Genomic responses to socio-sexual environment in male *Drosophila melanogaster* exposed to conspecific rivals

IRINA MOHORIANU*, AMANDA BRETMAN*, DAMIAN T. SMITH, EMILY FOWLER, TAMAS DALMAY and TRACEY CHAPMAN

SUPPLEMENTAL MATERIAL 1, 2

SUPPLEMENTAL MATERIAL 1 - ANALYSIS OF FUNCTIONAL ENRICHMENT

A standard method for examining whether particular sets of genes are up or down regulated is to perform a GO enrichment analysis (GO reference genome project 2009). We chose an equivalent analysis using ‘stripped down’ keywords to eliminate redundancy and improve the focus. We first conducted a manual search of the lists of genes showing DE for the major biological processes and molecular functions represented. A subsequent cross-reference to standard GO enrichment revealed that no important functions were missed (see below).

Our selection of ‘keywords’ based on this search was equivalent to a standard GO analysis but gave a tighter description with less redundancy. We focussed on biological processes and conducted a manual keyword curation to reduce redundant terms and minimise some potentially problematic issues that can occur with GO enrichment (e.g. Davies et al. 2010; Khatri et al. 2012). The GO hierarchy is an interconnected network and GO analysis necessarily yields enrichment scores for categories of genes with different levels of resolution – i.e. genes with specific functions can be assigned to higher level terms that are general and relatively uninformative (e.g. ‘transcription’, ‘translation’, ‘RNA’). An example from our data was that under a standard GO enrichment, seminal fluid and accessory proteins appeared under ‘transcription’. It would be more biologically informative if they were contained in a specific functional category. Seminal fluid genes were not captured by any biologically specific term under standard GO, but the creation of a new ‘seminal / accessory’ keyword category achieved this. The assignment of genes to any of our keyword categories was supported statistically in the same way as for standard GO analysis. A second example is that many genes appeared in the ‘response to stimulus’ category, which covers a range of different responses (to heat / cold / smell / shock / sound). It may be more informative if specific categories and functions are highlighted over the general ones.

In many cases standard GO analysis returned categories containing the same genes, which potentially introduced unnecessary redundancy. For example, the cellular component category (GO:0005576 extracellular region, GO:0005615 extracellular space and GO:0044421 extracellular region part) picked up many seminal fluid protein genes in the abdomen tests and all three GO categories contained many genes in common. GO therefore failed to detect the significant, specific contribution of these genes to male reproductive responses – as a result of placing them with all secreted transcripts. There were many other redundancies, e.g. ‘system
process’ and ‘neurological system process’ returned exactly the same genes as ‘sensory perception of chemical stimulus’. The latter was more informative and the use of the keyword ‘sensory’, eliminated considerable redundancy.

Our strategy was to (i) conduct a standard GO analysis to check that we had not overlooked major functions with the keywords approach, and (ii) calculate standard GO enrichment support for our chosen keyword categories.

(i) Summary of GO analysis. We summarise the GO terms for which there was evidence of significant enrichment below. In bold are the functions also captured by our keywords analysis (note that some keywords were eliminated from our analysis if they returned too few genes):

**HeadThorax:**
- 2h (P values from GO enrichment analysis: $8 \times 10^{-14} < P < 0.009$): **GPCR**, receptor signalling, signal transduction, regulation of cellular processes, single organism signalling, signalling, regulation of transcription, regulation of biosynthetic processes, regulation of transcription, regulation of biological processes, cell communication, biosynthetic processes, transcription, RNA biosynthesis, biological regulation, regulation of metabolic processes, **response to stimulus**, neurotransmitter / neuropeptide / peptide receptor, transmembrane signalling, chitin binding, sensory perception, NADH, pheromone binding, cell surface receptor signalling, regulation of cellular processes, single multicellular organism process, transcription factor activity, DNA binding.
- 26h (3 x $10^{-7} < P < 0.008$): detection of **pheromone**, sensory perception of chemical stimulus, response to pheromone, sensory perception, detection of chemical stimulus, extra cellular region / space (largely chemosensory genes), system / neurosystem process (chemosensory genes).
- 50h (0.001 < P < 0.009): membrane potential, fucose binding (2 genes only), neurological system process (chemosensory genes), various metabolic categories, system process (chemosensory genes), response to abiotic stimulus.

