# Lineage-specific, fast-evolving GATA-like gene regulates zygotic gene activation to promote endoderm specification and pattern formation in the Theridiidae spider

Sawa Iwasaki-Yokozawa[1], Ryota Nanjo[1,2], Yasuko Akiyama-Oda[1,3,4], Hiroki Oda[1,2,*]

[1]Laboratory of Evolutionary Cell and Developmental Biology, JT Biohistory Research Hall, Takatsuki, Osaka 569-1125, Japan

[2]Department of Biological Sciences, Graduate School of Science, Osaka University

[3]PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan

[4]Department of Microbiology and Infection Control, Faculty of Medicine, Osaka Medical and Pharmaceutical University

*Correspondence to: Hiroki Oda (hoda@brh.co.jp)

## Abstract

**Background:** The process of early development varies across the species-rich phylum Arthropoda. Owing to the limited research strategies for dissecting lineage-specific processes of development in arthropods, little is known about the variations in early arthropod development at molecular resolution. The Theridiidae spider, *Parasteatoda tepidariorum,* has its genome sequenced and could potentially to contribute to dissecting early embryonic processes.

**Results:** We present genome-wide identification of candidate genes that exhibit locally restricted expression in germ-disc forming stage embryos of *P. tepidariorum*, based on comparative transcriptomes of isolated cells from different regions of the embryo. A subsequent pilot screen by parental RNA interference identifies three genes required for body axis formation. One of them is a GATA-like gene that has been fast evolving after duplication and divergence from a canonical GATA family gene. This gene is designated *fuchi nashi* (*fuchi*) after its knockdown phenotypes, where the cell movement toward the formation of a germ disc was reversed. *fuchi* expression occurs in cells outside a forming germ disc and persists in the endoderm. Transcriptome and chromatin accessibility analyses of *fuchi* pRNAi embryos suggest that early *fuchi* activity regulates chromatin state and zygotic gene activation to promote endoderm

specification and pattern formation. We also show that there are many uncharacterized genes regulated by *fuchi*

**Conclusions:** Our genome-based research using an arthropod phylogentically distant from *Drosophila* identifies a lineage-specific, fast-evolving gene with key developmental roles in one of the earliest, genome-wide regulatory events, and allows for molecular exploration of the developmental variations in early arthropod embryos.

## Background

The early processes of development and reproductive strategies of animals vary among species. In certain metazoan phyla, including Chordata and Arthropoda, a high degree of early developmental variations among species is observed, despite the identification of similar morphological traits and gene expression patterns are recognized during mid embryogenesis in each phylum [1,2,3,4,5,6,7]. Mechanisms of animal evolution that account for varied developmental trajectories preceding the phylotypic period are poorly understood.

The species-rich phylum Arthropoda, comprising four major groups (Chelicerata, Myriapoda, Crustacea, and Hexapoda), conspicuously shows the diversification of early developmental processes without the disruption of stable embryonic traits, such as the elongating germ band and repetitive body units. Genetic mutation screens and a range of molecular and genetic techniques are feasible for the model organism *Drosophila melanogaster*, which has contributed to the identification of numerous essential genes and their interactions in body axis formation, germ layer formation, and segmentation in dipteran insects [8]. Establishing the *Drosophila* paradigm of genetic programs for early embryogenesis has allowed for candidate-gene approaches toward elucidating the evolution of developmental mechanisms in other insects or arthropod species. Such candidate-gene approaches, however, are biased with repsect to the detection of key developmental genes; having relatively low consideration for lineage-specific, fast evolving, or orphan genes [9,10]. Conversely, comparative genomics in Arthropoda and other metazoan phyla has revealed lineage-specific genomic events, including gain/loss, expansion, or divergence of a gene or a gene family [11,12,13,14]. *bicoid, spätzle*, and *gurken* in *Drosophila*, which are key regulators in the organization of the major embryonic axes, are examples of lineage-

64 specific or fast-evolving genes [15,16,17]. Considering these *Drosophila* cases and
65 similar ones in different phyla [18,19,20], an unbiased way of identifying key
66 developmental genes is required. Therefore, several studies that used non-*Drosophila*
67 insect species have performed genetic mutation or gene knockdown screens by
68 assessing larval cuticle phenotypes [21,22,23,24]. However, the identification and
69 characterization of novel genes with key functions in specific processes of early
70 embryogenesis are hardly feasible in any non-*Drosophila* arthropods, particularly in
71 non-insect arthropods.

72 In arthropods, many model species have continued to emerge, with the genome
73 sequenced [11,25]. There have been numerous technical merits with respect to the
74 study of early embryogenesis in one of them, the common house spider *Parasteatoda*
75 *tepidariorum*, and distinct differences from *Drosophila* have been observed in the early
76 embryogenesis mechanism [26]. These differences include early completion of
77 cellularization [27,28] and involvement of Hedgehog signaling in establishing the
78 polarity of the first embryonic axis and regulating symmetry-breaking movement of the
79 Dpp signaling center during the orthogonalization of the first and second embryonic
80 axes [29,30,31]. In the *P. tepidariorum* experimental system, genetic regulations during
81 early embryonic processes can be investigated using a simple gene knockdown
82 technique, that is, parental RNA interference (pRNAi) [30]. Research using *P.*
83 *tepidariorum* has been empowered by genomic and bioinformatic resources
84 [26,32,33,34,35,36,37], thus providing the foundation for identifying key genes
85 involved in *P. tepidariorum* development.

86 The egg of *P. tepidariorum* is spherical in shape (Fig. 1). As the development
87 starts, the egg nuclei synchronously divide, approaching the surface of the egg (stage
88 1). Cellular organization of the embryo is established by the time the nuclei increase to
89 16 [27]. The initial blastoderm comprises approximately 64 cells (stage 2; 11 h AEL),
90 showing spherical symmetry at the morphological level. This symmetry is broken at
91 approximately 15 h AEL by the emergence of uneven cell densities toward the
92 formation of a germ disc within a hemisphere of the egg (stage 3) (Fig. 1; Additional
93 file 1: Movie S1), which side is designated the embryonic side. At the pole on this side,
94 a small number of cells is apically constricted, thus forming a blastopore [26,29].
95 Zygotic gene activities are required for the process of germ-disc formation [38].

96 Although most blastoderm cells participate in the formation of the germ disc, cells
97 derived from around the abembryonic pole do not, and remain in the non-germ-disc
98 region (Fig. 1; see also Additional file 1: Movie S1). The formed germ disc is sharply
99 demarcated by the difference in cell density and comprises a single layer of more than
100 1,000 cuboidal epithelial cells [39]. Cells at the center of the germ disc, corresponding
101 to the blastopore, internalize to become the central endoderm (cEND) cells and the
102 cumulus mesenchymal (CM) cells [29,33]. The latter cell populations, which are
103 clustered, act as a dynamic source of Dpp signaling to promote the development of the
104 dorsal side of the embryo [29,30] (Fig. 1). Cells at and near the rim of the germ disc
105 internalize to become the peripheral endoderm (pEND) and mesoderm (pMES) cells
106 [33]. A line of cells along the rim of the germ disc express specific genes, including a
107 *P. tepidariorum hedgehog* homolog (*Pt-hh*), thereby contributing to the anterior-
108 posterior patterning of the embryonic field [31,40,41]. The germ disc is transformed
109 into a bilaterally symmetric germ band through orchestrated cell rearrangements, while
110 being progressively patterned [30,39] (Fig. 1).

111 The formation of the sharply demarcated germ disc is an early morphogenetic
112 event in the embryogenesis of Theridiidae and some other Araneoidea spiders
113 [28,29,38,42,43,44,45], but it is not a general feature of spider (Araneae) embryonic
114 development [46,47,48]. There are many other variations in early developmental
115 processes among spider species, including the presence/absence of a visible cumulus
116 and various modes of gastrulation [45,49]. The spider lineage provides diverse
117 developmental trajectories prior to reaching the embryonic traits typical of arthropods.

118 To explore the variation in the process of early embryonic development across
119 the phylum Arthropoda at molecular resolution, we conducted comparative
120 transcriptomes of cells isolated from different regions of the embryo at the germ-disc
121 forming stage in *P. tepidariorum*. Through a pRNAi-based functional screen of
122 differentially expressed gene candidates, we identified a lineage-specific GATA-like
123 gene, *fuchi nashi* (*fuchi*), whose knockdown hindered the completion of germ-disc
124 formation. We obtained genome-wide datasets that represented gene expressions and
125 chromatin accessibilities affected by *fuch*i knockdown in early embryos. Our findings
126 showed that *fuchi* regulates zygotic gene activation to promote endoderm specification
127 and pattern formation in the early stage of the Theridiidae spider embryo. This study

128 provides new molecular clues for exploring the mechanisms for diversifying early
129 developmental trajectories in the phylum Arthropoda.
130
131 **Results**
132 **Comparative transcriptome analyses of cells isolated from different regions of**
133 **early *P. tepidariorum* embryos**
134 To search for genes with locally-restricted expression in *P. tepidariorum* embryos at
135 stage 3, we performed RNA sequencing (RNA-seq) of small cell populations
136 (approximately 10-30 cells) isolated from central (c), intermediate (i), and peripheral
137 (p) regions of the nascent germ disc using glass capillary needles (Fig. 2A, B). RNA-
138 seq reads were mapped to the *P. tepidariorum* reference genome Ptep1.0
139 (GCA_000365465.1), and then the reads were counted against the AUGUSTUS gene
140 models (aug3.1) [32]. The read counts were compared using edgeR [50] by setting the
141 following three combinations: comparison I, c cells versus i/p cells; comparison II, p
142 cells versus c/i cells; and comparison III, i cells versus c/p cells (Fig. 2C-E). Based on
143 these comparisons, we genome-widely identified candidates of differentially expressed
144 genes (DEGs) (Additional file 2: Tables S1-S3). Using the values of false discovery
145 rate (FDR) and fold change (FC), the candidate DEGs were prioritized (lower FDR
146 with $\log_2 FC < -10$ [for comparisons I and II] or $\log_2 FC > 10$ [for comparison III]),
147 selecting the top 10 genes from comparison I as Group C (Fig. 2C), the top 5 genes
148 from comparison II as Group P (Fig. 2D), and the top 5 genes from comparison III as
149 Group CP (Fig. 2E). The top-ranked *g4238* in Group C was listed in Group CP, despite
150 its much stronger expression in c cells than in other cells. To avoid redundancy, *g4238*
151 was removed from Group CP. Therefore, a total of 19 genes were systematically
152 selected as high-priority candidate genes that might exhibit locally-restricted
153 expression in the stage-3 embryo.
154 To validate our gene-selection strategy, we examined the expression patterns
155 of the 19 selected genes in stage-3 embryos using chromogenic whole-mount in situ
156 hybridization (WISH) (Additional file 3: Fig. S1). Three of the ten Group C genes
157 (*g4238*, *g14287*, and *g16467*) showed specific expression at the embryonic pole.
158 Furthermore, three of the other Group C genes (*g15167*, *g5118*, *g20850*) and one of the
159 four Group CP genes (*g26874*) specific expression at the embryonic pole and in a broad

160 area on the abembryonic side. No other Group CP and Group P genes, however, showed
161 specific detectable signals.
162 Additionally, similar comparative transcriptome analyses of isolated cells were
163 applied to stage-4 and early stage-5 embryos. The resulting DEG lists (Additional file
164 2: Tables S4, S5) included some of the DEGs identified by the analyses of the stage-3
165 samples (e.g., *g4238*, and *g132*), as well as genes that had been known to show region-
166 specific expression at the corresponding and/or later stages (eg., *Pt-lab1* [*g7954*], *Pt-*
167 *BarH1* [*g8250*], *Pt-prd2* [*g18397*], and *Pt-hh* [*g4322*]).
168 These data suggest the effectiveness of our gene-selection strategy based on
169 comparative transcriptome analyses of isolated cells from early *P. tepidariorum*
170 embryos.
171
172 **Pilot functional screen identifies three genes required for germ-disc formation**
173 **and/or axis formation**
174 To identify genes with key functions in the early embryonic development of *P.*
175 *tepidariorum*, we screened the 19 candidate DEGs (from the stage-3 samples) using
176 parental RNAi (pRNAi). dsRNA synthesized for each gene was injected into at least
177 two adult females. Aliquots of eggs deposited by the animals were placed in oil to
178 monitor the developmental process from early stages under the stereomicroscope.
179 Through this screening, we identified three genes whose knockdown resulted in
180 uncommon defects associated with germ disc formation (*g26874* and *g7720*) and/or
181 cumulus movement (*g26874*, *g7720*, and *g4238*) (Additional file 4: Table S6)[51]. To
182 verify the RNAi specificity, we injected two or three dsRNAs prepared from non-
183 overlapping regions of each positive gene and confirmed that the same phenotypes were
184 produced depending on the genes (Fig. 3A; Additional file 3: Fig. S2; Additional file
185 5: Movie S2)[51]. In addition, we confirmed through WISH that the transcript levels
186 of the target genes were reduced in the corresponding pRNAi embryos (Fig. 3B).
187 Time-lapse microscopy and cell tracking revealed that *g26874* pRNAi
188 embryos showed normal blastoderm formation (stage 2) and subsequent initiation of
189 cell density shift toward the formation of a germ disc, in a rather normal way, but they
190 failed to stabilize the boundary of the forming germ disc (25 h AEL), with the cell
191 density shift being reversed (Fig. 3A; Additional file 5: Movie S2; Additional file 6:

192 Movie S3). In all of such severely affected *g26874* pRNAi embryos, cell thickening
193 took place at the embryonic pole. In 64% of the *g26874* pRNAi embryos (*n* = 101 from
194 10 egg sacs), however, CM cell migration was faint or failed to occur. The defects in
195 germ-disc formation were not recovered at later stages, showing no clear sign of cell
196 movements toward the formation of a germ band. In the most severe cases of
197 *g26874* pRNAi embryos, cells that had initially shifted toward forming a germ disc
198 reversed their movement to spread across the embryo, followed by abrupt breakdown
199 of the surface cell layer and extrusion of the yolk materials (Additional file 5: Movie
200 S2)[51].
201 *g7720* pRNAi embryos showed normal blastoderm formation followed by cell-
202 density shift (stages 2 and 3), but its development was arrested around the end of stage
203 3, showing a gradually degenerating germ disc (Additional file 5: Movie S2)[51].
204 Although cumulus-like cell thickening was formed at the embryonic pole, it persisted
205 there, with the overall morphology failing to develop further.
206 *g4238* pRNAi embryos were morphologically indistinguishable from wild-
207 type and *gfp* pRNAi (control) embryos until early stage 5. The earliest visible
208 difference of *g4238* pRNAi embryos from normal embryos was their failure to initiate
209 migration of the CM cells at early stage 5, followed by their disassembly and dispersal
210 (Additional file 5: Movie S2)[51]. In many of the embryos, surface cells around the
211 closed blastopore formed a tail-like protrusion. Despite experiencing cumulus defects,
212 some *g4238* pRNAi embryos displayed asymmetric cell movements toward the
213 formation of a germ band.
214
215 **Molecular characterization of the three identified genes**
216 Developmental transcript profiling in whole wild-type embryos, based on public
217 RNA-seq datasets [37], revealed that there is little or no maternal supply of *g26874*
218 and *g4238* transcripts and increasing levels of the zygotic transcripts prior to peaking
219 at stage 4 (Fig. 4D). In contrast, substantial levels of *g7720* transcript appeared to be
220 maternally supplied, with the zygotic transcript expressed at rather uniform levels
221 through development.
222 Reciprocal blastp/tblastn searches using *Drosophila* and mouse reference
223 sequence (RefSeq) protein databases at National Center for Biotechnology

