Deep sequencing analysis of the circadian transcriptome of the jewel wasp *Nasonia vitripennis*

Nathaniel J. Davies and Eran Tauber[†]

Dept. of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

Running title: Sequencing the *Nasonia* circadian transcriptome

Keywords: Circadian clock; transcriptomics; *Nasonia;* RNAseq

[†] For correspondence. Dept. Genetics, University of Leicester, University Road, Leicester LE1 7RH, United Kingdom. Email: et22@le.ac.uk

Tel: +44-116- 252- 3455; Fax; +44 116 252 378

Abstract

The study of the circadian clock has benefited greatly from using *Drosophila* as a model system. Yet, accumulating evidence suggests that the fly might not be the canonical insect model. Here, we have analysed the circadian transcriptome of the Jewl wasp *Nasonia vitripennis* by using RNA-seq in both constant darkness and constant light (in contrast to flies, the wasps are rhythmic under continuous light). We identify approximately 6% of the transcriptome as cycling under constant conditions, revealing a bimodal distribution of phases and low cycling amplitude. We examine the functions under circadian control in *Nasonia*, identifying clock control of functions such as metabolism, light response, and a variety of neural processes, drawing comparisons between *Nasonia* and *Drosophila*. We characterise the transcriptional differences underlying phenotypic differences in free-running circadian behaviour in constant darkness and constant light, revealing significant down-regulation of catabolic processes in constant light. We also profile levels of opsins transcription, gaining insight into how *Nasonia* responds to light, which is a key question in circadian research in the Hymenoptera. Although there was little similarity between cycling genes in *Drosophila* and *Nasonia*, the functions fulfilled by cycling transcripts were similar in both species. Of the known *Drosophila* core clock genes, only *pdp1e*, *shaggy* and *Clock* showed a significant cycling in *Nasonia*, underscoring the importance of studying the clock in non-model organisms.

24 ## Introduction

25 The circadian clock regulates fundamental biological processes such as sleep
26 (Huang, *et al.* 2011), metabolism (Huang, *et al.* 2011), and the immune system
27 (Scheiermann, *et al.* 2013), and has implications for a wide range of human
28 diseases. Notable examples of diseases linked to the circadian clock include
29 cancer (Kelleher, *et al.* 2014), Alzheimer's disease (Musiek, *et al.* 2015),
30 cardiovascular disease (Takeda and Maemura, 2011), obesity (Maury, *et al.*
31 2010), diabetes (Maury, *et al.* 2010), and depression (Quera Salva, *et al.* 2011). A
32 primary output of the clock is circadian regulation of transcription, a trait which
33 has been demonstrated in mammals (Hughes, *et al.* 2009), insects (McDonald
34 and Rosbash, 2001a), plants (Schaffer, *et al.* 2001), and even bacteria (Woelfle
35 and Johnson, 2006). Therefore, analysing transcriptional oscillations in clock-
36 controlled genes (CCGs) is a key step in understanding how the daily rhythms
37 produced by the clock are ultimately linked to behavioural phenotypes.

38 The genetic mechanisms underlying the animal circadian clock were first
39 elucidated through studies of model animals; primarily the fruit fly *Drosophila*.
40 The first clock gene to be identified, *period (per)*, was discovered through
41 mapping the genetic basis of *Drosophila* mutants with aberrant locomotor and
42 eclosion rhythms (Konopka and Benzer, 1971). The discovery of *period* was
43 followed by the discovery of its heterodimeric partner *timeless* (*tim*) (Sehgal, *et*
44 *al.* 1994). These two genes are joined by a roster of other genes working together
45 to produce robust internal rhythms.

46 The discoveries made in *Drosophila* have been instrumental for
47 understanding the mechanisms of the circadian clock in mammals (Yu and
48 Hardin, 2006). As the principal insect model, *Drosophila* has been used to great

49 effect to model circadian phenomena in humans (Rosato, *et al.* 2006). However,
50 as circadian research into non-drosophilid insects has advanced, several
51 alternative clock models have been proposed (Yuan, *et al.* 2007), some of which
52 may better model aspects of the mammalian clock than *Drosophila*.

53 For example, a major difference between the various clock models in
54 insects concerns the light input pathway. The main light input to the clock in
55 *Drosophila* is mediated through *cryptochrome* (*cry1*) which is activated in
56 response to light (Ceriani, *et al.* 1999), binds to and promotes the degradation of
57 *tim* (Busza, *et al.* 2004), ultimately resulting in the degradation of *per* (Ko, *et al.*
58 2002,Grima, *et al.* 2002). In contrast, mammalian-like *cryptochrome* (*cry2*) is not
59 light-sensitive (Yuan, *et al.* 2007), but is a part of the core transcriptional
60 feedback loop suppressing its own transcription (and that of *per*) by interfering
61 with the actions of the CLK-BMAL1 heterodimer (Kume, *et al.* 1999,Jin, *et al.*
62 1999). Mammals also lack a homolog for *timeless*, possessing only a homolog of
63 the *Drosophila* gene *timeout* (Benna, *et al.* 2000), a gene whose potential role in
64 the clock is less clear and less crucial than that of *timeless* (Gustafson and Partch,
65 2015,Benna, *et al.* 2010).

66 The Lepidoptera harbour both types of *cryptochrome* (*Drosophila*-like
67 *cry1* and mammal-like *cry2*) (Tomioka and Matsumoto, 2010), as well as
68 homologs of *timeless* and *timeout* (Tomioka and Matsumoto, 2015). The two
69 cryptochromes have been shown to act in a similar way to their *Drosophila* and
70 mammal counterparts; *cry1* functions as a light receptor and *cry2* serves as a
71 transcriptional repressor (Zhu, *et al.* 2008).

72 Of the major insect orders, the Hymenoptera arguably possess the most
73 mammalian-like core clock architecture, possessing *cry2* and *timeout* but neither

74 *cry1* nor *timeless* (Tomioka and Matsumoto, 2015,Yuan, *et al.* 2007). In addition
75 to these molecular similarities, there is evidence that the transcriptional profiles
76 of these genes match more closely the mammalian model than the *Drosophila*
77 model (Rubin, *et al.* 2006). Light-entrained circadian rhythms have been
78 demonstrated in the Hymenoptera, but the question of light detection in the
79 Hymenopteran clock remains an open one.

80 *Nasonia vitripennis* is a parasitoid wasp, which as a research model offers
81 advantages over other hymenopterans, including a fully sequenced genome
82 (Werren, *et al.* 2010), systemic RNAi (Lynch and Desplan, 2006), a robust and
83 well-characterised circadian response (Bertossa, *et al.* 2013), a fully functional
84 DNA methylation kit (Park, *et al.* 2011), and a history as a model for
85 photoperiodism (Saunders, 1969).

86 In this study, we advance *Nasonia* as an alternative circadian model by
87 using RNA-seq to profile whole-transcriptome gene expression in the *Nasonia*
88 head. As the *Nasonia* clock free-runs in both constant darkness and constant light
89 (Figure 1), we profiled both of these conditions to examine how the two
90 circadian transcriptomes differ. To our knowledge, this is the first circadian
91 RNA-seq study performed in an insect other than *Drosophila*, and the first study
92 to profile the circadian transcriptome oscillating under constant light.

93 Results

94 Identifying rhythmic transcription

95 We first performed an unbiased clustering analysis to ascertain the kinds of
96 expression patterns present in the data. To this end, Mfuzz (Kumar and E
97 Futschik, 2007) was used to carry soft c-means clustering, a method which is less

98 sensitive to biological noise than traditional clustering (Futschik and Carlisle,
99 2005). After filtering (see Methods), thirty clusters were generated for each
100 condition (Supplementary figures S1 and S2), revealing a variety of potentially
101 rhythmic and non-rhythmic expression trends. Potential asymmetric wave forms
102 were detected in LL (e.g. Supplementary figure S2, clusters 22 and 26).

103 To identify rhythmic transcripts, we used the RAIN algorithm (Thaben
104 and Westermark, 2014). At false discovery rate (FDR) threshold of 0.1 we
105 identified 1,057 rhythmic transcripts in DD and 929 in LL (Table S1, S2).