**Abdomen:**
- 2h (5 x $10^{-10} < P < 0.02$): **Immunity**, peptidases, extracellular space (~seminal fluid proteins)
- 26h (8 x $10^{-11} < P < 0.04$): **Immunity**, extra cellular space (~seminal fluid proteins)
- 50h (1 x $10^{-7} < P < 0.04$): **Immunity**, extra cellular space (~seminal fluid proteins)

(ii) GO enrichment support for keyword categories. For each keyword we calculated the number of differentially expressed genes in each body part as a function of the total (15,513), and the number of DE genes / total number of DE genes at each time point and body part, with the
associated standard GO enrichment $P$ value. The results show that the selected keywords have appropriate statistical support as enriched functional categories.

| olfactory          |          |          |
|--------------------|----------|----------|
|                    | count    | $P$ value| count | $P$ value |
| HT                 | 155/15512| AB       | 29/15512|
| 02h                | 14/359   | 3.14E-05 | 2/180 | 1.50E-01 |
| 26h                | 3/113    | 2.25E-02 | 3/164 | 3.34E-03 |
| 50h                | 4/153    | 2.52E-02 | 1/129 | 5.44E-01 |

| odorant            |          |          |
|--------------------|----------|----------|
|                    | count    | $P$ value| count | $P$ value |
| HT                 | 121/15512| AB       | 33/15512|
| 02h                | 13/359   | 8.71E-04 | 2/180 | 1.46E-01 |
| 26h                | 5/113    | 1.39E-06 | 2/164 | 5.11E-01 |
| 50h                | 6/153    | 3.16E-04 | 2/129 | 2.69E-01 |

| pheromone          |          |          |
|--------------------|----------|----------|
|                    | count    | $P$ value| count | $P$ value |
| HT                 | 111/15512| AB       |
| 02h                | 9/359    | 2.10E-02 |        |
| 26h                | 9/113    | 4.11E-18 |        |
| 50h                | 8/153    | 1.16E-07 |        |

| sensory            |          |          |
|--------------------|----------|----------|
|                    | count    | $P$ value| count | $P$ value |
| HT                 | 527/15512| AB       | 82/15512|
| 02h                | 43/359   | 9.82E-31 | 2/180 | 3.76E-01 |
| 26h                | 19/113   | 4.14E-74 | 2/164 | 1.54E-01 |
| 50h                | 15/153   | 1.22E-18 | 1/129 | 7.40E-01 |

| trichogen-all      |          |          |
|--------------------|----------|----------|
|                    | count    | $P$ value| count | $P$ value |
| HT                 | 599/15512| AB       | 155/15512|
| 02h                | 19/359   | 5.93E-02 | 9/180 | 3.08E-07 |
| 26h                | 4/113    | 6.54E-01 | 6/164 | 3.40E-04 |
| 50h                | 7/153    | 4.05E-01 | 3/129 | 2.51E-02 |
| Protein         | HT      | AB      | P value  | count | count | P value  |
|-----------------|---------|---------|----------|-------|-------|----------|
| trichogen-sensory | 75/15512 | 15/15512 |          | 0.02h | 0.02h |          |
|                 | 3/359   | 0/180   | 6.47E-01 |       |       | NA       |
|                 | 2/113   | 0/164   | 1.51E-01 |       |       | NA       |
|                 | 1/153   | 0/129   | 9.12E-01 |       |       | NA       |
| chitin          | 125/15512 | 36/15512 |          | 0.02h | 0.02h |          |
|                 | 21/359  | 5/180   | 3.66E-13 | 7.39E-05 |       |          |
|                 | 4/113   | 4/164   | 1.28E-03 | 1.53E-03 |       |          |
|                 | 5/135   | 2/129   | 3.88E-04 | 7.80E-02 |       |          |
| cuticle         | 165/15512 | 25/15512 |          | 0.02h | 0.02h |          |
|                 | 15/359  | 4/180   | 1.87E-04 | 1.31E-04 |       |          |
|                 | 2/113   | 4/164   | 6.43E-01 | 2.00E-02 |       |          |
|                 | 6/153   | 2/129   | 2.86E-04 | 3.60E-02 |       |          |
| GPCR            | 83/15512 | 40/15512 |          | 0.02h | 0.02h |          |
|                 | 25/359  | 7/180   | 2.19E-16 |       |       |          |
|                 | 0/113   | 4/164   | 9.62E-04 |       |       |          |
|                 | 5/153   | 7/129   | 4.95E-04 |       |       |          |
| seminal         | 40/15512 |          |          | 0.02h | 0.02h |          |
|                 | 7/180   | 3.83E-04 |          |       |       |          |
|                 | 4/164   | 9.62E-04 |          |       |       |          |
|                 | 7/129   | 4.53E-11 |          |       |       |          |
| proteolysis     | 243/15512 | 209/15512 |          |       |       |          |
In conclusion, there was strong overlap between standard GO terms and our keywords. This suggests that the keyword selection concisely captured informative biological categories. Statistical support for this assertion comes from: (i) the GO enrichment tests on the selected keyword terms, shown above, and (ii) the analysis of the distribution of DE in genes within each keyword category against a randomly drawn control distribution (Table S5).