224 Informataion (NCBI) and the *P. tepidariorum* aug3.1 transcript sequences [32]
225 revealed that *g7720* encoded an orthologue of *Drosophila* proximal sequence element
226 A (PSEA)-binding protein 49kD (Pbp49), and mouse small nuclear RNA-activating
227 complex subunit 3 (Snapc3)[52], and *g4238* an orthologue of *Drosophila* Ets98B (D-
228 ets4), and mouse SAM pointed domain-containing Ets transcription factor (Fig. 3C)
229 [53]. *g4238* was identical to *Pt-Ets4*, as reported in a previous study [54], and
230 consistent with our observations.
231 In contrast to the *g4238 and g7720* cases, reciprocal blastp/tblastn searches
232 failed to confirm the presence of *Drosophila* and mouse orthologues for *g26874*. The
233 best-hit *Drosophila* and mouse proteins were Pannier (Pnr), a GATA transcription
234 factor, and GATA-4, respectively (Fig. 3C), but these proteins were significantly
235 closer to GATA family members encoded by *g8336* and several other genes in the *P.*
236 *tepidariorum* genome. Based on these results, along with those obtained from
237 comprehensive sequence analyses described below, we concluded that *g26874*
238 encodes a novel GATA-like gene, which we named "*fuchi nashi*" (abbreviated *fuchi*)
239 after its knockdown phenotypes. This Japanese word means rimless.
240
241 ***fuchi* is a lineage-specific, fast-evolving GATA-like gene**
242 The family of GATA transcription factors are characterized by the possession of a
243 specific DNA-binding domain, which comprises two GATA-type zinc finger (ZF)
244 motifs (ZF1 and ZF2), typically $CX_2CX_{17}CX_2C$, and two basic regions followed by
245 each ZF motif (BR1 and BR2) (Fig. 4A, B)[55][56]. Fuchi had this GATA-type DNA-
246 binding domain, although its sequence was highly divergent from those of *Drosophila*
247 Pnr and mouse GATA-4. To comprehensively understand the diversity of GATA family
248 proteins among metazoans and the phylogenetic origin of *fuchi*, we exhaustively
249 collected and manually aligned amino acid sequences of GATA family proteins from *P.*
250 *tepidariorum*, other spider and non-spider chelicerate species as well as from other
251 selected arthropod and non-arthropod metazoan species using publicly available
252 bioinformatic resources (Additional file 7: Tables S7, S8). The resulting sequence
253 alignment revealed that the metazoan GATA family members were classified under
254 canonical and noncanonical types. The canonical type was characterized by a BR1 of
255 29 amino acid residues with high sequence similarity, and the noncanonical type by a

256 BR1 of more, or less, than 29 amino acid residues with a varying degree of sequence
257 divergence. The canonical GATA family members included proteins encoded by five *P.*
258 *tepidariorum* genes (designated Pt-GATA1 to Pt-GATA5, encoded by *g20514*, *g1834*,
259 *g8336*, *g25261*, and *g8337*, respectively) and their counterparts from other Araneae
260 species, as well as *Drosophila* Serpent (Srp), the only *Nematostella vectensis* GATA
261 (Nv-GATA), two echinoderm GATAs, and mouse GATA-1 to GATA-6 (Fig. 4B-D;
262 Additional file 7: Table S8). Phylogenetically widespread distribution of the specific
263 sequence features suggested that the canonical GATA family members represent the
264 ancestral state for metazoan GATA proteins.

265 The sequence alignment of the noncanonical GATA family members, including
266 Fuchi, guided the identification of at least three groups specifically found within the
267 Araneae lineage, based on gap insertion patterns, cysteine residue spacing, and
268 sequence signatures (Fig. 4C, D; Additional file 7: Table S8). These three groups were
269 designated as GATA-like 1 to 3 (GATAL1, GATAL2, and GATAL3). GATAL1, which
270 included a *P. tepidariorum* protein encoded by *g26871* (Pt-GATAL1), was relatively
271 close to the canonical GATA family members at the sequence level, but they had a BR1
272 of 34 amino acid residues. GATAL2, which included Fuchi and another *P. tepidariorum*
273 protein encoded by *g26875* (designated Pt-GATAL2b), was specifically identified by a
274 signature sequence QCV(K/R)CG at the N-terminal end of ZF1, varied cysteine residue
275 spacing in ZF1 ($CX_2CX_{17\text{-}20}CX_2C$), and a highly varied number (23-44) of amino acid
276 residues in BR1. GATAL3, which was only found in spiders of RTA clade, was
277 specifically identified by conserved unique sequences and cysteine residue spacing
278 ($CX_2CX_{18}CX_2C$) in ZF1. The signature sequence unique to GATAL2, QCV(K/R)CG,
279 corresponded to ECVNCG, conserved in most canonical GATA family members
280 (including Pt-GATA1 to 4) and GATAL1, to QC(A/V)(M/V)CG, conserved in Pt-
281 GATA5 and its counterparts, and to ECANCG, conserved in GATAL3.

282 To investigate the phylogenetic relationship between canonical and
283 noncanonical GATA family proteins, we performed maximum likelihood (ML) analysis
284 using unambiguously aligned 75 amino acid sites from a total of 118 proteins with both
285 ZF1 and ZF2 representing the typical cysteine residue spacing ($CX_2CX_{17}CX_2C$) (Fig.
286 4E, F; Additional file 7: Table S8). Results showed Araneae-specific expansions of
287 GATA protein family members, through gene duplication and divergence. ML tree

288 topology supported the orthology of each of the GATA5, GATAL1, and GATAL2
289 groups. Furthermore, branch lengths indicated that the sequences of the GATAL2
290 proteins had been evolving much faster than those of the canonical GATA family and
291 GATAL1 proteins (Fig 4F). Owing to their atypical cysteine residue spacing in ZF1
292 ($CX_2CX_{18\text{-}20}CX_2C$), Fuchi and other Araneoidea GATAL2 proteins could not be
293 included in the ML analysis (Fig. 4C; Additional file 7: Table S8). Importantly,
294 GATAL2 proteins with the typical ZF1 ($CX_2CX_{17}CX_2C$), which were used in the ML
295 analysis, were present in a broader phylogenetic range (including Leptonetidae,
296 Eresidae, and RTA clade) (Fig. 4C, E, F; Supplementary Table 4). This observation
297 suggested that variously divergent sequences in ZF1 of the Araneoidea GATAL2
298 proteins, except the specific signature sequence, were likely due to rapid sequence
299 evolution from the GATAL2 ancestral state.
300 Unlike the other two of the three genes judged as positive in the functional
301 screen, *fuchi* was characterized as a lineage-specific gene. We therefore decided to
302 focus our later analyses on this gene.
303
304 ***fuchi* is expressed in endoderm cells originating from both sides of the early**
305 **embryo**
306 Among the eight GATA family genes in *P. tepidariorum*, *fuchi* was the only gene
307 which was transcribed at substantial levels during the germ-disc forming stage (stage
308 3) (Fig. 3C; Additional file 3: Fig. S3). We investigated the expression of *fuchi*
309 transcript in early embryos at cellular resolution, using the combination of FISH with
310 antibody staining for β-catenin, which serves as a marker for regions of cell-cell contact,
311 in addition to conventional WISH (Fig. 5). Owing to technical limitations in the
312 fixation of stage-1 embryos, the earliest stage of development examined was stage 2,
313 the blastoderm stage (13 h AEL), when there is no morphological sign of asymmetry
314 reflecting the future embryonic axes. At this stage, localized signals for *fuchi*
315 expression were detected on one side of the embryo (Fig. 5A, A', B, B') but could not
316 be related with the axis that would emerge a little later (15 h AEL).
317 At early stage 3 (15 h AEL), the embryo begins to display morphological
318 asymmetries, with an emerging embryonic-abembryonic (Em-Ab) axis (Fig. 1). At this
319 stage and later, the *fuchi* transcript was detected in a small number of cells condensed

320 at the embryonic pole (Fig. 5I, I', I") and in a larger number of cells showing a
321 progressively wider distribution on the abembryonic side of the embryo (Fig. 5J, J', J").
322 Using a flat-preparation of a representative early stage 3 (15 h AEL) embryo stained
323 for *fuchi* transcript, the number of *fuchi*-expressing cells was counted. There were 8
324 and 20 *fuchi*-expressing cells around the embryonic pole and on the abembryonic side,
325 respectively, among a total of approximately 300 cells (Additional file 3: Fig. S4).
326 Notably, signals for *fuchi* transcript in the nuclei were, in many cases, observed as
327 paired dots (Fig. 5I', J'), indicating that the zygotic transcription was already active at
328 early stage 3. In the equatorial area of stage 3 embryos (15 and 21 h AEL), there was a
329 transition from cells expressing *fuchi* transcript, on the abembryonic side, to those not
330 expressing it, on the embryonic side (Fig. 5J, J', J", K, K', K"). In the embryo at 21 h
331 AEL, the *fuchi*-expressing cells had larger apical surfaces than the *fuchi*-negative cells
332 (Fig. 5K').
333 In embryos at stage 4 (25 and 27 h AEL), most *fuchi*-expressing cells on the
334 abembryonic side had further enlarged surfaces with less tensed lines of cell-cell
335 contact, which contrasted with a tightly packed organization of *fuchi*-negative cells
336 constituting the nascent germ disc on the embryonic side (Fig. 5D, L, M, M', M"). A
337 few *fuchi*-expressing cells, however, were identified as part of the germ disc at its
338 border (Fig. 5M', M"). The germ disc forms as a monolayered epithelium consisting of
339 uniform cells, but its center, corresponding to the embryonic pole, acts as a site of
340 blastopore formation (Fig. 1), through which cEND cells and CM cells are internalized
341 [33]. *fuchi* expression was observed in all these internalizing and internalized cells (Fig.
342 5L, L', L").
343 In embryos at stage 5 (32 and 38 h AEL), all cells internalized through the
344 blastopore continued to express *fuchi* (Fig. 5E, E', F, F', N, N1', N1", O, O1").
345 Peripheral populations of *fuchi*-expressing cells appeared to be internalized a little later
346 than the central populations. At early stage 5 (32 h AEL) *fuchi*-expressing cells were
347 found on the embryo surface not only in the non-germ disc region but also at the rim
348 of the germ disc (Fig. 5N, N2', N2"). At late stage 5 (38 h AEL), there were *fuchi*-
349 expressing cells that were internalized/internalizing from or through the rim of the germ
350 disc (Fig. 5F, F', O, O2', O2"). Using a DNA stain probe, SPY555-DNA, we tracked
351 some cell nuclei in the germ-disc/non-germ-disc transition area from early stage 5 to

352 stage 8 in a live embryo (Additional file 4: Movie S4), which directly visualized two
353 phases of cell internalization. Phase 1 of cell internalization occurred from mid to late
354 stage 5, where cells that were not part of the germ disc were approaching the rim of the
355 germ disc and then internalizing to below the surface cell layer of the germ disc. These
356 were most likely to be *fuchi*-expressing cells. Phase 2 of cell internalization occurred
357 subsequently, where cells at the rim of the germ disc were internalizing. Double
358 staining for *fuchi* and *Pt-hh* transcripts revealed certain cells expressing both *fuchi* and
359 *Pt-hh* at and near the rim of the germ disc at early stage 5 (Fig. 5P, P'). Upon
360 internalization, however, *fuchi*-expressing cells were exclusively *Pt-hh*-negative (Fig.
361 5Q, Q').
362 To determine whether *fuchi*-expressing cells internalized from the peripheral
363 side of the germ disc are endoderm or mesoderm, double staining of late stage 5 and
364 stage 6 embryos for *fuchi* and *Pt-fkh* (a marker for pMES plus pEND; [57]) or *003_J01*
365 (a marker for pMES)[58] was performed. The use of *003_J01* enabled us to identify
366 pMES cells prior to internalization, which was not the case with *Pt-twi*, another pMES
367 marker[33,57,59]. Observations indicated that all internalized, *fuchi*-expressing cells
368 were *Pt-fkh*-positive and most of them were *003_J01*-negative (Additional file 3: Fig.
369 S5). It was also indicated that *003_J01*-positive cells appeared to be internalized later
370 than *fuchi*-expressing cells, which presumably corresponded to Phase 2 of cell
371 internalization. *fuchi* expression in embryos at stages 6 and 7 never showed a spatially
372 periodic pattern along the emerging body axis (Fig. 5G, G', H, H'), contrasting with
373 mesoderm marker expression [59]. Taken together, these data suggested that *fuchi* is
374 expressed in endoderm cells originating from both sides of the early embryo. In
375 addition, *fuchi* expression occurred in association with extraembryonic differentiation
376 (Fig. 5G, G', H, H'). This *fuchi* expression was already initiated in the germ disc at late
377 stage 5 (Fig. 5O, O1").
378
379 ***fuchi* knockdown prevents the demarcation of a forming germ disc**
380 To investigate the defects following *fuchi* knockdown, we analyzed the expression of
381 *Pt-hh*, *Pt-fkh*, and *fuchi* transcripts in wild-type and *fuchi* pRNAi embryos at stage 4
382 and early stage 5 at cellular resolution. In normal germ-disc development, a line of cells
383 expressing *Pt-hh* and *Pt-fkh* was organized to demarcate the boundary of the forming