106 Rhythmic transcripts ($q < 0.1$) were sorted by phase, peak shape, and
107 significance, and plotted (Figure 2A). Examining the phase distribution (Figure
108 2B), it is apparent that the majority of transcripts show peak expression early in
109 the subjective morning/afternoon or in the subjective night, with fewer
110 transcripts peaking at intermediate times. This disparity in phase is greater in
111 the transcripts which show rhythmic expression in both DD and LL; less than
112 12% of transcripts in DD and less than 5% in LL show peak expression at
113 intermediate times (Figure 2B). The majority of these transcripts (~87%) exhibit
114 a similar (+-4 hrs) phase in LL to their phase in DD.

115 Similarly to *Drosophila* (Hughes, *et al.* 2012) and mammals (Hughes, *et al.*
116 2009), the majority of transcripts show only small cyclic changes in expression
117 amplitude over the day; over 80% of reliably quantified (see Methods)
118 transcripts in both conditions have amplitudes (peak expression divided by
119 trough expression) of 2 or less. In both DD and LL, transcripts with exceptionally
120 high amplitudes (> 4) are transcripts with unusually low or high measurements
121 at isolated time-points with no obvious specific shared function. This is in
122 contrast with results in *Drosophila* and mammals, where some core clock genes

123 exhibit very high amplitude oscillations (Hughes, *et al.* 2009,Hughes, *et al.*
124 2012,Li, *et al.* 2015).
125
126 **Canonical clock genes and comparison with *Drosophila***
127 The canonical clock genes were examined for rhythmicity both at the transcript
128 level and via an additional RAIN analysis at the gene level. The q-values (FDR
129 adjusted p-values) for the canonical clock genes are shown in supplementary
130 table S3. We found a rather limited evidence for rhythmicity in these genes
131 which included *pdp1e* (q ~ 0.1, LL and DD), *shaggy* (q < 0.1, DD), and *Clok* (q ~
132 0.1, LL). At a less stringent FDR (q< 0.2), *per*, *cyc*, *Dbt* and *cwo* were rhythmic in
133 DD, while *cry* and *cyc,* were oscillating in LL. For comparison between splice
134 variants and conditions, median expression levels of the canonical clock genes
135 and their transcripts for both DD and LL are shown in supplementary table S4.
136 We compared the transcripts identified as cycling in *Nasonia* heads with
137 the transcripts identified as cycling in *Drosophila* heads. For these purposes, we
138 used a list of genes identified in a meta-analysis study of *Drosophila* circadian
139 microarray data as being rhythmically expressed in either LD or DD (Keegan, *et*
140 *al.* 2007). Of 173 genes identified as rhythmic in *Drosophila*, 33 genes
141 (Supplementary table S5) were found to also be rhythmic in *Nasonia* (either in
142 LL or DD, q < 0.1), no more than would be expected by chance (p = 0.11,
143 hypergeometric test).
144
145 **Functions of rhythmic genes**
146 To capture the general functions that rhythmic genes may fulfil in *Nasonia*, we
147 tested a broader set of rhythmic genes (FDR < 0.2 in RAIN) for GO term

148 overrepresentation (Davies and Tauber, 2015a), revealing 94 GO terms
149 overrepresented for genes rhythmic in DD (including 'response to light stimulus',
150 'proteasome complex', and 'generation of neurons', Supplementary table S6) and
151 123 terms for genes rhythmic in LL (including 'locomotion','proteasome
152 complex', and 'response to external stimulus', Supplementary table S7), 25 of
153 which were shared between both conditions (Figure 3). Shared terms include
154 terms related to neurons, signal transmission, and responses to stimuli. Notably,
155 all four Nasonia opsins were found to exhibit similar transcriptional profiles in
156 LL and DD, with low expression in the morning and high expression in the
157 evening.
158 It has previously been demonstrated that the timing of different (or indeed
159 opposing) biological processes can be controlled through the circadian
160 regulation of groups of genes (Sancar, *et al.* 2015,Zhang, *et al.* 2014).
161 Unsupervised clustering methods have previously been established as a useful
162 method for functional characterisation of circadian genes (Nguyen, *et al.* 2014).
163 To establish whether temporal separation of functions occurs in *Nasonia*, we
164 therefore returned to the expression clustering analysis. Firstly, we employed
165 hypergeometric tests to identify clusters with an overrepresentation of rhythmic
166 genes (Figure 4, Supplementary table S8 and S9). Clusters which were found to
167 have a significant rhythmic component ($q < 0.05$, supplementary tables S8 and
168 S9) were analysed for overrepresented GO terms. Examples of clusters with
169 enriched functions include clusters DD7 and LL20 which are significantly
170 enriched for catalytic activity GO terms, especially genes involved in the
171 proteasome, and clusters DD24 and LL6 which are both involved in circadian
172 and neural processes. Other clusters (DD1 and DD2) did not turn up any

173 overrepresented GO terms and are thus likely comprised of genes with a wide
174 range of functions.
175
176 **Transcriptional differences between constant darkness and constant light**
177 To examine whether differences in circadian period seen in locomotor activity
178 between DD and LL could also be detected in transcriptional rhythms, we fitted
179 parametric models with a range of periods to transcripts rhythmic in both
180 conditions ($q < 0.1$). For those transcripts with statistically significant fits to the
181 model in both conditions ($q < 0.1$, see Methods), we took the period with the best
182 fit and compared these periods between conditions. Overall, transcripts in LL
183 showed a significantly ($p < 3.9e-09$, Wilcoxon rank sum test) shorter (median
184 24) period than those in DD (median 25.4), mirroring the behavioural
185 differences in period.
186 We have also tested for differential expression between DD and LL. In the
187 absence of biological replicates, we analysed differential expression using a fold-
188 change approach. We used 1.5 fold change as a cut-off for differential expression
189 (Dalman, *et al.* 2012), yielding 1,488 genes expressed higher in DD than LL and
190 971 genes expressed higher in LL than DD (Figure 5). Genes more highly
191 expressed in DD were significantly enriched ($q < 0.01$) for genes involved in
192 various forms of catalytic activity (Supplementary table S10), including the vast
193 majority of proteasome genes (>75%). Genes more highly expressed in LL were
194 enriched for a small number of terms including 'plasmalemma' and 'sequence-
195 specific DNA binding' (Supplementary table S11).

196 ## Discussion

197 This study provides the first insights into global transcriptional oscillation in
198 *Nasonia*. With RNA-seq, we profiled the circadian transcription of >26,000
199 transcripts in *Nasonia* in either DD or LL. At a relatively stringent FDR ($q < 0.1$),
200 we identified 1,057 cycling transcripts in DD and 929 cycling transcripts in LL.
201 These transcripts correspond to a cycling fraction of 6.7% and 5.9% of all
202 transcripts analysed in DD and LL respectively. These figures are comparable to
203 cycling fractions reported in various organisms and tissues, generally between
204 2% and 10% of the transcriptome (Michael and McClung, 2003).

205 In both conditions, cycling transcripts were found to cycle at low
206 amplitudes (mostly < 2 fold) and with a limited, bimodal, range of phases. This is
207 in contrast to microarray/RNA-seq studies in *Drosophila*, where transcripts were
208 found to cycle with a broader range of phases (Rodriguez, *et al.* 2013) and
209 studies in both mammals and *Drosophila*, which have identified a group of high-
210 amplitude (> 4-fold) cycling genes among the low-amplitude majority (Akhtar, *et*
211 *al.* 2002). High amplitude cyclers typically include clock genes (Akhtar, *et al.*
212 2002,Hughes, *et al.* 2012). The low oscillations of the *Nasonia* head
213 transcriptome render the expression profiles of the canonical clock genes
214 difficult to resolve (Covington, *et al.* 2008). This issue may also contribute to the
215 discordance between the various circadian microarray studies in *Drosophila*
216 (Keegan, *et al.* 2007).