REFERENCES

Davies MN, Meaburn EL, Schalkwyk LC. 2010. Gene set enrichment; a problem of pathways. Brief Funct Genomics 9:385-390.

Gene Ontology Reference Genome Project 2009. The Gene Ontology's Reference Genome Project: a unified framework for functional annotation across species. PLoS Comput Biol 5:e1000431.

Khatri P, Sirota M, Butte AJ. 2012. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol 8:e1002375.
SUPPLEMENTAL MATERIAL 2: ANALYSIS OF OVERLAP WITH PREVIOUS STUDIES

We tested for overlap of our gene expression data with related published work in 5 studies, as described below.

(1) Gene expression responses of whole males to a short, 5-minute burst of conspecific courtship delivery by males was previously analysed using array technology (Carney 2007). Ten genes were reported as up- and 33 down-regulated. For the closest comparability with our dataset we therefore tested for overlap of DE genes at 2h in both the HT and A.

(2) A second related study (Ellis and Carney 2011) provided evidence for DE of genes in male heads, again assessed using arrays. Males were subjected to 20 minutes of each of: (i) courting males (male placed with a cauterized female that could not mate), (ii) male-male exposure, and (iii) lone control male. We tested for overlap between these genes and our data, using the 2h_HT samples.

(3) We also tested whether we obtained gene expression changes in response to rivals in the set of genes identified to encode the non sperm male ejaculate proteins transferred to females during mating (Findlay et al. 2008).

(4) Our data could also usefully be compared with Fedorka et al. (2011) in which the expression of 3 seminal fluid protein and 4 testis genes in response to male-male competition was tested. qRT-PCR was conducted on 7 genes in whole males under low (1 male per vial) and high competition (4 males per vial). We made a direct comparison of these results to the 2, 26 and 50h_A gene expression data from our study.

(5) Finally, we compared the gene expression patterns we found with those of Wigby et al. (2009) in which increased transfer into females of the gene products of Acp26Aa and Acp70A following exposure of males to rivals was reported. The most relevant tissue and time periods for direct comparison were the 2 and 26h_A gene sets.

RESULTS

Validation by overlap of DE genes with previous studies (Table S8).

(1) There was little consistency in terms of the direction of difference for DE genes at the 2h timepoint in both the HT and A in our data and the genes changing in response to a 5-minute burst
of conspecific courtship delivery by males, as reported previously (Carney 2007) (Table S8A). For example, in the 33 genes denoted as significantly down-regulated in Carney (2007), the most consistent pattern in our data was in the opposite direction, i.e. there was some evidence for upregulation of 3 of these genes in the HT, with no evidence for any down regulated genes. Overall, the genes expressed by males in response to rivals by us were not generally the same as those previously detected by Carney (2007). This indicates that genes that change expression in males upon the short-term initiation of courtship are not the same as those used in responding to conspecific rivals.

(2) Among the 16 up regulated courtship-specific genes reported by Ellis and Carney (2011) we found 4 up regulated genes in common (data not shown). Among the socially responsive genes (those changing in response to males and females, but not courtship specific, Table S8B) there was minimal overlap. For example, among the 185 down regulated genes in Ellis and Carney (2011), we detected overlap but in the opposite direction, with evidence for up regulation in 6, and near threshold up regulation in a further ~17 genes, in the HT. As expected, there was little overlap with A genes. For the reported 80 up regulated genes (Table S8B), we found evidence for significant up regulation in 2 of these genes, and near threshold up regulation in a further ~5 genes, in the HT, with scant evidence for any down regulated genes and no evidence for overlap with our A genes.