384 germ disc during stage 4 and early stage 5 (Fig. 6A, A', C, C', E, E', G, G'). In the
385 recognizable germ-disc/non-germ-disc transition area of *fuchi* pRNAi embryos at stage
386 4, the initial *Pt-hh* expression was suppressed, while *Pt-fkh* expression was initiated in
387 certain cells with *fuchi*-positive nuclei and their neighboring cells (Fig. 6B, B', D, D').
388 At early stage 5, faint *Pt-hh* and *Pt-fkh* expression were detected in small numbers of
389 cells, but they did not contribute to shaping the germ disc (Fig. 6F, F', H, H'). In *fuchi*
390 pRNAi embryos at a later stage (late stage 5), there was no clear separation between
391 cell populations on the embryonic and abembryonic sides, with cell internalizations
392 being highly limited (Fig. 6I, I1', I1", I2', I2"). The only cells situated below the surface
393 layer were cells at or near the embryonic pole (Fig. 6I1', I1"), which were associated
394 with β-catenin concentrations marking the CM cell cluster. These observations
395 suggested that *fuchi* knockdown prevented the demarcation of a forming germ disc and
396 interrupted the formation of the germ layers.
397
398 ***fuchi* is required for zygotic activation of endodermal and patterning genes**
399 To explore the role of *fuchi* in regulating early *P. tepidariorum* development, we
400 conducted comparative transcriptome analyses of *fuchi* pRNAi versus untreated
401 embryos at stages 2, 3, and early stage 5 using RNA-seq. These analyses identified 6,
402 139, and 785 DEGs for stages 2, 3 and early stage 5, respectively (FDR<0.01; Fig. 7A,
403 B; Additional file 9: Tables S9-S11). The experiments were validated by detecting
404 lowered levels of *fuchi* transcript and unaffected levels of the α-catenin (*catA*)
405 transcript in the *fuchi* pRNAi samples (Fig. 7C). Except for the *fuchi* gene itself, there
406 was no overlap between the stage-2 and stage-3 DEGs but there was an overlap of 93
407 genes between the stage-3 and early stage-5 DEGs (Fig. 7A-C; Additional file 3: Fig.
408 S6; Additional file 9: Tables S9-S11). Transcripts for the stage-2 DEGs appeared to be
409 maternally supplied, which contrasted with those for highly ranked genes from the
410 stage-3 and early stage-5 DEGs lists (sorted by FDR value). They tended to show a
411 sharp increase in the transcript level during stages 2 to 4, with no or little transcript
412 supply from the mother (Fig. 7C, D). All the top 20 genes from the stage-3 DEG list
413 were included in the stage-5 DEG list (Fig. 7D; Additional file 9: Table S10).
414 Transcripts from 8 of the top 10 DEGs (*g1125*, *g2114*, *g20850*, *g17950*, *g12453*,
415 *g27539*, *g486*, and *g26914*) were detected in all or part of the internalized endoderm at

416 late stage 5, and those from 3 of the 8 genes (*g17950*, *g12453*, *g20850*) were detected
417 in cells on the abembryonic side at stage 3 (Additional file 3: Fig. S7). Notably, the
418 *g20850* was identified as one of the Group C genes in the initial comparative
419 transcriptome analysis of isolated cells (Fig. 2C; Additional file 3: Fig. S1).
420 The stage-3 *fuchi* pRNAi DEGs list also included several genes previously
421 characterized as key regulators of embryonic patterning in *P. tepidariorum*, such as *Pt-*
422 *hh* (*g4322*) [31], *Pt-sog* (*g23966*) [30], and *Pt-Delta* (*g25248*)[33] (Additional file 3:
423 Fig. S8; Additional file 9: Table S10). Despite the effects of *fuchi* knockdown on the
424 expression of these patterning genes, certain patterning events involving *Pt-msx1* and
425 *Pt-Delta* expression were initiated around the blastopore in *fuchi* pRNAi embryos
426 (Additional file 3: Fig. S8). *fuchi* knockdown led to a lack of circular patterns of *Pt-hh*
427 and *Pt-otd* expression associated with anterior patterning, but its effects on the *Pt-hh*
428 and *Pt-otd* expression levels in whole embryos at early stage 5 were limited.
429 Taken together, these results suggested that *fuchi* is required for the zygotic
430 activation of a set of endodermal genes and of some patterning genes at stage 3 or
431 shortly before. In addition, *Pt-GATA4* (*g25261*), the likely counterpart of *Drosophila*
432 *srp* (Feitosa et al., 2017), and *Pt-GATA5* (*g8337*) were included in the stage-5 *fuchi*
433 pRNAi DEG list but not in the stage-3 one (Additional file 9: Tables S9, S10).
434
435 ***fuchi* is involved in regulating chromatin accessibility in specific genomic regions**
436 **in early embryos**
437 Certain GATA family members are known to act as pioneer transcription factors that
438 can influence chromatin structure [60]. Therefore, we considered the possibility that
439 *fuchi* controls the transcription of downstream genes through chromatin structure
440 regulation. To investigate this possibility, we conducted assay for transposase-
441 accessible chromatin using sequencing (ATAC-seq) of wild-type (untreated), *fuchi*
442 pRNAi, and *Pt-hh* pRNAi embryos at stage 3 (18h AEL), obtaining genome-wide
443 datasets on chromatin accessibility. Read counts in ATAC-seq peaks were subjected to
444 comparative analyses, and 317 genomic regions that were differentially accessible
445 between the *fuchi* pRNAi and untreated samples but only 3 between the *Pt-hh* pRNAi
446 and untreated samples were identified (Fig. 8A; Additional file 10; Tables S12, S13;
447 Additional file 11; Tables S14, S15; FDR<0.05). This indicated that the *Pt-hh* pRNAi

448 samples served as a negative control for assessing the impact of *fuchi* knockdown. The
449 two comparisons revealed only one shared genomic region (Fig. 8A; Additional file 10;
450 Table S13; Additional file 11; Table S15), which potentially reflected a nonspecific
451 effect stemming from pRNAi treatment. Therefore, 316 genomic regions where
452 chromatin accessibility was significantly affected following *fuchi* knockdown in stage-
453 3 embryos were identified (Fig. 8B).

454 Most of these regions (276/316, 93%) showed suppressed accessibility upon
455 *fuchi* knockdown (Fig. 8A, B), and 30 of them were found to be located closer to (or
456 within) genes included in the *fuchi* pRNAi DEG lists (for stage-3 and/or early stage-5)
457 than to any other annotated genes (Fig. 8B; Additional file 10: Table S13). Moreover,
458 motif discovery search revealed that motif sequences RNWGATAAGAVW, similar to
459 typical GATA-binding motif sequences, were significantly enriched among the
460 sequences of the *fuchi*-dependent chromatin accessibility peaks (Fig. 8C). *fuchi*-
461 dependent chromatin accessibility peaks containing these motif sequences were
462 actually present within or close to some of the genes identified by the comparative
463 transcriptome analyses of *fuchi* pRNAi versus untreated embryos (Fig. 8B).

464 Notable examples of such genes were the top-ranked gene from the stage-3
465 *fuchi* pRNAi DEGs list (*g1125*), the key patterning gene *Pt-hh* (*g4322*), and the
466 canonical GATA family member *Pt-GATA5* (*g8337*) (Fig. 9A, B, C). *g1125* was
467 predicted to encode an uncharacterized product containing two potential
468 transmembrane domains (XP_015915605.1; Fig. 9D). Genes encoding homologous
469 proteins to this product were detected in the genomes of at least three other Araneoidea
470 species, *Araneus ventricosus* (GBO06802.1), *Argiope bruennichi* (KAF8764632.1),
471 and *Oedothorax gibbosus* (KAG8179838.1)(Fig. 9D), but any non-Araneoidea protein
472 sequences that showed similarity to these products (E-value < 1) were not found in the
473 NCBI non-redundant protein sequences collection (March 7, 2022). *Pt-GATA5*
474 transcript, like *fuchi*, was observed in cEND, CM, and pEND cells at stage 5 and widely
475 distributed endoderm cells at stage 7 (Fig. 9E, F). Altogether, these results suggested
476 that *fuchi* is required for establishing "open" states of chromatin structure in specific
477 genomic regions, some of which are associated with endodermal and patterning genes
478 expressed in the early stages.
479

480 **Detection of consistent endodermal gene expression based on single-nucleus and**
481 **single-cell RNA-seq datasets**
482 In a recent separate study, we performed single-nucleus and single-cell RNA-seq
483 analysis of late stage 5 embryos in *P. tepidariorum* [61]. In this study, clearly isolated
484 endodermal cell clusters, which were marked by a known early endoderm marker
485 012_A08 [57], were identified (Additional file3: Fig. S9). Strong *fuchi* expression was
486 detected in these clusters, with weaker levels of *fuchi* expression being additionally
487 detected in mesodermal (marked by *Pt-twi*)[59] and dorsal germ-disc (marked by *Pt-*
488 *GATA1*)[61] cell populations. At the same time, the endoderm-specific expression of
489 *Pt-GATA4* and *Pt-GATA5* was confirmed. In contrast to *fuchi*, and the two canonical
490 GATA genes, *Pt-hh* expression was missing or largely reduced in the endodermal cell
491 populations at late stage 5. These data were consistent with the staining data shown in
492 Fig. 5 and Fig. S5 (Additional file 3). The single-nucleus and single-cell RNA-seq
493 datasets further confirmed that most of the high-ranking genes from stage-3 *fuchi*
494 pRNAi DEG list showed endoderm-specific expression at late stage 5 (Additional file
495 3: Fig. S10).
496
497 **Discussion**
498 We conducted genome-based research using the spider *P. tepidariorum* to explore the
499 variations in the mechanisms of early embryonic development in the phylum
500 Arthropoda. A key initial attempt in this study was to analyze transcriptomes of cells
501 isolated from different regions of the early spider embryo. This allowed for the
502 genome-wide identification of candidate genes whose expression might be locally
503 restricted along the emerging first embryonic axis, without relying on knowledge from
504 other organisms. Although similar approaches to identifying localized transcripts in
505 early animal embryos have been reported in non-arthropod animals [18,62,63], early
506 arthropod embryos are rarely studied in such a way.
507 A benefit of using *P. tepidariorum* is the applicability of a simple gene-
508 knockdown technique, that is, pRNAi. This allowed us to carry out a functional screen
509 of some of the listed genes. Gene-specific embryonic phenotypes were obtained using
510 3 of the 19 genes tested (15%). Compared with similar screening efforts we previously
511 made following sequence-based gene selection, this small proportion is not surprising

512 [31]. Although similar RNAi-based gene knockdown is available in certain arthropods,
513 such as *Tribolium castaneum*, it is important to note that the *P. tepidariorum* system
514 allows for the easy monitoring of early embryonic developmental processes, even
515 following pRNAi treatment. Time-lapse recording of the phenotypes from early stages
516 gave sharable, rich, objective information on gene functions and the timing of their
517 actions. Early defects often lead to reduced characteristic morphologies at later stages.
518 Exclusion of common spontaneous defects is required for effective functional
519 screening. Direct detection of the earliest developmental defects following knockdown
520 treatments in the initial screening could simplify the structuring of our interests.
521 Despite the small scale of the functional screen we conducted, incomparable with the
522 large-scale, unbiased RNAi screen in *Tribolium* [24], our results may encourage
523 additional gene-function screens in *P. tepidariorum*.

524 Unexpectedly, our pRNAi screen found that the cell movement toward the
525 formation of a germ disc was reversed following the knockdown of a GATA-like
526 transcription factor gene, which we named *fuchi nashi* (*fuchi*). This phenotype suggests
527 the presence of a shape-stabilizing phase in the germ-disc development, which likely
528 corresponds to stage 4, during which a line of cells along the border is organized to
529 gain specific gene expressions. *fuchi* expression, however, is initiated much earlier than
530 this stage, in cells outside a forming germ disc, and later continues in all internalizing
531 and internalized endoderm cells. Our comparative transcriptome and chromatin
532 accessibility analyses of *fuchi* pRNAi versus untreated embryos suggested that early
533 *fuchi* activities have broad impacts on zygotic gene activation around stage 3. Lack of
534 the *fuchi* activities appeared to hinder the specification of endoderm fates. The
535 abnormal reversed cell movements in *fuchi* pRNAi embryos could be, at least in part,
536 explained if we assume that the different cell types being specified exhibit different
537 mechanical properties. The earliest morphological defects of *fuchi* pRNAi embryos
538 possibly result from a failure to establish the differences in adhesion tension between
539 cell populations.

540 We showed that early *fuchi* activities influence the expression of some of the
541 key signaling genes involved in embryonic development and pattern formation. These
542 include *Pt-hh*, *Pt-Delta*, and *Pt-sog*. *Pt-hh* and *Pt-Delta*, down-regulated in response to
543 *fuchi* knockdown at stage 3, are known to show expression patterns similar to *fuchi* in

544 early embryos (stages 3 and 4)[31,33], and *Pt-sog*, up-regulated oppositely, to be
545 expressed in broad domains of the stage-5 germ disc [30]. Notably, early *Pt-hh*
546 expression initiated from the abembryonic side contributes to the formation of the
547 global polarity in the germ disc and subsequent anterior patterning. In *fuchi* pRNAi
548 embryos, however, no patterning events that should occur from the abembryonic side
549 of the egg were identified. These findings suggest that early *fuchi* activities specify
550 endoderm fates and promote pattern formation in a wide area of the embryo, at the
551 same time. Similar GATA factor-mediated regulation of *hh* expression has been
552 documented in several mammalian and *Drosophila* developmental events [64,65,66].
553 The family of GATA transcription factors is known to regulate cell fate
554 specification and cell differentiation during various developmental processes in a wide
555 range of animals [68]. Our functional and phylogenetic characterization of *fuchi*
556 provide the opportunity to reconsider the relationship between the evolution of GATA
557 family members and their developmental roles. We defined "canonical" members of
558 the GATA family, as they have perfectly alignable, close sequences to each other
559 throughout the double zinc-finger DNA-binding domain regardless of the phylogenetic
560 distance of the species. The canonical members include all six GATA paralogs in
561 vertebrates, GATA-1 to GATA-6, the only *Nematostella* GATA, *Drosophila* Serpent
562 (Srp), and five *P. tepidariorum* GATA paralogs, Pt-GATA1 to Pt-GATA5. The
563 vertebrate GATAs are phylogenetically classified into two subgroups: GATA-1/2/3
564 and GATA-4/5/6, while Srp is closer to GATA-4/5/6 [56][69]. *srp* functions for
565 endoderm specification in early *Drosophila* embryos [70], and vertebrate GATA-4/5/6
566 genes have developmental roles in endodermal cell lineages [71,72,73,74]. Notably, in
567 mammals, GATA-6, in cooperation with GATA-4, is essential in the specification of
568 the primitive and definitive endoderm, through chromatin accessibility regulation
569 [75,76]. Based on these lines of evidence, the idea that the roles of GATA-4/5/6 genes
570 in endoderm specification and development are conserved across bilaterians has been
571 widely accepted. Consistent with this idea, single-cell and single-nucleus
572 transcriptomes indicated that the likely *srp* ortholog *Pt-GATA4* (*Pt-srp*) appears to be
573 expressed in an endoderm-specific manner at least at late stage 5 of *P. tepidariorum*
574 development (Additional file 3: Fig. S9), although a previous report mentioned that
575 "*Pt-srp* expression starts at stage 8 in the presumptive extraembryonic cells." [77]