217 An emerging property of the circadian transcriptome in *Nasonia* is the
218 temporal separation of function by phase (Fig 2). Notably, genes involved in
219 catalytic activity were strongly overrepresented in morning-peaking transcripts.
220 This is in line with other studies which show catalytic activity confined to the

221 morning in fungi (Sancar, *et al.* 2015), in agreement with a general observation
222 that an important (or even primary) function of circadian clocks (Hurley, *et al.*
223 2015) is to temporally separate catabolism and anabolism. Although we did not
224 detect an overrepresentation of anabolic genes within the cyclic transcripts,
225 expression clusters DD10 and LL24 (Supplementary figures S1 and S2) did show
226 strong overrepresentation (Supplementary tables S12 and S13) for genes
227 involved in cytosolic ribosomal genes (q < 3.e-56) and cellular anabolism (q < 2e-
228 06). These clusters exhibit an antagonistic expression pattern to the expression
229 clusters containing the catabolic genes, suggesting that catabolism and
230 anabolism are indeed separated by the circadian clock in *Nasonia*.
231 The comparison of expression between LL and DD reveals that a majority
232 of genes involved in the proteasome and a broader set of genes involved in
233 catabolism, are more highly expressed in DD than LL. As turnover rates of clock
234 proteins have shown to be coupled with changes in the circadian period (Syed, *et*
235 *al.* 2011,He and Liu, 2005), up-regulation of the proteasome may provide an
236 explanation for differences in period observed between DD and LL.
237 Although the genes which cycle in *Drosophila* largely differ from those
238 cycling in *Nasonia*, the functions fulfilled by CCGs in *Nasonia* are similar to the
239 functions filled by CCGs in *Drosophila*. Examples of functions shared by CCGs in
240 the *Drosophila* and *Nasonia* heads are: various aspects of metabolism (Rodriguez,
241 *et al.* 2013,Ueda, *et al.* 2002,Ceriani, *et al.* 2002,Claridge-Chang, *et al.* 2001),
242 phototransduction (Ueda, *et al.* 2002,Rodriguez, *et al.* 2013), synaptic/nervous
243 functions (McDonald and Rosbash, 2001b,Ceriani, *et al.* 2002,Claridge-Chang, *et*
244 *al.* 2001), oxidoreductase activity (Claridge-Chang, *et al.* 2001), mating behaviour

245 (Rodriguez, *et al.* 2013), and immunity (McDonald and Rosbash, 2001b,Ceriani,
246 *et al.* 2002).
247 We identified cycling of genes involved in response to light, particularly
248 all four *Nasonia* opsins. These opsins, along with associated gPCRs, cycle with a
249 similar phase and are all more highly expressed in LL than in DD (Supplementary
250 figure S6). Daily and circadian changes in opsin expression have been
251 demonstrated in other organisms (e.g. mice (Bowes, *et al.* 1988) , zebrafish (Li, *et*
252 *al.* 2005), honeybee (Sasagawa, *et al.* 2003)), and opsin expression is generally
253 found to be up-regulated in response to light (Yan, *et al.* 2014). Characterising
254 the opsins in *Nasonia* is likely to provide insights into the light input pathway
255 into the clock, particularly as *Nasonia* does not possess other obvious light input
256 candidate genes such as *Drosophila*-like *CRY1* (Bertossa, *et al.* 2014) or *Pteropsin*
257 (Velarde, *et al.* 2005) (Supplementary figure S6).

258 **Data availability**

259 We have made the expression profile for each transcript in both conditions
260 available on WaspAtlas (Davies and Tauber, 2015b). Data have been archived in
261 the NCBI short read archive (SRA), with accession number PRJNA318159.

262 Methods

263 **Maintenance and sample collection**

264 Stocks of *Nasonia vitripennis* (strain AsymCX) were maintained at 25°C on
265 blowfly pupal hosts in 12:12 light:dark cycles. To obtain male wasps for
266 experiments, groups of eight females were isolated at the yellow pupal stage and
267 transferred onto fresh hosts upon eclosion. The resulting male progeny were
268 collected upon eclosion and moved onto vials with a 30% sucrose agar medium,

269 in groups of 20. During entrainment (four full days in an LD 12:12 cycle) and
270 collection, wasps were kept in four light boxes in the same incubator at 19°C.
271 Starting at CT1, wasps were collected every four hours and snap-frozen in liquid
272 nitrogen and immediately transferred to -80°C. Wasps were collected
273 sequentially from light box to light box every four hours to minimise disturbance
274 of wasps, and so that wasps were collected from each light box once every 16
275 hours, thereby minimising the effect of variations within light boxes.
276 Temperature and light recordings were taken during the experiment, and can be
277 viewed in Supplementary file S2. To verify that wasps entrained correctly to the
278 experimental conditions and that free-running behaviour was as expected,
279 individual male wasps were isolated and locomotor activity was monitored.
280 Behavioural recordings of individual male wasps in experimental conditions can
281 be seen in Supplementary figure S7, ruling out behavioural differences caused by
282 inter light box variations in light intensity in LL, though not transcriptional
283 differences.
284
285 **RNA extraction, sequencing, and read mapping**
286 RNA was extracted from pooled groups of 50 heads for each sample, using Trizol
287 RNA extraction protocol, and followed by clean-up using the RNAeasy spin
288 column kit (Qiagen). Samples were polyA selected and sequenced at Glasgow
289 Polyomics (University of Glasgow, United Kingdom) on the Illumina NextSeq500
290 platform, resulting in approximately 20 million 75bp paired-end reads per
291 sample.
292 Read mapping was achieved with Tophat2 (v2.1.0)(Trapnell, *et al.* 2012)
293 against the *Nasonia* Nvit_2.1 NCBI annotation. As the purpose of this study was

294 not to identify novel splice variants or improve on existing annotation, novel
295 junction detection was disabled for accurate quantification of known transcripts.
296 Mean mapping efficiency was above 90% for both conditions (Supplementary
297 table S14). Read quantification was performing using the DEseq normalisation
298 method (Anders and Huber, 2010). All 24 samples from both conditions were
299 grouped together to allow comparison between as well as within conditions.
300
301 **Expression profile clustering**
302 Isoform expression profiles were first filtered to include only those isoforms
303 with no missing values at any time-point in either condition. Expression values
304 were standardised using the 'Standardise' function in Mfuzz (Kumar and E
305 Futschik, 2007). The 'cselection' function in Mfuzz was used to select an
306 appropriate c-value for the c-means clustering (default parameters; m=1.25).
307 Based on this analysis, thirty fuzzy clusters were generated for each condition
308 using the fuzzification parameter m=1.25.
309
310 **Rhythmic expression analysis**
311 RAIN (Thaben and Westermark, 2014) was used on all filtered isoforms (i.e.
312 those with no missing values at any timepoint) in either condition to detect
313 rhythmic isoforms at a period of 24 hours. As a non-parametric method, RAIN
314 only facilitates detection of rhythmic isoforms with periods which are a multiple
315 of the sample resolution (in this case 4 hr). The p-values produced by RAIN were
316 corrected to q-values using the Benjamini-Hochberg method (Benjamini and
317 Hochberg, 1995). This method was repeated using expression values for genes

318 rather than transcripts for the clock gene analysis (i.e. the summed expression
319 values for all known transcripts of a particular gene).
320 Maximum fold changes in expression were calculated by normalising per-
321 condition expression values by the median value and calculating the ratio from
322 the lowest expression over 48 hours to the highest. Reliably quantified
323 transcripts are defined as those those transcripts where the absolute FPKM
324 value is 5 or above at all timepoints, the threshold for this set at a similar level to
325 other analyses (Hughes, *et al.* 2012).
326 To analyse the period of rhythmic transcripts, we fitted parametric
327 waveforms with a variety of periods (20 to 28 hrs in steps of 0.2 hrs) to all
328 transcripts identified as rhythmic ($q < 0.1$) in both conditions. This FDR
329 threshold is in line with, or more strict, than thresholds chosen in other similar
330 studies (Hughes, *et al.* 2012,Huang, *et al.* 2013,Keegan, *et al.* 2007). Those
331 transcripts (85 in total) which showed a significant ($q < 0.1$) fit to the model in
332 both conditions were analysed in terms of their best fitting period.
333 GO term overrepresentation was performed in WaspAtlas (Davies and
334 Tauber, 2015b) using the Nvit_2.1 NCBI annotation dataset. All hypergeometric
335 tests were performed within R using the 'phyper' function. Clusters with
336 rhythmic components were identified by collapsing the fuzzy clusters into hard
337 clusters using the 'cluster' property of the Mfuzz object, performing
338 hypergeometric tests to identify clusters with enrichment for rhythmic
339 transcripts. Thirty tests were performed for each condition (i.e. for all clusters),
340 and were corrected per-condition using the Benjamini-Hochberg method in R (R
341 Development Core Team, 2008).