Of the 145 genes significantly down regulated in response to 20 min of male-male contact in Ellis and Carney (2011) we found some overlap (Table S8C). However, as for the socially responsive genes, the direction of DE was opposite, with evidence for 2 (and a further ~3 near threshold) genes being up regulated at 2h in the HT, and minimal evidence of down regulation and no overlap of DE in the A. Of the 95 genes reported as up regulated following male-male exposure we found slightly better agreement, with evidence for overlap with 2 (and a further near threshold ~5-6) genes with increased expression in our 2h HT samples. There was scant evidence for overlap of with differential expression with the A genes.

Overall, there was limited overlap between the genes we identified as DE in response to rivals and those identified as socially responsive in the array study of Ellis and Carney (2011). Any overlap of DE was not consistently in the same direction. The degree of overlap of DE between the two studies was appropriate given the differences in temporal (20 mins versus 2h) and tissue (male heads versus male head + thorax) sampling. Reassuringly, we saw little overlap with our A sample data.

(3) We found ejaculate protein genes, including those described in Findlay et al. (2008) among our DE genes in the A samples, and a few also in the HT (Table S8D). For example, in the 2h A sample we found evidence for DE in at least 1 replicate in response to rivals in 7 of the seminal fluid protein genes in Findlay et al. (2008) (including Acp26Ab and 70A). At 26h, we found
evidence for DE in at least 1 replicate in 7 seminal fluid protein genes (including \textit{Acp36DE} and \textit{Acp76A}). By 50h the response in the genes reported in Findlay et al. (2008) included evidence for DE in at least 1 replicate in 6 of the genes in our data (including \textit{PebII}, \textit{Acp24A4} and \textit{CG33259}). The `seminal fluid protein' genes found in our HT dataset comprised a few odorant receptors and odorant binding proteins, suggesting that these genes have dual reproductive functions in and outside the male reproductive tract.

(3) We also assessed gene expression in the 7 genes studied by Fedorka et al. (2011) in our A data across all time points (Table S8E). In contrast to Fedorka et al. (2011) we found no increase in gene expression over time from 2 to 50h for any of the genes, in fact for two of the three seminal fluid protein genes there was a decrease over time. Consistent with Fedorka et al. (2011), there was no evidence for differential expression in the 4 sperm genes between males exposed or not to rivals. For the two seminal fluid protein genes (\textit{Acp26Aa}, \textit{Acp62F}) that developed decreased gene expression over time in the presence of rivals, we also found lower gene expression with rivals, although in our experiment this difference was present at all 3 time points. Also in line with Fedorka et al. (2011) we found no consistent pattern of DE in \textit{Acp70A} in response to competition / rivals. There was therefore reasonable concordance between the datasets.

(4) We compared the two genes studied by Wigby et al. (2009) with same two genes in our A samples at 2 and 26h (Table S8F). There was no difference in \textit{Acp26Aa} gene expression in our data. Hence the increased transfer of \textit{Acp26Aa} protein reported by Wigby et al. (2009) may be regulated post transcriptionally. For \textit{Acp70A} we found increased gene expression in the presence of rivals at 2h in one replicate set and at 26h in the other. This suggests that the increase in \textit{Acp70A} gene product transferred into females may at least be partly subject to transcriptional regulation.

Overall across the 5 studies compared, we therefore found appropriate agreement where expected, providing independent validation of our findings.

REFERENCES

Carney GE. 2007. A rapid genome-wide response to \textit{Drosophila melanogaster} social interactions. \textit{BMC Genomics} \textbf{8}:288.

Ellis LL, Carney GE. 2011. Socially-responsive gene expression in male \textit{Drosophila melanogaster} is influenced by the sex of the interacting partner. \textit{Genetics} \textbf{187}:157-169.

Fedorka KM, Winterhalter WE, Ware B. 2011. Perceived sperm competition intensity influences seminal fluid protein production prior to courtship and mating. \textit{Evolution} \textbf{65}:584-590.
Findlay GD, Yi X, Maccoss MJ, Swanson WJ. 2008. Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biol* 6:e178.

Wigby S, Sirot LK, Linklater JR, Buehner N, Calboli FC, et al. 2009. Seminal fluid protein allocation and male reproductive success. *Current Biology* 19:751-757.