576 Evidently, developmental transcriptomes gave a quantitative indication that there is
577 little *Pt-GATA4* expression at stage 4 and earlier but substantial expression at late stage
578 5 and later (Additional file 3: Fig. S3). We suggested that instead of *Pt-GATA4*, *fuchi*
579 and *Pt-GATA5* are key GATA family genes contributing to endoderm specification and
580 development in the earlier stages of *P. tepidariorum* development. Although *Pt-GATA5*
581 is a canonical member of the GATA family, it appears to have no orthologous
582 counterpart outside the Araneae. Fuchi is not a canonical GATA factor, having a highly
583 diverged sequence with retaining the dual zinc fingers. *fuchi* appears to act on *Pt-*
584 *GATA5*, possibly through chromatin accessibility regulation. The *fuchi*-dependent
585 ATAC-seq peak detected in the *Pt-GATA5* locus is located in an intron between exons
586 encoding the two zinc fingers. Sequential regulation of multiple GATA factors in early
587 endoderm specification and the subsequent differentiation process is a shared feature
588 among distantly related bilaterians, including mouse, *C. elegans* and *Drosophila*
589 [76][78][79]. However, the genetic components responsible for this regulation exhibit
590 no orthologous relationships between the species, even within the Arthropoda,
591 suggesting independent modifications of the GATA-mediated regulatory system for
592 endoderm specification and development in the respective animal lineages.
593 *fuchi* and its orthologs (GATAL2) were found in sub lineages of the Araneae,
594 including many Araneomorphae taxa but not non-Araneomorphae. These GATA-like
595 genes, regardless of the cysteine spacing patterns in ZF1, have been rapidly evolving
596 their sequences compared to most other GATA and GATA-like genes. Despite this
597 divergence trend, we could define the likely orthologous GATAL2 group using the
598 conservative signature sequence QCV(K/R)CG. Because the amino acid residues
599 homologous to this sequence are involved in binding to specific cofactors, such as the
600 vertebrate Friend of GATA (FOG) and *Drosophila* U-shaped (Ush), in canonical
601 GATA factors [80,81,82], Fuchi possibly binds to a unique cofactor via this sequence.
602 In relation with GATA cofactors, one of two *P. tepidariorum* homologs of *Drosophila*
603 Ush (*g14866*; the other homolog is *g16893*) was included in both the gene lists obtained
604 from the comparative RNA-seq and ATAC-seq analyses using stage-3 *fuchi* pRNAi
605 embryos (Fig. 8; Additional file 9: Table S10; Additional file 10: Table S13). Certain
606 GATA cofactors are known to exert an inhibitory effect to complicate the GATA
607 factor-mediated transcriptional regulation [64,65,83]. Such an inhibitory effect of a

608 GATA cofactor potentially accounts for the transient *Pt-hh* expression in the early
609 abembryonic cell population. High degrees of sequence divergence in the DNA-
610 binding domain of Fuchi and other GATAL2 proteins may be an indication of evolving
611 interactions with cofactors and DNA-binding sites. Previous studies have suggested
612 that N-terminal zinc fingers (ZF1) influence selectivity of DNA-binding sequences
613 [82,84]. Alternatively, the sequence divergences might have simply resulted from
614 reducing degrees of functional constrains. Future studies will focus on testing these
615 possibilities.

616 The phylogenetic origin of the *fuchi* activity in early endoderm specification
617 and those of the regulatory connection between *fuchi* and *Pt-GATA5*, is an intriguing
618 issue that provides molecular clues regarding the diversification mechanism of the
619 process of early embryonic development. The formation of a sharply demarcated germ
620 disc is not a common feature of early spider development. In Theridiidae spider
621 embryos, the formation of the germ disc involves an appearance of two distinct
622 endoderm or endoderm-like cell populations derived from both polar regions of the
623 early embryo, and the formed sharp boundary of the germ disc serves as sites of
624 endoderm and mesoderm cell internalization, as well as sites of sending patterning
625 signals [31,33,42,45]. These developmental strategies have potential merits of gaining
626 parallel processes, which may allow for a more rapid development. In the embryos of
627 many other spider species, including three distinct examples from outside the
628 Araneoidea, *Cupiennius salei* [46], *Pholcus phalangioides* [47], and *Hasarius adansoni*
629 [85], cell movements causing the germ layers only occur from one of the polar regions,
630 which corresponds to the embryonic pole or the center of the germ disc in the *P.*
631 *tepidariorum* embryo. The technical term "germ disc" results in potential confusions in
632 documenting the Theridiidae embryonic development. This term is, in certain cases,
633 used to indicate the multicell-layered region of the embryo that results from germ layer
634 formation in spiders and other arthropods [86]. The initial germ disc in Theridiidae
635 embryos, however, is a form of a single epithelial cell layer that precedes the formation
636 of the germ layers [27,28,38]. *fuchi* serves as a molecular marker for the cells outside
637 the forming germ disc. *fuchi* homologs may aid in identifying the existence of an
638 evolutionarily conserved abembryonic cell type in other spider species. Complicatedly,
639 the *P. tepidariorum* genome has a paralog of *fuchi* that was presumably generated by

640 recent tandem gene duplication. No corresponding paralog was detected in other
641 Theridiidae and non-Theridiidae spiders. The *fuchi* paralog, *Pt-GATAL2b*, shows little
642 expression at least during the early and mid embryogenesis of *P. tepidariorum*. Studies
643 on the expression and function of *fuchi*/*Pt-GATAL2b* and *Pt-GATA5* orthologs in the
644 early embryos of other spider species enable us to further address the problems on the
645 early developmental variations among the Araneae lineages.
646 This study has produced genome-wide datasets related to molecular and
647 genetic regulation of early embryonic development in *P. tepidariorum*, which can be
648 integrated with single-nucleus and single-cell RNA-seq datasets established in another
649 study [61]. These data resources, as well as the research strategies presented, will
650 facilitate our efforts to dissect the early developmental processes at molecular
651 resolution and establish an independent paradigm of genetic programs for early
652 arthropod development. Importantly, many of the genes revealed in this study encode
653 proteins not yet characterized, thus underscoring the importance of taking genome-
654 based approaches to studying the evolution of animal development.
655
656 **Conclusions**
657 Using a Theridiidae spider empowered by genome sequencing, we combined
658 comparative transcriptomes of isolated cells from different regions of early embryos
659 and a pRNAi-based functional screen of genes. This research strategy allowed us to
660 identify a lineage-specific, fast-evolving GATA-like gene, *fuchi*, that is essential for
661 early embryonic development in the spider. *fuchi* is expressed in future endodermal cell
662 poupulations in the early embryo. Further genome-wide analyses suggested that *fuchi*
663 regulates chromatin state and zygotic gene activation to promote endoderm
664 specification and pattern formation. The presented datasets have rich information about
665 the molecular regulation of early embryonic development in the Theridiidae spider.
666 The presence of many uncharacterized genes under the control of *fuchi* has been
667 revealed. Our genome-based research using a chelicerate arthropod phylogentically
668 distant from *Drosophila* provides the foundation for molecular and genetic exploration
669 of the variations in early development across Arthropoda.
670
671 **Materials and Methods**

672 **Spiders**

673 Laboratory stocks of *P. tepidariorum* were maintained at 25 °C in 16 h light/8 h dark
674 cycles. Developmental stages of the embryo were determined according to the previous
675 descriptions [29,59]. To determine the precise time of egg laying, female behaviors
676 were recorded every 5 min using a trail camera. The animal experimentation described
677 was approved by the institutional committee for animal care and use (No. 2020-1) and
678 conducted according to JT Biohistory Research Hall Regulation on Animal
679 Experimentation.

680

681 **Cell isolation**

682 Embryos at stage 3 were dechorionated with 50% commercial bleach for 1 min and
683 rinsed several times with distilled water. After removing water, the embryos were
684 placed on a glass slide with double-sticky tape and immediately covered with
685 halocarbon oil 700 (Sigma-Aldrich). Glass capillaries (2-000-075; Drummond) were
686 prepared in advance by pulling with a puller (PN-3; Narishige). The capillary tip was
687 broken using forceps, just before use, and filled with lysis buffer supplied in Dynabeads
688 mRNA DIRECT Kit (Ambion) by capillary action. Upon the manual manipulation of
689 the capillary under the stereomicroscope, approximately 10–30 cells were sucked from
690 a central, an intermediate, or a peripheral region of the stage-3 embryo and transferred
691 into 0.2 ml tubes containing 10 µl of lysis buffer. The samples were stored at -80 °C
692 until they were retrieved for RNA extraction.

693

694 **RNA-seq of isolated cells**

695 Poly(A) mRNA was extracted using the Dynabeads mRNA DIRECT Kit (Ambion).
696 After fragmentation of the extracted RNA for 5 min at 94 °C, first-strand cDNA
697 synthesis and subsequent cDNA amplification by 22 cycles of polymerase chain
698 reaction (PCR) were performed using the SMARTer stranded RNA-seq kit (Takara).
699 The synthesized first-strand cDNA and amplified cDNA were purified using Ampure
700 XP Beads (Beckman Coulter). Constructed libraries for sequencing were quantified
701 using the Agilent 2100 BioAnalyzer high sensitivity DNA kit (Agilent). Two or three
702 libraries for each region of the stage-3 embryo were sequenced in single-end runs in
703 the antisense direction using the MiSeq reagent kit V3 (150 cycles) on the Illumina

704 MiSeq platform. Over 10 million reads were obtained from each library. Following
705 essentially the same procedure, RNA-seq of cells from central and peripheral regions
706 of stage 4 and 5 embryos was performed.
707
708 **Processing of RNA-seq data**
709 The adapter and primer sequences, as well as the first three bases derived from the
710 SMARTer Stranded Oligo, were trimmed from the MiSeq raw reads using CLC
711 Genomics Workbench 7.0.3 (Qiagen). Quality trimming was also performed at the
712 following parameter settings: quality score, limit = 0.05; trim ambiguous nucleotides,
713 maximum number of ambiguities = 2; and filter on length, discard reads below length
714 = 30. Trimmed reads were aligned to the *Parasteatoda tepidariorum* genome (Ptep_1.0,
715 GCA_000365465.1) using the BLAT algorithm [87] in the DDBJ Read Annotation
716 Pipeline (https://p.ddbj.nig.ac.jp/pipeline/). Output alignments were filtered using the
717 PERL script filterPSL.pl, which is accessible from the AUGUSTUS 3.0.1 scripts folder
718 (http://github.com/Gaius-Augustus/Augustus), with the following settings: 60%
719 coverage; 90% identity; uniqueness threshold 0.96. The filtered alignments (pslx-
720 format) were converted to sam-format. Based on these sam-format data, the number of
721 mapped reads was counted against the AUGUSTUS gene models (aug3.1;
722 https://i5k.nal.usda.gov/content/data-downloads)[32] using htseq-count v.0.6.1p1 [88]
723 with default settings. The RNA-seq datasets, including the gene count matrix, were
724 deposited in the NCBI Gene Expression Omnibus (GEO) database (GSE193511).
725
726 **Comparative analyses of RNA-seq gene count datasets**
727 To assess the consistency of RNA-seq gene count datasets between biological
728 replicates of each sample group, the coefficients of determination were calculated using
729 the gene count matrix, and confirmed to be exceed 0.85. Out of the 27,990 annotated
730 genes, 10,862 genes showed one or more counts per million (cpm) for at least one of
731 the eight samples (p1-3, i1-3, c1, and c2). Count data matrices on these genes (10,862
732 genes × 8 samples) were used for comparative analyses. To identify differentially
733 expressed genes (DEGs) using a Bioconductor package edgeR (version 3.8.6) [50],
734 three types of comparisons were set up: comparison I, c1 and c2 vs. p1-3 and i1-3;
735 comparison II, p1-p3 vs. c1, c2, and i1-3; and comparison III, i1-3 vs. c1, c2, and p1-

736 3. Two criteria were applied to determine the high-priority candidate genes: first,
737 $log_2FC < -10$ in comparisons I and II, and $log_2FC > 10$ in comparison III; and second,
738 lower FDR values.
739
740 **cDNA cloning**
741 Full-length or partial cDNAs of newly identified genes were obtained from our
742 collections of expressed sequence tag (EST) clones previously described [33][40], or
743 were isolated by PCR amplification. cDNA fragments cloned in pBluescript II SK(+)
744 (Agilent), pTriplEx2 (Takara) or pZL1 (Invitrogen) were used to synthesize double-
745 stranded RNA (dsRNA) for pRNAi and/or RNA probes for in situ hybridization. The
746 primers and cDNA clones used are summarized in Table S16 (Additional file 12).
747
748 **Parental RNAi**
749 For dsRNA synthesis, DNA templates with the T7 promoter sequence at both ends were
750 prepared by PCR using appropriate primers depending on the plasmid type used.
751 dsRNAs were synthesized using a Megascript Kit (Ambion) as described previously
752 [30]. One to two µl of the dsRNA solution at a concentration of 2.0 µg/µl was injected
753 into the abdomen of adult females four times at 2–3 day-intervals using pulled glass
754 capillary tubes. In the pilot pRNAi screen, dsRNA against each gene was injected into
755 two or four females. When egg laying took place, more than 10 eggs were randomly
756 selected from each egg sac and transferred into halocarbon oil 700, which allowed us
757 to nondestructively monitor the embryonic development through cleared chorion.
758 For three genes (*g26874*, *g7720*, *g4238*) that were judged as positive in the
759 pilot screen, 2 or 3 additional dsRNAs were prepared from non-overlapping regions of
760 each of the cDNAs, and each dsRNA was injected into at least 2 females. Development
761 of eggs laid by the injected females were monitored using time-lapse recording, or with
762 occasional visual inspection, under the stereo microscope. *gfp* dsRNA was used as the
763 control.
764
765 **Time-lapse microscopy of live embryos**
766 Embryos dechorionated with 50 % commercial bleach were placed on a glass slide with
767 double-sticky tape and covered with halocarbon oil 700. When the observation was