342 For comparison to microarray studies, orthologs for *Drosophila*
343 *melanogaster* were obtained from a meta-study of circadian microarray data
344 (Keegan, *et al.* 2007). The 214 obtained FlyBase identifiers were converted to the
345 latest identifiers using the validation tool, resulting in 218 unique identifiers (the
346 increase in identifiers can be attributed to previous identifiers referring to
347 multiple genes in the current annotation). Orthologs for these *Drosophila* genes
348 were obtained through WaspAtlas, retrieving orthologs for 135 genes which
349 mapped to 173 unique *Nasonia* genes due to gene duplications, etc. This set of
350 173 genes was compared with the number of genes with rhythmic transcripts
351 that would be expected by chance using a hypergeometric test.
352
353 **Phylogenetic analysis of opsin genes**
354 Opsin genes were searched for using NCBI BLASTP using six species; *Apis*
355 *mellifera*, *Bombyx mori*, *Drosophila melanogaster*, *Mus musculus*, *Nasonia*
356 *vitripennis*, and *Homo sapiens*, using the *Nasonia Lop1* protein sequence as a
357 query. BLAST results were inspected and 7e-19 was chosen as an appropriate
358 cut-off to include all opsin sequences. Sequences were aligned by ClustalW in
359 MEGA (Tamura, *et al.* 2007) and a maximum likelihood tree generated using
360 default parameters. Duplicated sequences were manually removed, and
361 sequences renamed for display on the tree. Full protein name to shortened
362 display name translations can be found in supplementary table S15.
363

364 ## References

365 Akhtar RA, Reddy AB, Maywood ES, Clayton JD, King VM, Smith AG, Gant TW,
366 Hastings MH, Kyriacou CP. 2002. Circadian cycling of the mouse liver
367 transcriptome, as revealed by cDNA microarray, is driven by the
368 Suprachiasmatic Nucleus. *Current Biology,* **12:** 540-550.

369 Anders S and Huber W. 2010. Differential expression analysis for sequence count
370 data. *Genome Biol.* **11:** R106-2010-11-10-r106. Epub 2010 Oct 27.

371 Benjamini Y and Hochberg Y. 1995. Controlling the False Discovery Rate: A
372 Practical and Powerful Approach to Multiple Testing. *Journal of the Royal*
373 *Statistical Society.Series B (Methodological)* **57:** 289-300.

374 Benna C, Bonaccorsi S, Wulbeck C, Helfrich-Forster C, Gatti M, Kyriacou CP, Costa
375 R, Sandrelli F. 2010. Drosophila timeless2 is required for chromosome stability
376 and circadian photoreception. *Curr. Biol.* **20:** 346-352.

377 Benna C, Scannapieco P, Piccin A, Sandrelli F, Zordan M, Rosato E, Kyriacou CP,
378 Valle G, Costa R. 2000. A second timeless gene in Drosophila shares greater
379 sequence similarity with mammalian tim. *Curr. Biol.* **10:** R512-3.

380 Bertossa RC, van de Zande L, Beukeboom LW, Beersma DG. 2014. Phylogeny and
381 oscillating expression of period and cryptochrome in short and long
382 photoperiods suggest a conserved function in Nasonia vitripennis. *Chronobiol.*
383 *Int.* **31:** 749-760.

384 Bertossa RC, van Dijk J, Diao W, Saunders D, Beukeboom LW, Beersma DG. 2013.
385 Circadian rhythms differ between sexes and closely related species of Nasonia
386 wasps. *PLoS One* **8:** e60167.

387 Bowes C, van Veen T, Farber DB. 1988. Opsin, G-protein and 48-kDa protein in
388 normal and rd mouse retinas: developmental expression of mRNAs and proteins
389 and light/dark cycling of mRNAs. *Exp. Eye Res.* **47:** 369-390.

390 Busza A, Emery-Le M, Rosbash M, Emery P. 2004. Roles of the two Drosophila
391 CRYPTOCHROME structural domains in circadian photoreception. *Science* **304:**
392 1503-6.

393 Ceriani MF, Darlington TK, Staknis D, Mas P, Petti AA, Weitz CJ, Kay SA. 1999.
394 Light-dependent sequestration of TIMELESS by CRYPTOCHROME. *Science* **285:**
395 553-6.

396 Ceriani MF, Hogenesch JB, Yanovsky M, Panda S, Straume M, Kay SA. 2002.
397 Genome-Wide Expression Analysis in Drosophila Reveals Genes Controlling
398 Circadian Behavior. *J. Neurosci.* **22:** 9305-9319.

399 Claridge-Chang A, Wijnen H, Naef F, Boothroyd C, Rajewsky N, Young MW. 2001.
400 Circadian regulation of gene expression systems in the *Drosophila* head. *Neuron*
401 **32:** 657-671.

402 Covington MF, Maloof JN, Straume M, Kay SA, Harmer SL. 2008. Global
403 transcriptome analysis reveals circadian regulation of key pathways in plant
404 growth and development. *Genome Biol.* **9:** R130.

405 Dalman MR, Deeter A, Nimishakavi G, Duan ZH. 2012. Fold change and p-value
406 cutoffs significantly alter microarray interpretations. *BMC Bioinformatics* **13**
407 **Suppl 2:** S11-2105-13-S2-S11.

408 Davies NJ and Tauber E. 2015a. WaspAtlas: a *Nasonia vitripennis* gene database
409 and analysis platform. *Database (Oxford)* **2015:** bav103. Print 2015.

410 Davies NJ and Tauber E. 2015b. WaspAtlas: a Nasonia vitripennis gene database
411 and analysis platform. *Database (Oxford)* **2015:** 10.1093/database/bav103. Print
412 2015.

413 Futschik ME and Carlisle B. 2005. Noise-robust soft clustering of gene expression
414 time-course data. *J. Bioinform Comput. Biol.* **3:** 965-988.

415 Grima B, Lamouroux A, Chelot E, Papin C, Limbourg-Bouchon B, Rouyer F. 2002.
416 The F-box protein slimb controls the levels of clock proteins period and timeless.
417 *Nature* **420:** 178-82.

418 Gustafson CL and Partch CL. 2015. Emerging models for the molecular basis of
419 mammalian circadian timing. *Biochemistry* **54:** 134-149.

420 He Q and Liu Y. 2005. Degradation of the Neurospora circadian clock protein
421 FREQUENCY through the ubiquitin-proteasome pathway. *Biochem. Soc. Trans.*
422 **33:** 953-956.

423 Huang W, Ramsey KM, Marcheva B, Bass J. 2011. Circadian rhythms, sleep, and
424 metabolism. *J. Clin. Invest.* **121:** 2133-2141.

425 Huang Y, Ainsley JA, Reijmers LG, Jackson FR. 2013. Translational profiling of
426 clock cells reveals circadianly synchronized protein synthesis. *PLoS Biol.* **11:**
427 e1001703.

428 Hughes ME, DiTacchio L, Hayes KR, Vollmers C, Pulivarthy S, Baggs JE, Panda S,
429 Hogenesch JB. 2009. Harmonics of circadian gene transcription in mammals.
430 *PLoS Genet.* **5:** e1000442.

431 Hughes ME, Grant GR, Paquin C, Qian J, Nitabach MN. 2012. Deep sequencing the
432 circadian and diurnal transcriptome of Drosophila brain. *Genome Res.* **22:** 1266-
433 1281.

