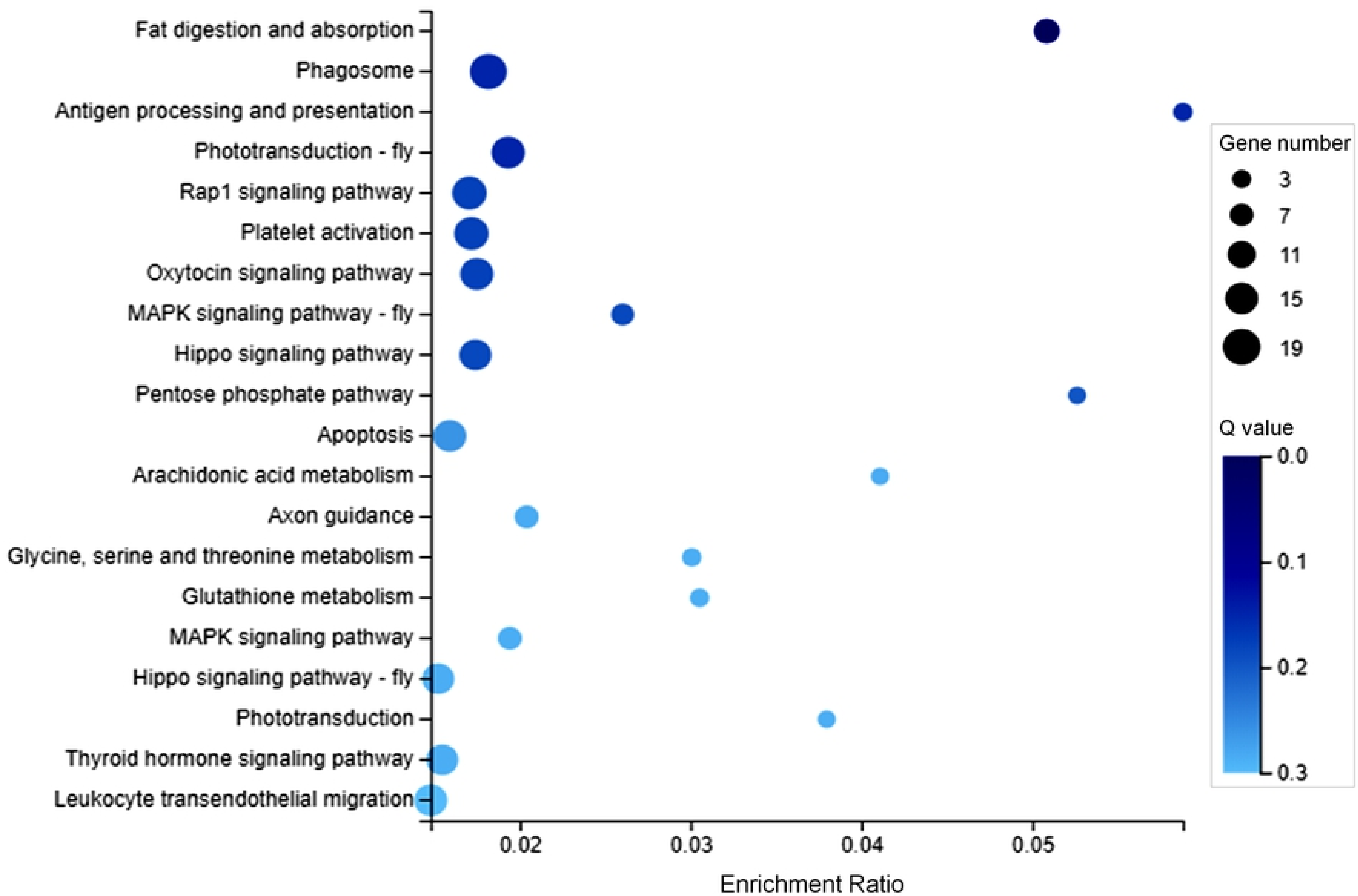


Fig 6C