768 started from stage 1, the dechorionation step was omitted to prevent possible effect on
769 the development. Images were taken every 5 or 10 min using a stereo microscope
770 (SZX12, Olympus) equipped with a color CCD camera (C7780-10, Hamamatsu
771 Photonics) controlled by AquaCosmos software (Hamamatsu Photonics) or other stereo
772 microscopes (M165C, Leica; or SZX7, Olympus) equipped with a color CMOS camera
773 (WRAYCAM-G200 or -NF300, WRAYMER).
774 For live imaging of nuclei around the germ-disc/non-germ-disc transition area,
775 dechorionated stage-4 embryos were placed on a glass slide with double-sticky tape,
776 and an embryo appropriately oriented was selected and skewered with a thin glass
777 needle to prevent the embryo from rotating during later development. A 1,000× stock
778 solution of a cell-permeable dye SPY555-DNA (Spirochrome), which was prepared in
779 50 µl of dimethylsulfoxide, was microinjected into the perivitelline space of the
780 skewered egg. The injecting volume should be low to avoid affecting the development
781 of the embryo from being affected. Approximately 1 hour after dye injection, the
782 embryo started to be examined under an Olympus BX50 fluorescence microscope
783 equipped with a cooled CCD camera (CoolSNAP HQ, Roper Scientific) controlled by
784 MetaMorph.version 6.1 (Universal Imaging). Using 10 × objective lens (UPlanAPO
785 10×/0.40, Olympus), bright-field and fluorescence 500 × 500 pixel images were
786 acquired every 5 min at exposure times of 10 msec and 100 msec, respectively. To
787 reduce the excitation light, two neutral density filters (U-ND6-2 and U-ND25-2,
788 Olympus) were used. A single focal plane, which was manually adjusted during
789 approximately 30 h of observation, was recorded at each time point. The observed
790 embryo was checked for its normal development. Time-lapse images were analyzed
791 using Imaris version 7.6.5, where cells were manually tracked, and ImageJ version
792 1.51d.
793
794 **Molecular phylogenetic analysis of GATA family members**
795 To collect amino acid sequences containing GATA-type DNA binding domain from
796 many arachnid species and other arthropod and nonarthropod metazoan representative
797 species, we used nucleotide sequence resources presented in Table S7 (Additional file
798 7), which included coding sequences from sequenced genomes, transcriptome
799 assemblies, and RNA-seq raw data, retrieved from GenBank (NCBI) or GigaScience

800 database. RNA-seq of *Hasarius adansoni* mid-stage embryos was performed in this
801 study, as described below. The RNA-seq data were subjected to de novo assembling
802 using Trinity (version 2.11.0). To identify amino acid sequences coding for GATA-
803 type DNA binding domains in the retrieved or assembled nucleotide sequences, blastp
804 searches were performed using the Fuchi amino acid sequence as a query with a cut-
805 off e-value of 1e-2, followed by the systematic selection of sequences that contained at
806 least one GATA-type zinc-finger motif sequence ($CX_2CX_{16\text{-}20}CX_2C$). The identified
807 amino acid sequences were manually aligned and classified based on the detection of
808 signature sequences, excluding GATA-like proteins that had only one GATA-type zinc
809 finger motif. For molecular phylogenetic analysis, 75 amino acid sites from 118
810 sequences, which were unambiguously aligned, were used (Additional file 7: Table S8).
811 The molecular phylogeny was inferred using the Maximum Likelihood (ML) method
812 and JTT matrix-based model [89] on the software MEGA11 [90].
813
814 **RNA-seq of whole embryos**
815 For RNA-seq analyses of *fuchi* pRNAi embryos, mated adult females were injected
816 with approximately 1–2 µl of *fuchi* dsRNA solution (2 µg/µl) four times at 2- to 3-day
817 intervals. For RNA extraction, 50–100 embryos at stage 2, 3, and early stage 5 that
818 were derived from egg sacs produced by the injected females 2days before (untreated)
819 and 18–19 days after the first injection of the dsRNA were used. Two biological
820 replicates were prepared for each sample type. The mRNA extraction and library
821 construction for sequencing were performed as previously described [37]. The libraries
822 were quantified using the Library quantification kit (Takara) and Thermal cycler Dice
823 Realtime TP800 (Takara) and sequenced in in single-end runs (150 cycles) in the
824 antisense direction on the Illumina MiSeq platform. The sequence reads were processed
825 as previously described [37], and the datasets were deposited in the GEO database
826 (GSE193650). Estimates of DEGs between untreated and *fuchi* pRNAi embryos at each
827 stage were performed using EdgeR (version 3.8.6).
828 For the *H. adansoni* RNA-seq, poly(A) mRNA was extracted using a
829 QuickPrep Micro mRNA purification kit (GE Healthcare) from 20 sibling mid-stage
830 embryos with extending limbs. Library construction for sequencing was performed as
831 previously described [37]. The library was sequenced in a paired-end run (150 × 2

832 cycles) on the Illumina Miseq platform, and the raw reads (SRR17326784) and the *H.*
833 *adansoni* transcriptome assembly (GJQJ00000000) were deposited in GenBank.
834
835 **ATAC-seq of whole embryos**
836 ATAC-seq was performed following the procedure described by Buenrostro et al.
837 (2015) [91] with modifications considering the procedure reported by Haines and
838 Eisen (2018) [92]. In brief, 40 whole embryos at stage 3 (18 h AEL) were
839 homogenized using disposable pellet pestles (12-141-368; Fisher Scientific) and 1.5
840 mL Eppendorf tubes (0030125150) in 50 µl lysis buffer (10mM Tris, 10mM NaCl,
841 3mM $MgCl_2$, adjusted to pH 7.4 with HCl) containing 0.15 mM spermidine, after
842 which 50 µl of lysis buffer and 1 µl of 10% Nonidet P-40 were added as the pestle
843 was rinsed. After a 10 min incubation, nuclei were spun down at 800 g and
844 resuspended in 17.5 µl water, followed by the addition of 25 µl of 2× TD buffer
845 (Illumina) and 7.5 µl of Nextra Tn5 transposase (Illumina). Transposed DNA was
846 purified using MinElute Reaction Cleanup Kit (Qiagen) and amplified with Illumina
847 Nextera Transposase Adaptors index PCR primers, and the amplified DNA was
848 purified using Ampure XP Beads (Beckman Coulter). Constructed libraries were
849 paired-end sequenced (151 bp reads × 2) on the Illumina HiSeq X Ten system.
850 Comparative analyses of ATAC-seq datasets from three different sample
851 types (wild-type, *fuchi* pRNAi, and *Pt-hh* pRNAi) were performed basically
852 following the procedure described by Reske et al. (2020) [93] with slight
853 modifications based on guidelines provided by [94] and [95]. Two biological
854 replicates were obtained from each sample type. ATAC-seq raw reads were trimmed
855 for adaptor and primer sequences and low-quality sequences using Trim Galore! with
856 default settings. Trimmed reads were then aligned to the *P. tepidariorum* genome
857 (Ptep_2.0) using bowtie2 v2.2.5 with default settings. Reads mapped with low quality
858 scores (less than 30) were removed using samtools (version 1.9). PCR duplicates
859 were removed using Picard MarkDuplicates (version 2.23.4). The alignments of
860 filtered reads to the genome were saved as bam files. Peak calling was performed
861 using the ATAC-seq mode of Genrich (version 6.0). Differential accessibility analysis
862 was performed using a Bioconductor package, *csaw* [96]. Peak regions called in both
863 replicates of each sample type were extracted, and the peak regions extracted from

the three sample types were merged. Reads were counted in peaks, with low abundance peaks (log CPM < -3) removed. Read counts were subject to normalization with the trimmed mean of *M* values (TMM) method [97]. Differential accessibility was estimated using the glmQLFit and glmQLFTest functions in EdgeR (version 3.32.0). Peaks with FDR < 0.05 were considered as significant. To discover motif sequences in the identified differential peak regions, STREME [98] was used. Wild-type whole embryos at stages 1 (9 h AEL) and 2 (12 h AEL) were similarly analyzed by ATAC-seq to detect genomic regions differentially accessible between the different stages (stages 1 to 3), but no clear results were obtained. All the ATAC-seq datasets were deposited in the GEO database (GSE193870).

**Staining of embryos**

Antisense RNA probes for in situ hybridization were prepared by *in vitro* transcription using Digoxigenin- or Dinitrophenyl-conjugated UTP (DIG, Roche 11277073910; DNP, PerkinElmer NEL555001EA) as previously described [57]. Whole-mount in situ hybridization with chromogenic substrate was performed in the previously described manner [30] with the exception of an additional tyramide-based signal amplification (TSA) process described previously [57]. In brief, after washing DIG-conjugated RNA probes, the samples were incubated with anti-DIG-POD antibody (11633716001, Roche; 1:500 dilution) and then reacted with DNP-tyramide (NEL747, PerkinElmer). In the detection step, anti-DNP-AP antibody (MB-3100, Vector; 1:50) and NBT/BCIP were used. Most samples were counterstained with 4',6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich).

Double-color FISH was performed as previously described [57]. To visualize the DIG- and DNP-conjugated probes, anti-DIG-POD (Roche 11633716001; 1:500) and anti-DNP-HRP (PerkinElmer FP1129; 1:100) were used in combination with DyLight488-tyramide and DyLight680-tyramide, respectively.

Single- and double-color FISH were combined with immunostaining for β-catenin in certain cases (Figs. 5 and 6), where ProK treatment prior to hybridization was omitted. After completing the FISH procedure, the embryos were incubated with rabbit anti-β-catenin antibody (C2206, Sigma-Aldrich 1:500) and then with Cy3-

895 conjugated anti-rabbit IgG antibody (AP187C, Chemicon; 1:200), followed by
896 counterstaining with DAPI.
897
898 **Image acquisition and processing**
899 Images for chromogenically stained WISH samples were obtained using a stereo
900 microscope (SZX12, Olympus), a fluorescence unit BT-ExSMOP (Biotools), and a
901 CCD camera (C7780-10, Hamamatsu Photonics). For confocal microscopy of
902 fluorescently stained embryos, the whole embryos were individually mounted on glass
903 slides with spacers of approximately 100–500 μm, which were adjusted depending on
904 the sample using sticky tape (CT-18S, Nichiban) and silicone vacuum grease (335148,
905 Beckman). To view different sides of a single embryo, it was rotated after being
906 mounted on the glass slide. The embryos were examined using a Leica TCS SPE
907 confocal system equipped with four laser sources (405, 488, 532, and 635 nm), which
908 were used to excite DAPI, DyLight488, Cy3, and DyLight680, respectively. Acquired
909 confocal stacks were processed and analyzed using Imaris version 7.6.5 (Bitplane) and
910 ImageJ version 1.51d. The Imaris snapshot feature and oblique slicer tool were used to
911 obtain images capturing the embryonic regions of interest. Linear signals of cell-cell
912 adherens junctions visualized by β-catenin staining were traced in Adobe Photoshop
913 CS5 ver. 12.0.4.
914
915 **Availability of data and materials**
916 Raw and processed data for the comparative RNA-seq and ATAC-seq analyses are
917 accessible from NCBI GEO (GSE193511, GSE193650, and GSE193870). The data
918 aligned to the reference genome can be viewed on our web database
919 (https://www.brh2.jp). The nucleotide sequences determined for *fuchi* and *Pt-GATA5*
920 are available under the accessions LC671429 and LC671430 in the
921 DDBJ/EMBL/GenBank International Nucleotide Sequence Databases. The sources of
922 sequence data used in the molecular phylogenetic analysis are listed in Additional file
923 7: Tables S7 and S8. Supplementary movies showing the pilot pRNAi screen and the
924 pRNAi spescificity validation are available on figshare [51].
925
926 **Funding**

927 This work was supported by JSPS Kakenhi (26440130, 17K07418, 20K06676 to YA;
928 15K07139 to HO) and JST PRESTO (JPMJPR2041 to YA).
929
930 **Acknowledgements**
931 We would like to thank Akiko Noda for technical assistance, and other members of JT
932 Biohistory Research Hall for discussion.
933
934 **Competing interests**
935 Authors declare no competing interests.
936
937 **Supplementary Information**
938 **Additional file 1: Movie S1.** Time-lapse video of sibling *Parasteatoda tepidariorum*
939 embryos. The left embryo is the same as shown in Figure 1. Time after egg laying
940 (AEL) is indicated. Scale bar, 100 µm.
941 **Additional file 2: Table S1.** List of DEG candidates identified by comparative
942 transcriptome analysis of c versus i/p cells from stage-3 embryo (comparison I). **Table**
943 **S2.** List of DEG candidates identified by comparative transcriptome analysis of p
944 versus c/i cells from stage-3 embryo (comparison II)**. Table S3.** List of DEG candidates
945 identified by comparative transcriptome analysis of i versus c/p cells from stage-3
946 embryo (comparison III)**. Table S4.** List of DEG candidates indentified by comparative
947 transcriptome analysis of c versus p cells from stage-4 embryo**. Table S5.** List of DEG
948 candidates identified by comparative transcriptome analysis of c versus p cells from
949 early stage-5 embryo.
950 **Additional file 3: Figure S1.** Chromogenic WISH of stage-3 embryos using probes for
951 19 selected DEG candidates. **Figure S2.** Validation of the specificity of the RNAi
952 effects for *g26874*, *g7720*, and *g4238*. **Figure S3.** Developmental transcript profiling
953 of *P. tepidariorum* GATA family genes in wild-type embryos. **Figure S4.** Cell counting
954 in a stage-3 embryo stained for *fuchi*. **Figure S5.** *fuchi* expression in pEND cells.
955 **Figure S6.** Characterization of 5 DEG candidates from the comparative transcriptome
956 analysis of stage-2 *fuchi* pRNAi versus untreated embryos. **Figure S7.** Characterization
957 of the top-10 DEG candidates from the comparative transcriptome analysis of stage-3
958 *fuchi* pRNAi versus untreated embryos. **Figure S8.** Effects of *fuchi* RNAi on

959 expression of selected genes. **Figure S9.** Single-nucleus transcriptome data showing
960 endodermal and mesodermal cell populations in late stage-5 embryos. **Figure S10.**
961 Single-nucleus and single-cell transcriptome data showing cell populations expressing
962 the genes regulated by *fuchi*.
963 **Additional file 4: Table S6.** Summary of a pilot pRNAi screen of 19 selected genes.
964 **Additional file 5: Movie S2.** Time-lapse video showing phenotypes *gfp*, *g26874*,
965 *g7720*, and *g4238* pRNAi embryos. Time after egg laying (AEL) and time after start of
966 recording are indicated. Movie starts are adjusted by the timing of blastoderm
967 formation. Some of the embryos are the same as those shown in Figure 3A. Scale bar,
968 100 μm.
969 **Additional file 6: Movie S3.** Tracking of cells in wild-type and *g26874* pRNAi
970 embryos. In the *g26874* pRNAi embryo, cells on the abembryonic side initially shift
971 toward forming a germ disc but, later, they return back. Time after start of recording is
972 indicated.
973 **Additional file 7: Table S7.** List of bioinformatic resources used for phylogenetic
974 characterization of Fuchi and other GATA family members in spiders. **Table S8.** Amino
975 acid sequence alignment and classification of GATA family members from spiders and
976 other metazoans.
977 **Additional file 8: Movie S4.** Time-lapse video showing cell internalizations taking
978 place around the germ-disc rim. Nuclei labeled by SPY555-DNA are presented in white.
979 The germ disc on the embryonic side is to the lower right, while the non-germ-disc
980 region on the abembryonic side is to the upper left. Internalizing cells are tracked. Cells
981 from the non-germ-disc region are moving to below the germ-disc epithelium. Later,
982 cells at the germ-disc rim internalize. Time [d : h : m] after start of recording is indicated.
983 **Additional file 9: Table S9.** List of DEGs identified by comparative transcriptome
984 analysis of *fuchi* pRNAi versus untreated embryos at stage 2. **Table S10.** List of DEGs
985 identified by comparative transcriptome analysis of *fuchi* pRNAi versus untreated
986 embryos at stage 3. **Table S11.** List of DEGs identified by comparative transcriptome
987 analysis of *fuchi* pRNAi versus untreated embryos at early stage 5.
988 **Additional file 10: Table S12.** Data from comparative analysis of read counts in
989 extracted ATAC-seq peaks between *fuchi* pRNAi and wild-type embryos at stage 3

990 using edgeR. **Table 13.** List of differential ATAC-seq peaks between *fuchi* pRNAi and
991 wild-type embryos at stage 3 (FDR < 0.05).
992 **Additional file 11: Table S14.** Data from comparative analysis of read counts in
993 extracted ATAC-seq peaks between *Pt-hh* pRNAi and wild-type embryos at stage 3
994 using edgeR. **Table S15.** List of differential ATAC-seq peaks between Pt-hh pRNAi
995 and wild-type embryos at stage 3 (FDR < 0.05).
996 **Additional file 12: Table S16.** Primers and cDNA clones used for RNAi and in situ
997 hybridization.
998
999 **References**


1000 1. Raff RA. The Shape of Life: Genes, Development, and the Evolution of Animal Form. Chicago: Univ.
1001 Chicago Press; 1996.