434 Hurley JM, Loros JJ, Dunlap JC. 2015. The circadian system as an organizer of
435 metabolism. *Fungal Genet. Biol.*

436 Jin X, Shearman LP, Weaver DR, Zylka MJ, de Vries GJ, Reppert SM. 1999. A
437 molecular mechanism regulating rhythmic output from the suprachiasmatic
438 circadian clock. *Cell* **96:** 57-68.

439 Keegan KP, Pradhan S, Wang JP, Allada R. 2007. Meta-analysis of Drosophila
440 circadian microarray studies identifies a novel set of rhythmically expressed
441 genes. *PLoS Comput. Biol.* **3:** e208.

442 Kelleher FC, Rao A, Maguire A. 2014. Circadian molecular clocks and cancer.
443 *Cancer Lett.* **342:** 9-18.

444 Ko HW, Jiang J, Edery I. 2002. Role for Slimb in the degradation of Drosophila
445 Period protein phosphorylated by Doubletime. *Nature* **420:** 673-678.

446 Konopka R and Benzer S. 1971. Clock mutants of *Drosophila melanogaster*. *Proc.*
447 *Natl. Acad. Sci. USA* **68:** 2112-2116.

448 Kumar L and E Futschik M. 2007. Mfuzz: a software package for soft clustering of
449 microarray data. *Bioinformation* **2:** 5-7.

450 Kume K, Zylka MJ, Sriram S, Shearman LP, Weaver DR, Jin X, Maywood ES,
451 Hastings MH, Reppert SM. 1999. mCRY1 and mCRY2 are essential components of
452 the negative limb of the circadian clock feedback loop. *Cell* **98:** 193-205.

453 Li, J., Grant, G. R., Hogenesch, J. B., and Hughes, M. E. 2015. Chapter Sixteen -
454 Considerations for RNA-seq Analysis of Circadian Rhythms. In *Methods in*
455 *Enzymology* (ed. Amita Sehgal), pp. 349-367. Academic Press, .

456 Li P, Temple S, Gao Y, Haimberger TJ, Hawryshyn CW, Li L. 2005. Circadian
457 rhythms of behavioral cone sensitivity and long wavelength opsin mRNA
458 expression: a correlation study in zebrafish. *J. Exp. Biol.* **208:** 497-504.

459 Lynch JA and Desplan C. 2006. A method for parental RNA interference in the
460 wasp *Nasonia vitripennis*. *Nat. Protoc.* **1:** 486-494.

461 Maury E, Ramsey KM, Bass J. 2010. Circadian Rhythms and Metabolic Syndrome:
462 From Experimental Genetics to Human Disease. *Circulation Research* **106:** 447-
463 462.

464 McDonald MJ and Rosbash M. 2001a. Microarray analysis and organization of
465 circadian gene expression in *Drosophila*. *Cell* **107:** 567-578.

466 McDonald MJ and Rosbash M. 2001b. Microarray analysis and organization of
467 circadian gene expression in Drosophila. *Cell* **107:** 567-578.

468 Michael TP and McClung CR. 2003. Enhancer trapping reveals widespread
469 circadian clock transcriptional control in Arabidopsis. *Plant Physiol.* **132:** 629-
470 639.

471 Musiek ES, Xiong DD, Holtzman DM. 2015. Sleep, circadian rhythms, and the
472 pathogenesis of Alzheimer disease. *Exp. Mol. Med.* **47:** e148.

473 Nguyen TT, Mattick JS, Yang Q, Orman MA, Ierapetritou MG, Berthiaume F,
474 Androulakis IP. 2014. Bioinformatics analysis of transcriptional regulation of
475 circadian genes in rat liver. *BMC Bioinformatics* **15:** 83-2105-15-83.

476 Park J, Peng Z, Zeng J, Elango N, Park T, Wheeler D, Werren JH, Yi SV. 2011.
477 Comparative analyses of DNA methylation and sequence evolution using
478 *Nasonia* genomes. *Mol. Biol. Evol.* **28:** 3345-3354.

479 Quera Salva MA, Hartley S, Barbot F, Alvarez JC, Lofaso F, Guilleminault C. 2011.
480 Circadian rhythms, melatonin and depression. *Curr. Pharm. Des.* **17:** 1459-1470.

481 R Development Core Team. 2008. *R: A Language and Environment for Statistical*
482 *Computing.* R Foundation for Statistical Computing, Vienna, Austria.

483 Rodriguez J, Tang CH, Khodor YL, Vodala S, Menet JS, Rosbash M. 2013. Nascent-
484 Seq analysis of Drosophila cycling gene expression. *Proc. Natl. Acad. Sci. U. S. A.*
485 **110:** E275-84.

486 Rosato E, Tauber E, Kyriacou CP. 2006. Molecular genetics of the fruit-fly
487 circadian clock. *Eur. J. Hum. Genet.* **14:** 729-38.

488 Rubin EB, Shemesh Y, Cohen M, Elgavish S, Robertson HM, Bloch G. 2006.
489 Molecular and phylogenetic analyses reveal mammalian-like clockwork in the
490 honey bee (Apis mellifera) and shed new light on the molecular evolution of the
491 circadian clock. *Genome Res.* **16:** 1352-1365.

492 Sancar C, Sancar G, Ha N, Cesbron F, Brunner M. 2015. Dawn- and dusk-phased
493 circadian transcription rhythms coordinate anabolic and catabolic functions in
494 Neurospora. *BMC Biol.* **13:** 17-015-0126-4.

495 Sasagawa H, Narita R, Kitagawa Y, Kadowaki T. 2003. The expression of genes
496 encoding visual components is regulated by a circadian clock, light environment
497 and age in the honeybee (Apis mellifera). *Eur. J. Neurosci.* **17:** 963-970.

498 Saunders DS. 1969. Diapause and photoperiodism in the parasitic wasp Nasonia
499 vitripennis, with special reference to the nature of the photoperiodic clock. *Symp.*
500 *Soc. Exp. Biol.* **23:** 301-29.

501 Schaffer R, Landgraf J, Accerbi M, Simon V, Larson M, Wisman E. 2001.
502 Microarray analysis of diurnal and circadian-regulated genes in Arabidopsis.
503 *Plant Cell* **13:** 113-123.

504 Scheiermann C, Kunisaki Y, Frenette PS. 2013. Circadian control of the immune
505 system. *Nat. Rev. Immunol.* **13:** 190-198.

506 Sehgal A, Price J, Man B, Young M. 1994. Loss of circadian behavioral rhythms
507 and per RNA oscillations in the Drosophila mutant timeless. *Science* **263:** 1603-
508 1606.

509 Syed S, Saez L, Young MW. 2011. Kinetics of doubletime kinase-dependent
510 degradation of the Drosophila period protein. *J. Biol. Chem.* **286:** 27654-27662.

511 Takeda N and Maemura K. 2011. Circadian clock and cardiovascular disease. *J.*
512 *Cardiol.* **57:** 249-256.

513 Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary
514 Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24:** 1596-1599.

515 Thaben PF and Westermark PO. 2014. Detecting rhythms in time series with
516 RAIN. *J. Biol. Rhythms* **29:** 391-400.

517 Tomioka K and Matsumoto A. 2010. A comparative view of insect circadian clock
518 systems. *Cell Mol. Life Sci.* **67:** 1397-1406.

519 Tomioka K and Matsumoto A. 2015. Circadian molecular clockworks in non-
520 model insects. *Current Opinion in Insect Science; Insect genomics * Development*
521 *and regulation* **7:** 58-64.

522 Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL,
523 Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of
524 RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7:** 562-578.

525 Ueda HR, Matsumoto A, Kawamura M, Iino M, Tanimura T, Hashimoto S. 2002.
526 Genome-wide Transcriptional Orchestration of Circadian Rhythms in Drosophila.
527 *J. Biol. Chem.* **277:** 14048-14052.

528 Velarde RA, Sauer CD, O. Walden KK, Fahrbach SE, Robertson HM. 2005.
529 Pteropsin: A vertebrate-like non-visual opsin expressed in the honey bee brain.
530 *Insect Biochem. Mol. Biol.* **35:** 1367-1377.

531 Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Nasonia
532 Genome Working Group, Werren JH, Richards S, Desjardins CA, et al. 2010.
533 Functional and evolutionary insights from the genomes of three parasitoid
534 *Nasonia* species. *Science* **327:** 343-348.

535 Woelfle MA and Johnson CH. 2006. No promoter left behind: global circadian
536 gene expression in cyanobacteria. *J. Biol. Rhythms* **21:** 419-431.

537 Yan S, Zhu J, Zhu W, Zhang X, Li Z, Liu X, Zhang Q. 2014. The expression of three
538 opsin genes from the compound eye of Helicoverpa armigera (Lepidoptera:
539 Noctuidae) is regulated by a circadian clock, light conditions and nutritional
540 status. *PLoS One* **9:** e111683.

541 Yu W and Hardin PE. 2006. Circadian oscillators of Drosophila and mammals. *J.*
542 *Cell. Sci.* **119:** 4793-4795.

543 Yuan Q, Metterville D, Briscoe AD, Reppert SM. 2007. Insect Cryptochromes:
544 Gene Duplication and Loss Define Diverse Ways to Construct Insect Circadian
545 Clocks. *Mol. Biol. Evol.* msm011.

546 Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. 2014. A circadian
547 gene expression atlas in mammals: implications for biology and medicine. *Proc.*
548 *Natl. Acad. Sci. U. S. A.* **111:** 16219-16224.

549 Zhu H, Sauman I, Yuan Q, Casselman A, Emery-Le M, Emery P, Reppert SM. 2008.
550 Cryptochromes define a novel circadian clock mechanism in monarch butterflies
551 that may underlie sun compass navigation. *PLoS Biol.* **6:** e4.

552
553

Figure legend

Figure 1. Free-run behavioural rhythm in *Nasonia*. Representative actograms of individual *Nasonia* males in DD (left) and LL (right). Activity counts were sorted into 30 minute bins and plotted in blue. Yellow and grey backgrounds indicate lights on and lights off respectively. Gray and black bars below the actogram indicate the 12 hr subjective day and night.

Figure 2. Circadian transcriptional rhythms. (**A**) Heatmap of median-normalised expression of rhythmic ($q < 0.1$) transcripts in both constant darkness and constant light. (**B**) Histograms and heatmap of phases of rhythmic transcripts ($q < 0.1$ in both conditions), showing bimodal phase distribution and overlap between the two conditions.

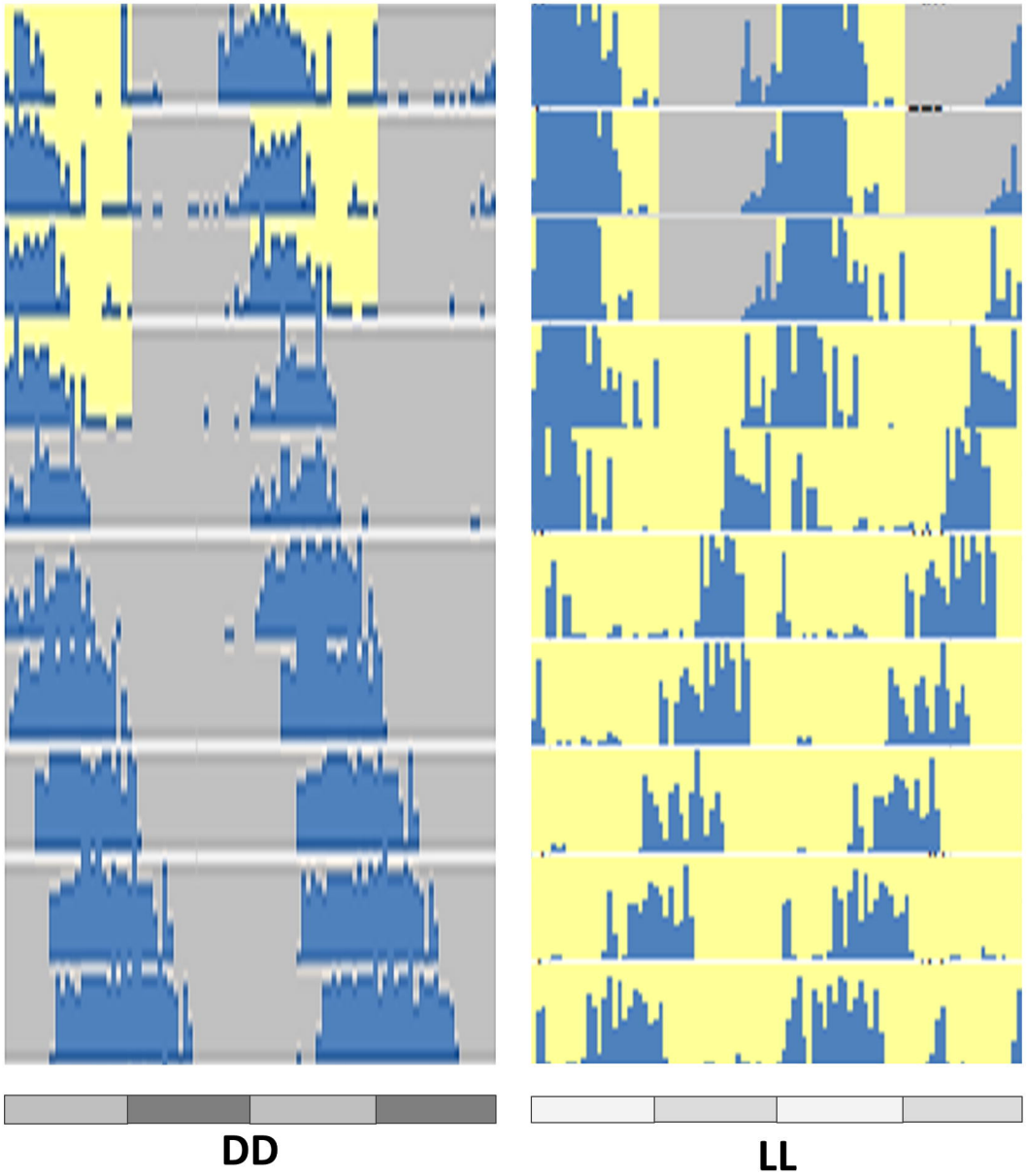
Figure 3. Enrichment of GO terms among cycling transcripts. (**A**) Bar plot of 10 top overrepresented GO terms (by gene proportion) for both DD and LL rhythmic genes. (**B**) Euler diagram showing the overlap of overrepresented terms in DD (blue) and LL (red).

Figure 4. Normalised expression of clusters with significant ($q < 0.01$) overrepresentations of rhythmic genes. Each transcript profile in each cluster is coloured by that gene's membership of the cluster.

Figure 5. Comparison of the DD and LL transcriptomes. (**A**) FPKM (log2) expression of transcripts in DD (x axis) and LL (y axis), showing genes classified

(> 1.5 median fold change) as differentially expressed up in DD (blue) and up in LL (red). (**B**) Selected overrepresented ($q < 0.01$) GO terms for genes more highly expressed in DD. (**C**) Heatmap showing median-normalised expression for differentially expressed transcripts, in DD (left) and in LL (right), sorted by fold change.