1002 2. Duboule D. Temporal colinearity and the phylotypic progression: a basis for the stability of a
1003 vertebrate Bauplan and the evolution of morphologies through heterochrony. Dev Suppl. 1994;135–42.

1004 3. Galis F, van Dooren TJM, Metz JAJ. Conservation of the segmented germband stage: robustness or
1005 pleiotropy? Trends Genet. 2002;18:504–9.

1006 4. Sander K. The evolution of patterning mechanisms: gleanings from insect embryogenesis and
1007 spermatogenesis. In: Goodwin BC Hold N Wylie CC. (eds) Development and evolution. Cambridge:
1008 Cambridge University Press; 1983. p. 137–60.

1009 5. Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, et al. Gene expression
1010 divergence recapitulates the developmental hourglass model. Nature. 2010;468:811–4.

1011 6. Richardson MK. Heterochrony and the Phylotypic Period. Dev Biol. 1995;172:412–21.

1012 7. Irie N, Kuratani S. Comparative transcriptome analysis reveals vertebrate phylotypic period during
1013 organogenesis. Nat Commun. 2011;2:248.

1014 8. Nüsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in Drosophila.
1015 Nature. 1980;287:795–801.

1016 9. Domazet-Loso T, Tautz D. An Evolutionary Analysis of Orphan Genes in Drosophila. Genome Res.
1017 2003;13:2213–9.

10. Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG. More than just orphans: are taxonomically-restricted genes important in evolution? Trends in Genetics. 2009;25:404–13.

11. Thomas GWC, Dohmen E, Hughes DST, Murali SC, Poelchau M, Glastad K, et al. Gene content evolution in the arthropods. Genome Biol. 2020;21:15.

12. Morino Y, Hashimoto N, Wada H. Expansion of TALE homeobox genes and the evolution of spiralian development. Nat Ecol Evol. 2017;1:1942–9.

13. Holland LZ, Albalat R, Azumi K, Benito-Gutiérrez È, Blow MJ, Bronner-Fraser M, et al. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res. 2008;18:1100–11.

14. Khalturin K, Anton-Erxleben F, Sassmann S, Wittlieb J, Hemmrich G, Bosch TCG. A Novel Gene Family Controls Species-Specific Morphological Traits in Hydra. PLOS Biol. Public Library of Science; 2008;6:e278.

15. Stauber M, J\textbackslashHackle H, Schmidt-Ott U. The anterior determinant bicoid of Drosophila is a derived Hox class 3 gene. Proc Natl Acad Sci U A. 1999;96:3786–9.

16. Lynch JA, Peel AD, Drechsler A, Averof M, Roth S. EGF signaling and the origin of axial polarity among the insects. Curr Biol. 2010;20:1042–7.

17. Lima LF, Torres AQ, Jardim R, Mesquita RD, Schama R. Evolution of Toll, Spatzle and MyD88 in insects: the problem of the Diptera bias. BMC Genomics. 2021;22:562.

18. Nishida H, Sawada K. macho-1 encodes a localized mRNA in ascidian eggs that specifies muscle fate during embryogenesis. Nature. 2001;409:724–9.

19. Fekany K, Yamanaka Y, Leung T, Sirotkin HI, Topczewski J, Gates MA, et al. The zebrafish bozozok locus encodes Dharma, a homeodomain protein essential for induction of gastrula organizer and dorsoanterior embryonic structures. Development. 1999;126:1427–38.

20. Bae S, Reid CD, Kessler DS. Siamois and Twin are redundant and essential in formation of the Spemann organizer. Dev Biol. 2011;352:367–81.

1043 21. Pultz MA, Zimmerman KK, Alto NM, Kaeberlein M, Lange SK, Pitt JN, et al. A Genetic Screen for
1044 Zygotic Embryonic Lethal Mutations Affecting Cuticular Morphology in the Wasp Nasonia vitripennis.
1045 Genetics. 2000;154:1213–29.

1046 22. Sulston IA, Anderson KV. Embryonic patterning mutants of Tribolium castaneum. Development.
1047 1996;122:805–14.

1048 23. Maderspacher F, Bucher G, Klingler M. Pair-rule and gap gene mutants in the flour beetle Tribolium
1049 castaneum. Dev Genes Evol. 1998;208:558–68.

1050 24. Schmitt-Engel C, Schultheis D, Schwirz J, Ströhlein N, Troelenberg N, Majumdar U, et al. The
1051 iBeetle large-scale RNAi screen reveals gene functions for insect development and physiology. Nat
1052 Commun. 2015;6:7822.

1053 25. Consortium i5K. The i5K Initiative: advancing arthropod genomics for knowledge, human health,
1054 agriculture, and the environment. J Hered. 2013;104:595–600.

1055 26. Oda H, Akiyama-Oda Y. The common house spider Parasteatoda tepidariorum. EvoDevo. 2020;11:6.

1056 27. Kanayama M, Akiyama-Oda Y, Oda H. Early embryonic development in the spider Achaearanea
1057 tepidariorum: microinjection verifies that cellularization is complete before the blastoderm stage.
1058 Arthropod Struct Dev. 2010;39:436–45.

1059 28. Suzuki H, Kondo A. Early embryonic development, including germ-disk stage, in the Theridiid
1060 spider Achaearanea japonica (Bös. et Str.). J Morphol. 1995;224:147–57.

1061 29. Akiyama-Oda Y, Oda H. Early patterning of the spider embryo: a cluster of mesenchymal cells at
1062 the cumulus produces Dpp signals received by germ disc epithelial cells. Development. 2003;130:1735–
1063 47.

1064 30. Akiyama-Oda Y, Oda H. Axis specification in the spider embryo: dpp is required for radial-to-axial
1065 symmetry transformation and sog for ventral patterning. Development. 2006;133:2347–57.

1066 31. Akiyama-Oda Y, Oda H. Cell migration that orients the dorsoventral axis is coordinated with
1067 anteroposterior patterning mediated by Hedgehog signaling in the early spider embryo. Development.
1068 2010;137:1263–73.

1069 32. Schwager EE, Sharma PP, Clarke T, Leite DJ, Wierschin T, Pechmann M, et al. The house spider
1070 genome reveals an ancient whole-genome duplication during arachnid evolution. BMC Biol. 2017;15:62.

1071 33. Oda H, Nishimura O, Hirao Y, Tarui H, Agata K, Akiyama-Oda Y. Progressive activation of Delta-
1072 Notch signaling from around the blastopore is required to set up a functional caudal lobe in the spider
1073 Achaearanea tepidariorum. Development. 2007;134:2195–205.

1074 34. Posnien N, Zeng V, Schwager EE, Pechmann M, Hilbrant M, Keefe JD, et al. A comprehensive
1075 reference transcriptome resource for the common house spider Parasteatoda tepidariorum. PLoS One.
1076 2014;9:e104885.

1077 35. Kono N, Nakamura H, Ito Y, Tomita M, Arakawa K. Evaluation of the impact of RNA preservation
1078 methods of spiders for de novo transcriptome assembly. Mol Ecol Resour. 2016;16:662–72.

1079 36. Sasaki M, Akiyama-Oda Y, Oda H. Evolutionary origin of type IV classical cadherins in arthropods.
1080 BMC Evol Biol. 2017;17:142.

1081 37. Iwasaki-Yokozawa S, Akiyama-Oda Y, Oda H. Genome-scale embryonic developmental profile of
1082 gene expression in the common house spider Parasteatoda tepidariorum. Data Brief. 2018;19:865–7.

1083 38. Pechmann M. Formation of the germ-disc in spider embryos by a condensation-like mechanism.
1084 Front Zool. 2016;13:35.

1085 39. Hemmi N, Akiyama-Oda Y, Fujimoto K, Oda H. A quantitative study of the diversity of stripe-
1086 forming processes in an arthropod cell-based field undergoing axis formation and growth. Dev Biol.
1087 2018;437:84–104.

1088 40. Kanayama M, Akiyama-Oda Y, Nishimura O, Tarui H, Agata K, Oda H. Travelling and splitting of
1089 a wave of hedgehog expression involved in spider-head segmentation. Nat Commun. 2011;2:500.

1090 41. Akiyama-Oda Y, Oda H. Hedgehog signaling controls segmentation dynamics and diversity via
1091 msx1in a spider embryo. Sci Adv. 2020;6.

1092 42. Montgomery J TH. The development of Theridium, an Aranead, up to the stage of reversion. J
1093 Morphol. 1909;20:297–352.

43. Chaw RC, Vance E, Black SD. Gastrulation in the spider Zygiella x-notata involves three distinct phases of cell internalization. Dev Dyn. 2007;236:3484–95.

44. Mittmann B, Wolff C. Embryonic development and staging of the cobweb spider Parasteatoda tepidariorum C. L. Koch, 1841 (syn.: Achaearanea tepidariorum; Araneomorphae; Theridiidae). Dev Genes Evol. 2012;222:189–216.

45. Edgar A, Bates C, Larkin K, Black S. Gastrulation occurs in multiple phases at two distinct sites in Latrodectus and Cheiracanthium spiders. Evodevo. 2015;6:33.

46. Wolff C, Hilbrant M. The embryonic development of the central American wandering spider Cupiennius salei. Front Zool. 2011;8:15.

47. Turetzek N, Prpic N-MM. Observations on germ band development in the cellar spider Pholcus phalangioides. Dev Genes Evol. 2016;226:413–22.

48. Pechmann M. Embryonic development and secondary axis induction in the Brazilian white knee tarantula Acanthoscurria geniculata, C. L. Koch, 1841 (Araneae; Mygalomorphae; Theraphosidae). Dev Genes Evol. 2020;230:75–94.

49. Sekiguchi K. Reduplication in spider eggs produced by centrifugation. Sci Rep Tokyo Bunrika Daigaku Sec B. 1957;8:227–80.

50. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.

51. Iwasaki-Yokozawa S, Akiyama-Oda Y, Oda H. Supplementary movies for “Lineage-specific, fast-evolving GATA-like gene regulates zygotic gene activation to promote endoderm specification and pattern formation in the Theridiidae spider” [Internet]. figshare. 2022 [cited 2022 Jun 8]. Available from: https://figshare.com/account/home#/collections/5963661

52. Su Y, Song Y, Wang Y, Jessop L, Zhan L, Stumph WE. Characterization of a Drosophila Proximal-Sequence-Element-Binding Protein Involved in Transcription of Small Nuclear RNA Genes. Eur J Biochem. 1997;248:231–7.

53. Chen T, Bunting M, Karim FD, Thummel CS. Isolation and characterization of five Drosophila genes that encode an ets-related DNA binding domain. Dev Biol. 1992;151:176–91.

1121 54. Pechmann M, Benton MA, Kenny NJ, Posnien N, Roth S. A novel role for Ets4in axis specification
1122 and cell migration in the spider Parasteatoda tepidariorum. Elife. 2017;6.

1123 55. Tang Y, Wei Y, He W, Wang Y, Zhong J, Qin C. GATA transcription factors in vertebrates:
1124 evolutionary, structural and functional interplay. Mol Genet Genomics. 2014;289:203–14.

1125 56. Patient RK, McGhee JD. The GATA family (vertebrates and invertebrates). Curr Opin Genet Dev.
1126 2002;12:416–22.

1127 57. Akiyama-Oda Y, Oda H. Multi-color FISH facilitates analysis of cell-type diversification and
1128 developmental gene regulation in the Parasteatoda spider embryo. Dev Growth Differ. 2016;58:215–24.

1129 58. Oda H, Akiyama-Oda Y. Microarray data on the comparison of transcript expression between normal
1130 and Pt-DeltaRNAi embryos in the common house spider Parasteatoda tepidariorum. Data Brief.
1131 2019;25:104350.

1132 59. Yamazaki K, Akiyama-Oda Y, Oda H. Expression patterns of a twist-related gene in embryos of the
1133 spider Achaearanea tepidariorum reveal divergent aspects of mesoderm development in the fly and
1134 spider. Zool Sci. 2005;22:177–85.

1135 60. Iwafuchi-Doi M, Zaret KS. Pioneer transcription factors in cell reprogramming. Genes Dev.
1136 2014;28:2679–92.

1137 61. Akiyama-Oda Y, Akaiwa T, Oda H. Reconstruction of the Global Polarity of an Early Spider Embryo
1138 by Single-Cell and Single-Nucleus Transcriptome Analysis. submitted.

1139 62. Rebagliati MR, Weeks DL, Harvey RP, Melton DA. Identification and cloning of localized maternal
1140 RNAs from xenopus eggs. Cell. 1985;42:769–77.

1141 63. Sasai Y, Lu B, Steinbeisser H, Geissert D, Gont LK, De Robertis EM. Xenopus chordin: a novel
1142 dorsalizing factor activated by organizer-specific homeobox genes. Cell. 1994;79:779–90.

1143 64. Kozhemyakina E, Ionescu A, Lassar AB. GATA6 is a crucial regulator of Shh in the limb bud. PLoS
1144 Genet. 2014;10:e1004072.

1145 65. Tokusumi Y, Tokusumi T, Stoller-Conrad J, Schulz RA. Serpent, Suppressor of Hairless and U-
1146 shaped are crucial regulators of hedgehog niche expression and prohemocyte maintenance during
1147 Drosophila larval hematopoiesis. Development. 2010;137:3561–8.