Figure 1



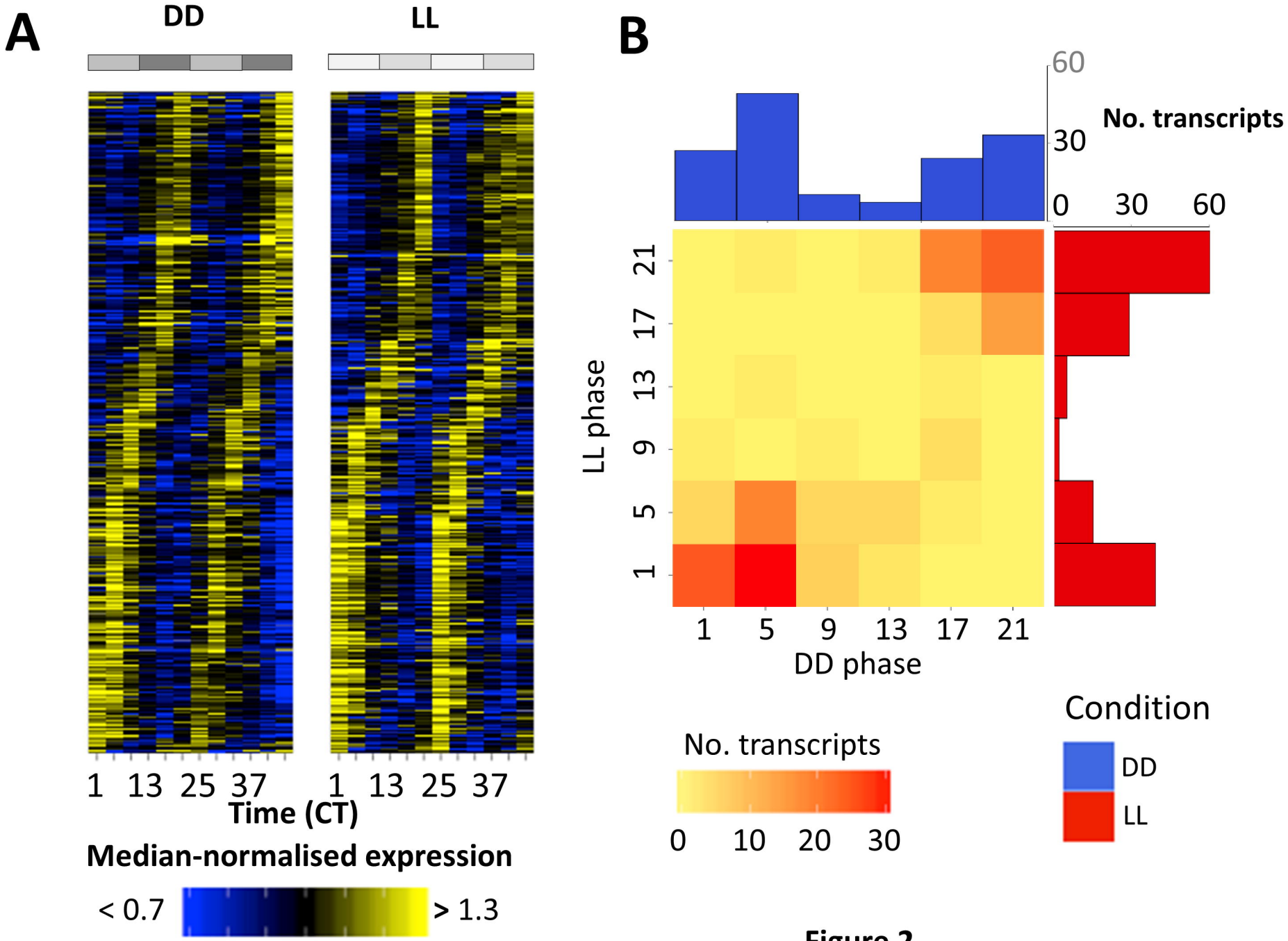

A
DD
LL
1 13 25 37
1 13 25 37
Time (CT)
Median-normalised expression
< 0.7
> 1.3
B
60
30
0
No. transcripts
0 30 60
LL phase
21
17
13
9
5
1
1 5 9 13 17 21
DD phase
No. transcripts
0 10 20 30
Condition
DD
LL

Figure 2

Figure 3

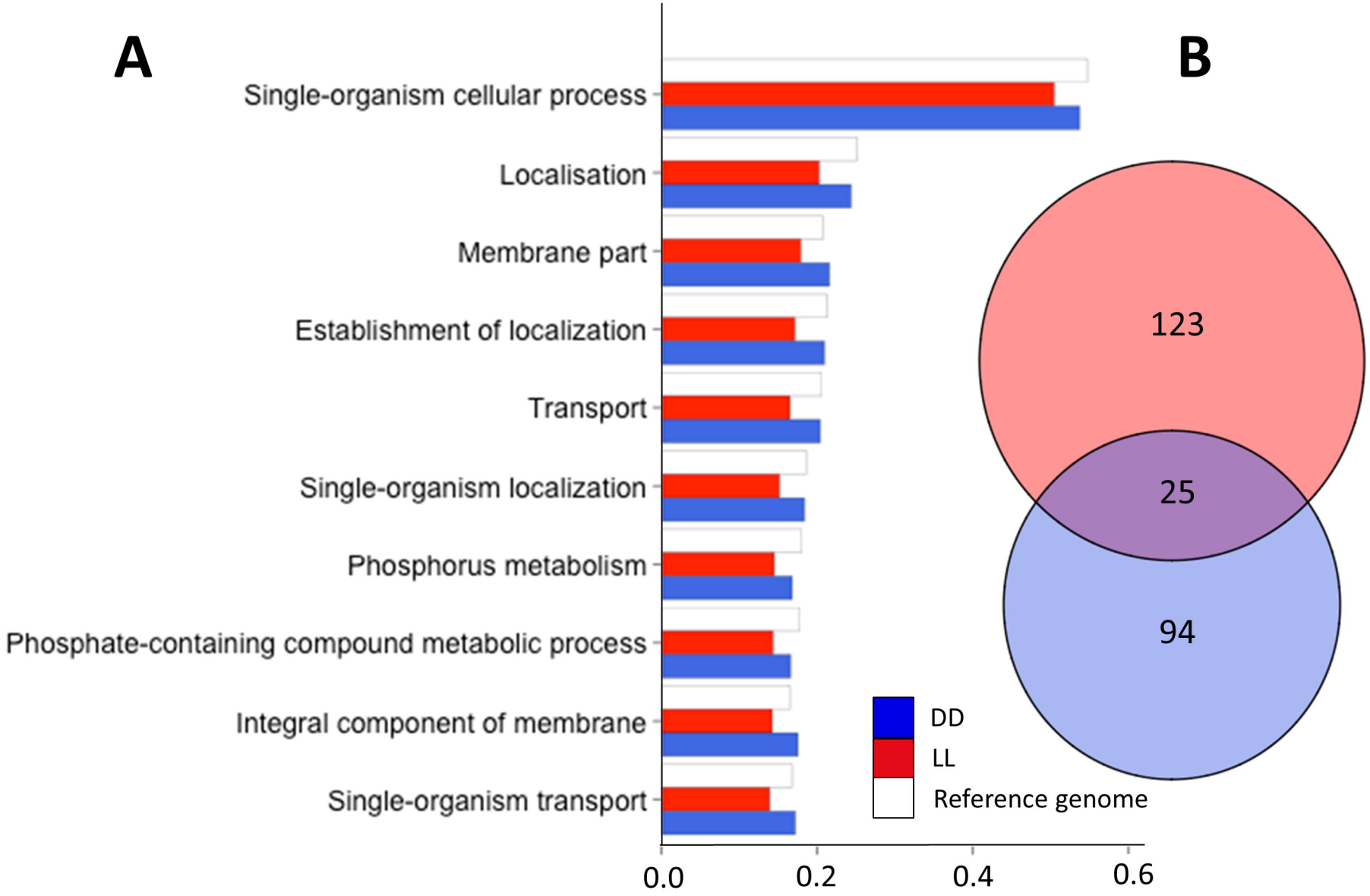
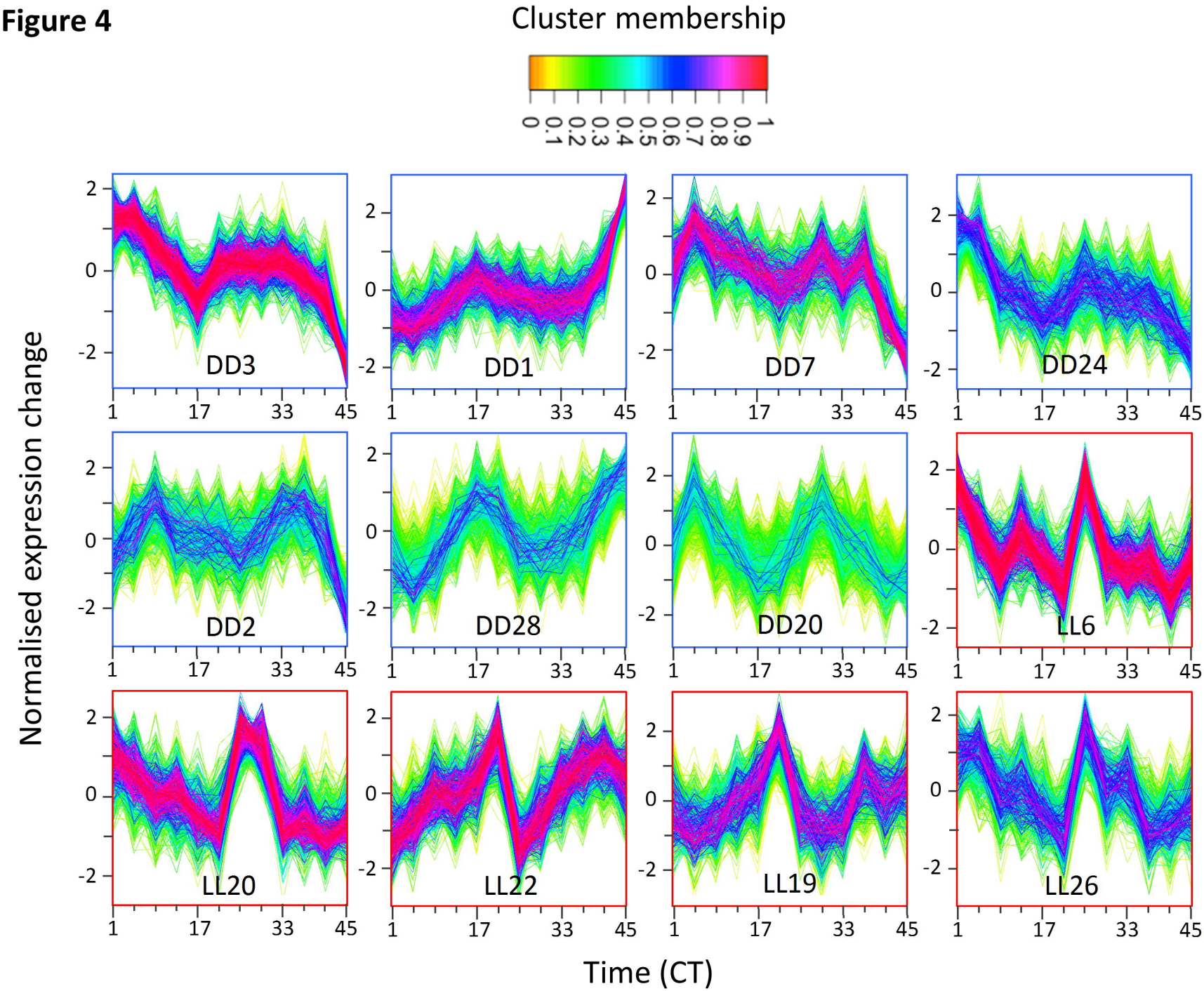

A
B
Single-organism cellular process
Localisation
Membrane part
Establishment of localization
Transport
Single-organism localization
Phosphorus metabolism
Phosphate-containing compound metabolic process
Integral component of membrane
Single-organism transport
0.0
0.2
0.4
0.6
DD
LL
Reference genome
123
25
94

Figure 4


Cluster membership
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Normalised expression change
Time (CT)
DD3
DD1
DD7
DD24
DD2
DD28
DD20
LL6
LL20
LL22
LL19
LL26
2
0
-2
1
17
33
45

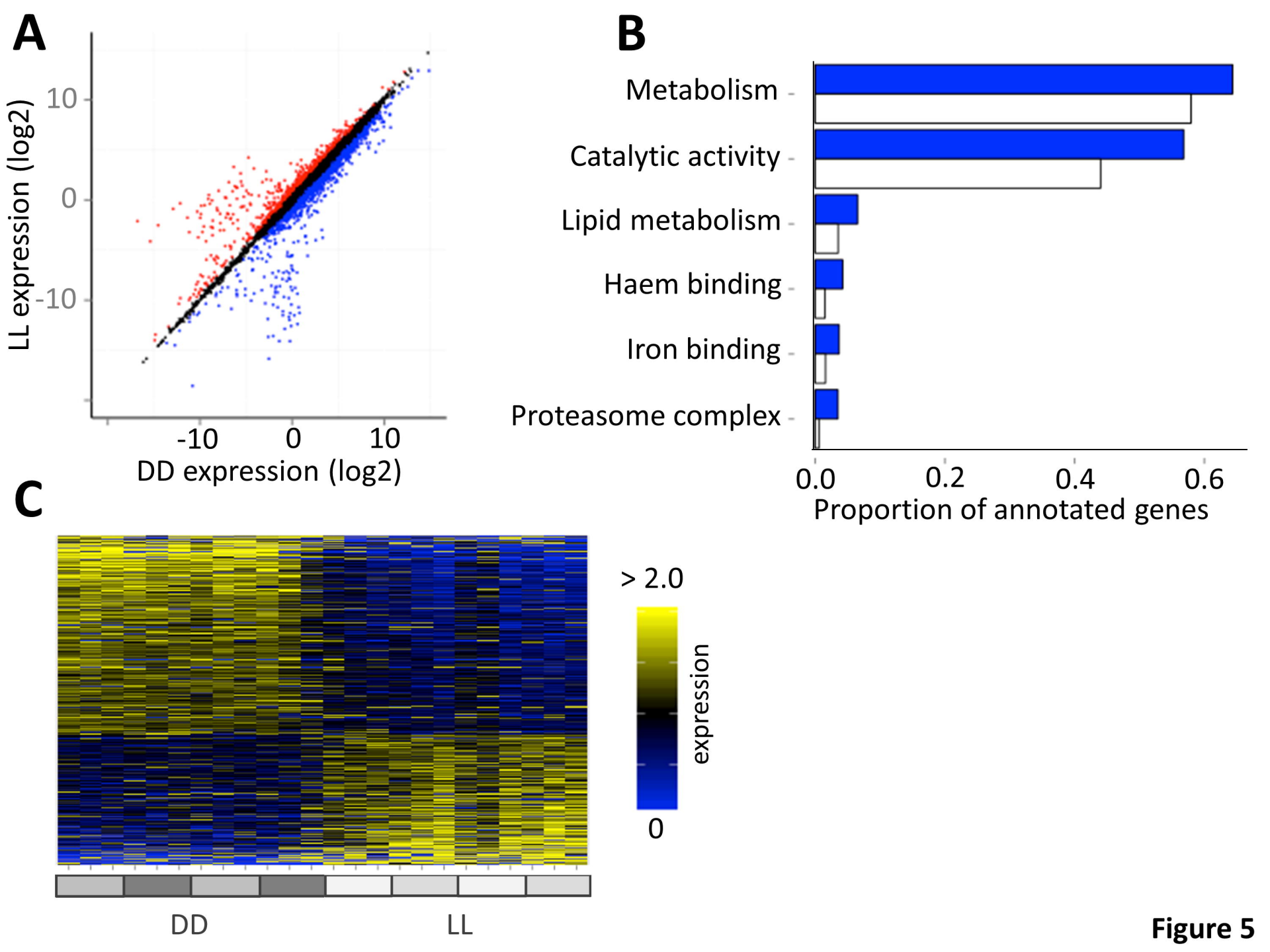

A
10
0
-10
LL expression (log2)
-10
0
10
DD expression (log2)
B
Metabolism
Catalytic activity
Lipid metabolism
Haem binding
Iron binding
Proteasome complex
0.0
0.2
0.4
0.6
Proportion of annotated genes
C
> 2.0
expression
0
DD
LL

Figure 5