1148 66. Xuan S, Sussel L. GATA4 and GATA6 regulate pancreatic endoderm identity through inhibition of
1149 hedgehog signaling. Development. 2016;143:780–6.

1150 67. Fujikura J, Yamato E, Yonemura S, Hosoda K, Masui S, Nakao K, et al. Differentiation of embryonic
1151 stem cells is induced by GATA factors. Genes Dev. 2002;16:784–9.

1152 68. Tremblay M, Sanchez-Ferras O, Bouchard M. GATA transcription factors in development and
1153 disease. Development. 2018;145:dev164384.

1154 69. Gillis WQ, Bowerman BA, Schneider SQ. The evolution of protostome GATA factors: molecular
1155 phylogenetics, synteny, and intron/exon structure reveal orthologous relationships. BMC Evol Biol.
1156 2008;8:112.

1157 70. Reuter R. The gene serpent has homeotic properties and specifies endoderm versus ectoderm within
1158 the Drosophila gut. Development. 1994;120:1123–35.

1159 71. Morrisey EE, Tang Z, Sigrist K, Lu MM, Jiang F, Ip HS, et al. GATA6 regulates HNF4 and is
1160 required for differentiation of visceral endoderm in the mouse embryo. Genes Dev. 1998;12:3579–90.

1161 72. Shivdasani RA. Molecular Regulation of Vertebrate Early Endoderm Development. Dev Biol.
1162 2002;249:191–203.

1163 73. Tam PP, Kanai-Azuma M, Kanai Y. Early endoderm development in vertebrates: lineage
1164 differentiation and morphogenetic function. Curr Opin Genet Dev. 2003;13:393–400.

1165 74. Fukuda K, Kikuchi Y. Endoderm development in vertebrates: fate mapping, induction and regional
1166 specification. Dev Growth Differ. 2005;47:343–55.

1167 75. Frankenberg S, Gerbe F, Bessonnard S, Belville C, Pouchin P, Bardot O, et al. Primitive Endoderm
1168 Differentiates via a Three-Step Mechanism Involving Nanog and RTK Signaling. Dev Cell.
1169 2011;21:1005–13.

76. Heslop JA, Pournasr B, Liu J-T, Duncan SA. GATA6 defines endoderm fate by controlling chromatin accessibility during differentiation of human-induced pluripotent stem cells. Cell Rep. 2021;35:109145.

77. Feitosa NM, Pechmann M, Schwager EE, Tobias-Santos V, McGregor AP, Damen WGM, et al. Molecular control of gut formation in the spider Parasteatoda tepidariorum. Genesis. 2017;55.

78. Maduro MF, Rothman JH. Making Worm Guts: The Gene Regulatory Network of the Caenorhabditis elegans Endoderm. Dev Biol. 2002;246:68–85.

79. Murakami R, Okumura T, Uchiyama H. GATA factors as key regulatory molecules in the development of Drosophila endoderm. Dev Growth Differ. 2005;47:581–9.

80. Tsang AP, Visvader JE, Turner CA, Fujiwara Y, Yu C, Weiss MJ, et al. FOG, a Multitype Zinc Finger Protein, Acts as a Cofactor for Transcription Factor GATA-1 in Erythroid and Megakaryocytic Differentiation. Cell. 1997;90:109–19.

81. Crispino JD, Lodish MB, MacKay JP, Orkin SH. Use of Altered Specificity Mutants to Probe a Specific Protein–Protein Interaction in Differentiation: the GATA-1:FOG Complex. Mol Cell. 1999;3:219–28.

82. Waltzer, Lucas, Bataille, Laetitia, Peyrefitte, Sandrine, Haenlin, Marc. Two isoforms of Serpent containing either one or two GATA zinc fingers have different roles in Drosophila haematopoiesis. EMBO J. 2002;21:5477–86.

83. Chlon TM, Doré LC, Crispino JD. Cofactor-mediated restriction of GATA-1 chromatin occupancy coordinates lineage-specific gene expression. Mol Cell. 2012;47:608–21.

84. Pedone, Paolo, V., Omichinski, James, G., Nony, Pascale, Trainor, Cecelia, Gronenborn, Angela, M., Clore, G.Marius, et al. The N-terminal fingers of chicken GATA-2 and GATA–3 are independent sequence-specific DNA binding domains. EMBO J. 1997;16:2874–82.

85. Oda H, Iwasaki-Yokozawa S, Usui T, Akiyama-Oda Y. Experimental duplication of bilaterian body axes in spider embryos: Holm's organizer and self-regulation of embryonic fields. Dev Genes Evol. 2020;230:49–63.

86. Holm $Å$. Experimentelle Untersuchungen über die Entwicklung und Entwicklungsphysiologie des Spinnenembryos. Zool BiDr Upps. 1952;29:293–424.

1197 87. Kent WJ. BLAT—The BLAST-Like Alignment Tool. Genome Res. 2002;12:656–64.

1198 88. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing
1199 data. Bioinformatics. 2015;31:166–9.

1200 89. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein
1201 sequences. Bioinformatics. 1992;8:275–82.

1202 90. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11.
1203 Mol Biol Evol. 2021;38:3022–7.

1204 91. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin
1205 Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015;109:21.29.1-9.

1206 92. Haines JE, Eisen MB. Patterns of chromatin accessibility along the anterior-posterior axis in the
1207 early Drosophila embryo. PLoS Genet. 2018;14:e1007367.

1208 93. Reske JJ, Wilson MR, Chandler RL. ATAC-seq normalization method can significantly affect
1209 differential accessibility analysis and interpretation. Epigenetics Chromatin. 2020;13:22.

1210 94. Gaspar JM. ATAC-seq Guidelines [Internet]. Harv. FAS Inform. 2019 [cited 2022 Apr 26]. Available
1211 from: https://informatics.fas.harvard.edu/./atac-seq-guidelines.html

1212 95. Delisle, L., Doyle, M., Heyl, F. Galaxy Training: ATAC-Seq data analysis [Internet]. 2021 [cited
1213 2022 Apr 26]. Available from: https://training.galaxyproject.org/training-
1214 material/topics/epigenetics/tutorials/atac-seq/tutorial.html

1215 96. Lun ATL, Smyth GK. csaw: a Bioconductor package for differential binding analysis of ChIP-seq
1216 data using sliding windows. Nucleic Acids Res. 2016;44:e45.

1217 97. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of
1218 RNA-seq data. Genome Biol. 2010;11:R25.

1219 98. Bailey TL. STREME: accurate and versatile sequence motif discovery. Bioinformatics.
1220 2021;37:2834–40.

1221 99. Fernández R, Kallal RJ, Dimitrov D, Ballesteros JA, Arnedo MA, Giribet G, et al. Phylogenomics,
1222 Diversification Dynamics, and Comparative Transcriptomics across the Spider Tree of Life. Curr Biol.
1223 2018;28:1489-1497.e5.

1224 100. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale
1225 protein function classification. Bioinformatics. 2014;30:1236–40.

1226 101. Akiyama-Oda Y, Akaiwa T, Oda H. Supplementary Data for "Reconstruction of the Global Polarity
1227 of an Early Spider Embryo by Single-Cell and Single-Nucleus Transcriptome Analysis." figshare; 2022
1228 [cited 2022 Jun 8]; Available from:
1229 https://figshare.com/collections/Supplementary_Data_for_Reconstruction_of_the_Global_Polarity_of
1230 _an_Early_Spider_Embryo_by_Single-Cell_and_Single-Nucleus_Transcriptome_Analysis_/5963571

1231

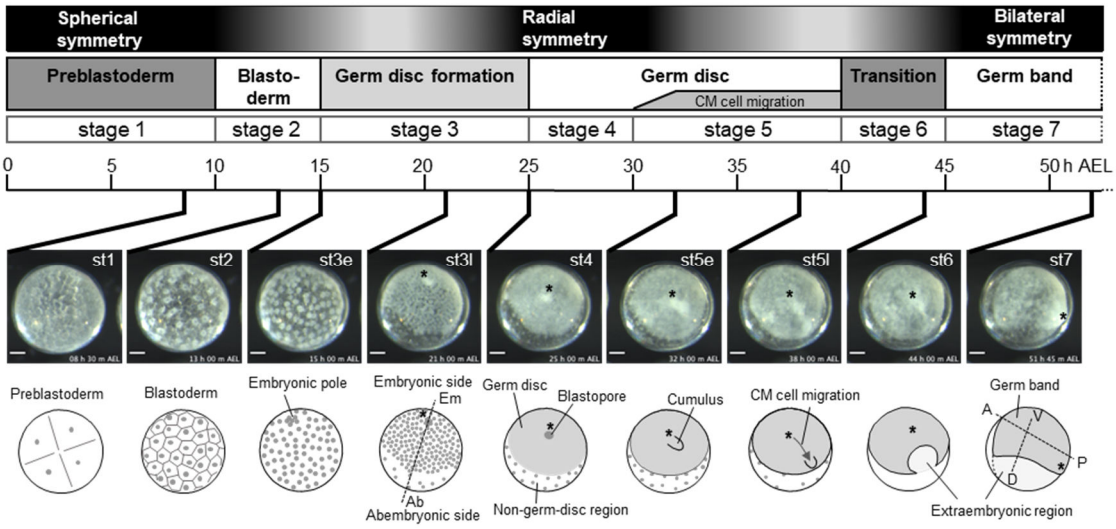



**Fig. 1. Early embryonic processes of *Parasteatoda tepidariorum*.** Selected time-lapse images of a live wild-type embryo (Additional file 1: Movie S1) are shown along with their timelines (time [h] after egg laying [AEL]), stages of development, and illustrations of morphological characteristics. The morphology of the developing *P. tepidariorum* embryo undergoes symmetry transitions as shown (top). An axis of radial symmetry (embryonic-abembryonic axis) and two axes of bilateral symmetry (anterior-posterior [A-P] and dorsal-ventral [D-V] axes) are indicated by broken lines. Asterisks mark cells around the embryonic pole, cells at the closed blastopore, and cells at the posterior pole, which are in the same lineage. Scale bars,100 µm.

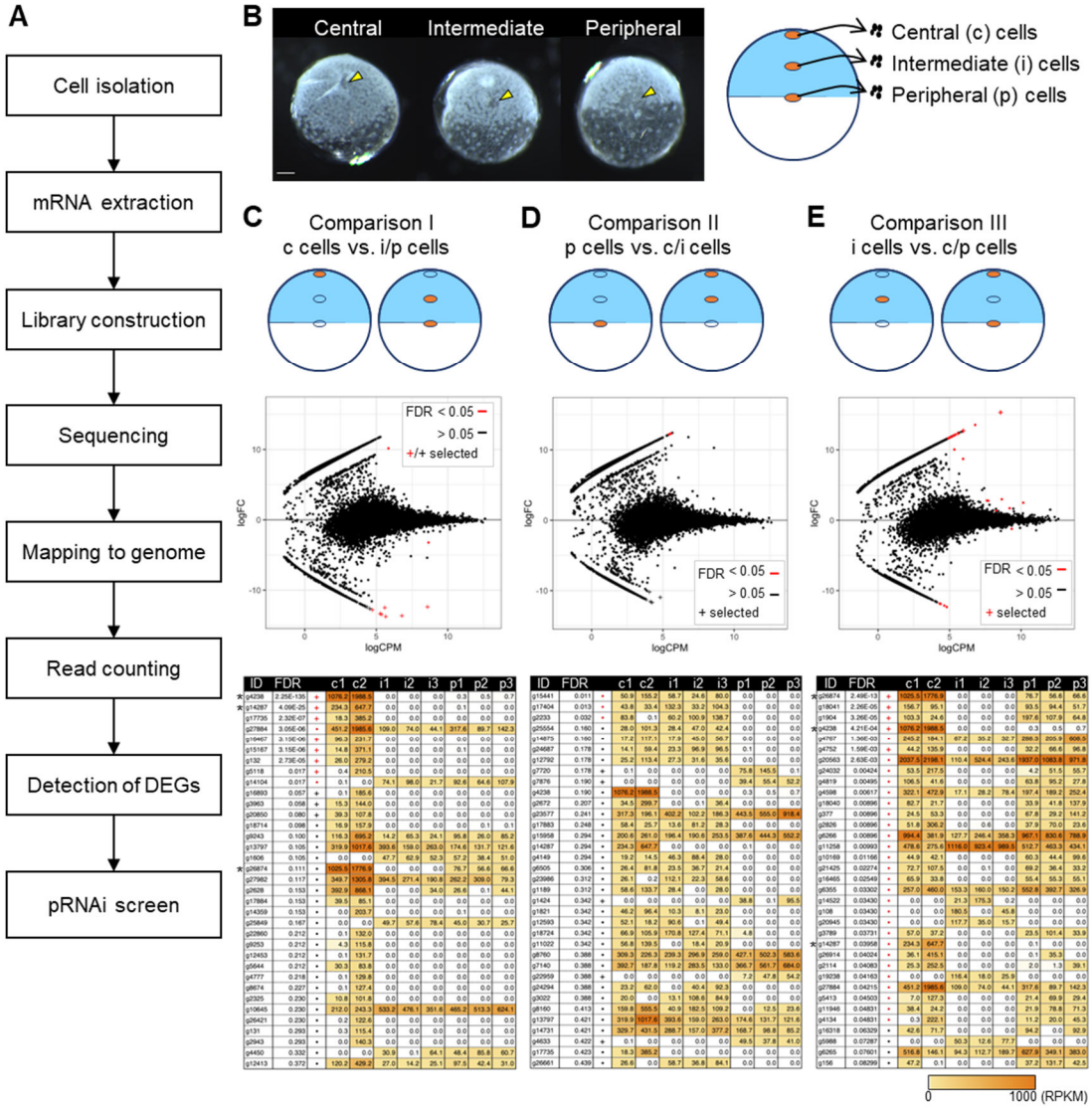
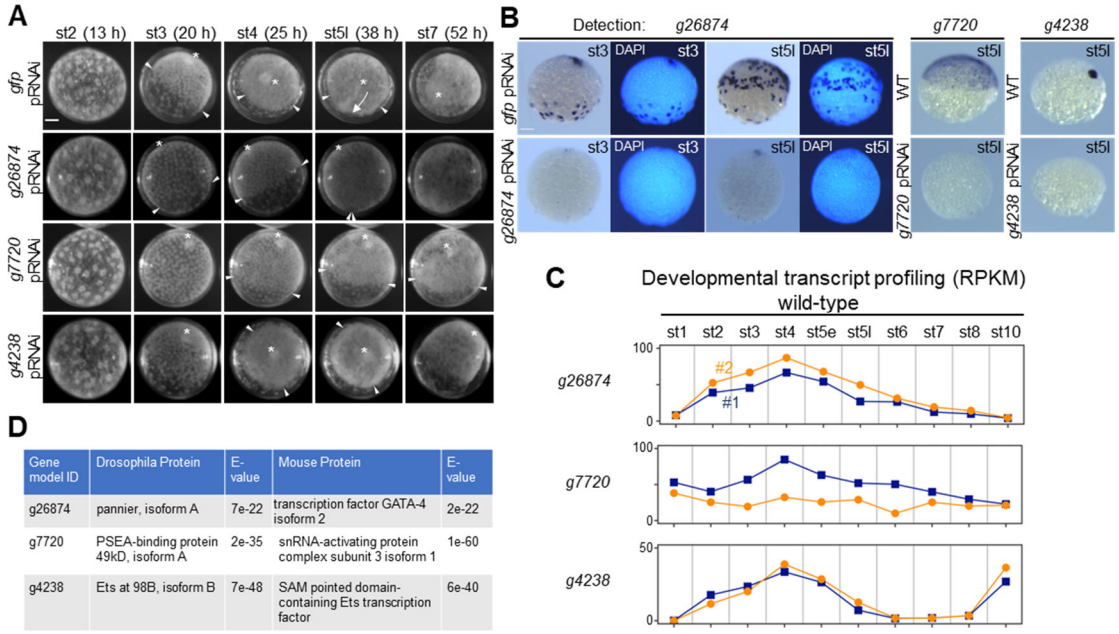



**Fig. 2. Genome-wide identification of candidate genes with locally restricted expression in stage-3 embryo. (A)** Outline of the experimental procedure. **(B)** Images showing live stage-3 embryos following isolation of cells from 3 different regions (yellow arrowheads; central [c], intermediate [i], and peripheral [p]) of the embryo, along with the schematic showing the isolated cells used for RNA-seq. Scale bar, 100 µm. **(C–E)** Three types of comparison using RNA-seq datasets from the c, i, and p cells (comparison I: c cells versus i/p cells [C]; comparison II: p cells versus c/i cells [D]; comparison III: i cells versus c/p cells [E]) detected candidates of differentially expressed genes (DEGs). Top: schematic showing the grouping of the datasets for comparison. Middle: MA-plot of $\log_2$ fold-change (FC) versus $\log_2$ average expression level (CPM; count per million) from 10,862 genes. Bottom: lists of DEG candidates

1255 with normalized expression levels (RPKM; reads per kilobase of transcript, per million
1256 mapped reads) from biological replicates of the three sample types (c1, c2, i1, i2, i3,
1257 p1, p2, and p3), which are sorted by FDR values. Genes with FDR < 0.05 are
1258 highlighted in red, and candidates of DEGs that were selected for a pilot pRNAi screen
1259 are indicated by plus signs in the MA-plots and tables. Each table displays the top 35
1260 genes; the full lists are presented in Additional file 2: Tables S1–S3. Note that genes
1261 marked by asterisks appear in multiple tables.
1262

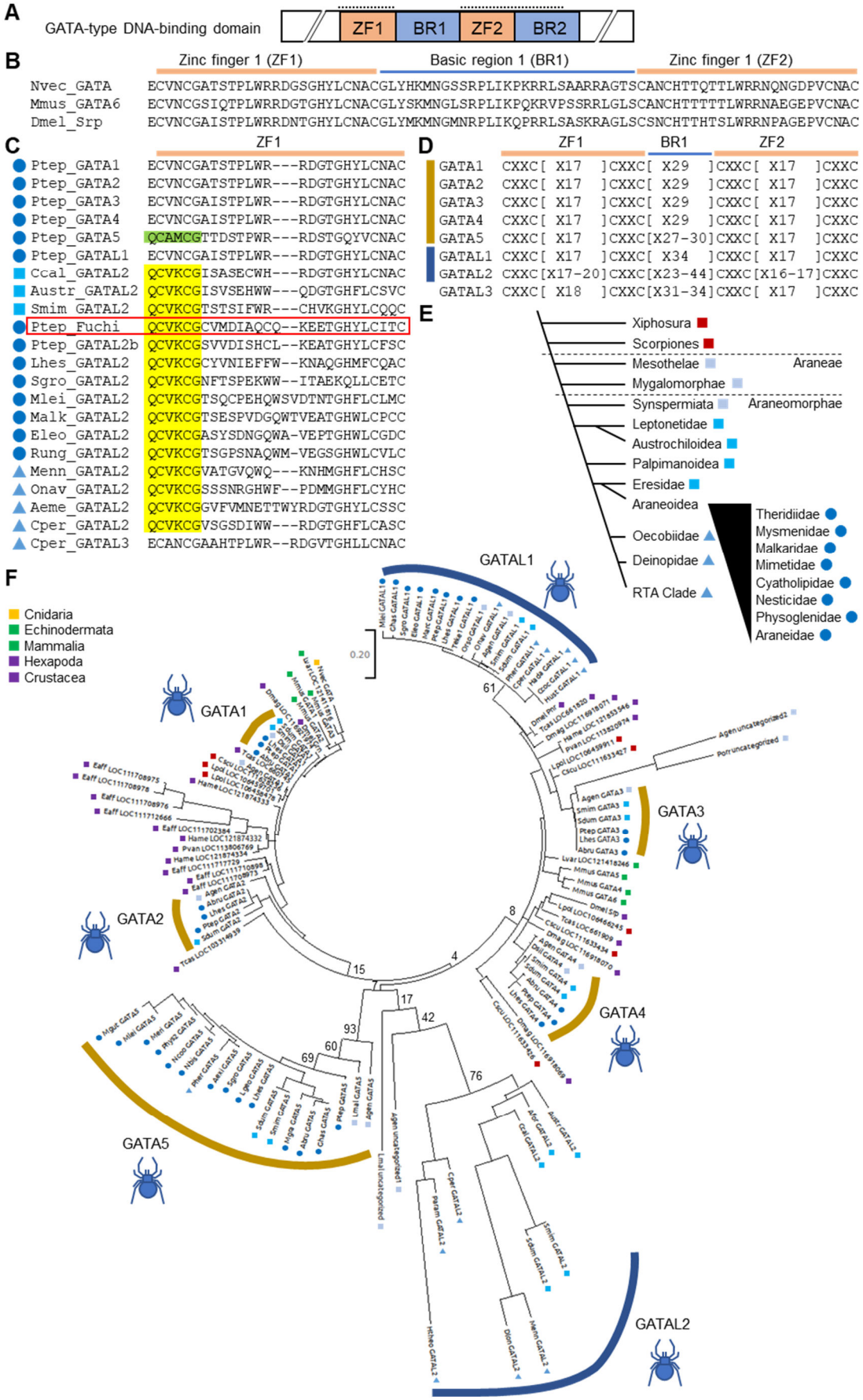
1263




D

| Gene model ID | Drosophila Protein | E-value | Mouse Protein | E-value |
|---|---|---|---|---|
| g26874 | pannier, isoform A | 7e-22 | transcription factor GATA-4 isoform 2 | 2e-22 |
| g7720 | PSEA-binding protein 49kD, isoform A | 2e-35 | snRNA-activating protein complex subunit 3 isoform 1 | 1e-60 |
| g4238 | Ets at 98B, isoform B | 7e-48 | SAM pointed domain-containing Ets transcription factor | 6e-40 |

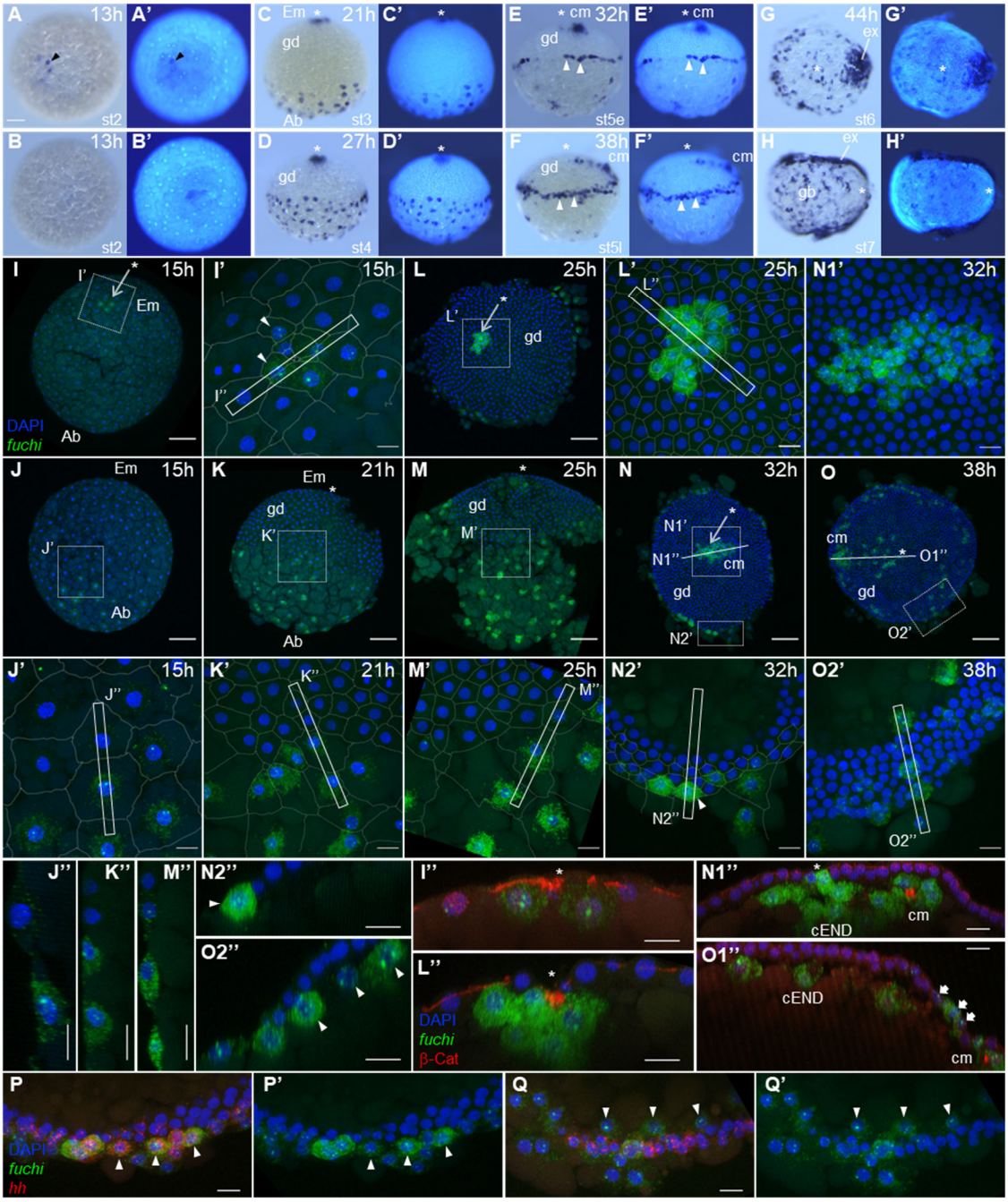
**Fig. 3. Identification of three genes in a pilot pRNAi screen of the DEG candidates.** **(A)** Images showing the development of live *gfp* pRNAi, *g26874* pRNAi, *g7720* pRNAi, and *g4238* pRNAi embryos. These images are related to Movie S2 (Additional file 4). Stages of development and time (h) after egg laying (AEL) are shown. Asterisks indicate the embryonic pole, closed blastopore, and posterior pole. Arrowheads indicate the rim of the forming/formed germ disc, or corresponding sites in morphologically affected embryos. The arrow indicates CM cell migration, which did not occur in the *g26874* pRNAi, *g7720* pRNAi, and *g4238* RNAi embryos. **(B)** Effect of *g26874*, *g7720*, or *g4238* pRNAi on the expression of the corresponding target gene at stage 3 and/or late stage 5, as revealed by WISH. *gfp* pRNAi or wild-type (WT) embryos were stained as controls. Counterstains with DAPI were displayed for the comparison between *g26874* and *gfp* pRNAi embryos. **(C)** Graphs showing developmental profiling of the transcript levels (RPKM) for *g26874*, *g7720*, and *g4238* in wild-type embryos, based on public datasets [37]. Two biological replicates are individually shown. **(D)** Results of blast searches against *D. melanogaster* and mouse RefSeq protein databases using predicted amino acid sequences of *g26874*, *g7720*, and *g4238* as queries. The top-hit protein from each search is shown. Scale bars, 100 µm in A and B.


A
GATA-type DNA-binding domain
ZF1
BR1
ZF2
BR2
B
Zinc finger 1 (ZF1)
Basic region 1 (BR1)
Zinc finger 1 (ZF2)
Nvec_GATA ECVNCGATSTPLWRRDGSGHYLCNACGLYHKMNGSSRPLIKPKRRLSAARRAGTSCANCHTTQTTLWRRNQNGDPVCNAC
Mmus_GATA6 ECVNCGSIQTPLWRRDGTGHYLCNACGLYSKMNGLSRPLIKPQKRVPSSRRLGLSCANCHTTTTTLWRRNAEGEPVCNAC
Dmel_Srp ECVNCGAISTPLWRRDNTGHYLCNACGLYMKMNGMNRPLIKQPRRLSASKRAGLSCSNCHTTHTSLWRRNPAGEPVCNAC
C
ZF1
Ptep_GATA1 ECVNCGATSTPLWR---RDGTGHYLCNAC
Ptep_GATA2 ECVNCGATSTPLWR---RDGTGHYLCNAC
Ptep_GATA3 ECVNCGAISTPLWR---RDGTGHYLCNAC
Ptep_GATA4 ECVNCGAISTPLWR---RDGTGHYLCNAC
Ptep_GATA5 QCAMCGTTDSTPWR---RDSTGQYVCNAC
Ptep_GATAL1 ECVNCGAISTPLWR---RDGTGHYLCNAC
Ccal_GATAL2 QCVKCGISASECWH---RDGTGHYLCNAC
Austr_GATAL2 QCVKCGISVSEHWW---QDGTGHFLCSVC
Smim_GATAL2 QCVKCGTSTSIFWR---CHVKGHYLCQQC
Ptep_Fuchi QCVKCGCVMDIAQCQ--KEETGHYLCITC
Ptep_GATAL2b QCVKCGSVVDISHCL--KEATGHYLCFSC
Lhes_GATAL2 QCVKCGCYVNIEFFW--KNAQGHMFCQAC
Sgro_GATAL2 QCVKCGNFTSPEKWW--ITAEKQLLCETC
Mlei_GATAL2 QCVKCGTSQCPEHQWSVDTNTGHFLCLMC
Malk_GATAL2 QCVKCGTSESPVDGQWTVEATGHWLCPCC
Eleo_GATAL2 QCVKCGASYSDNGQWA-VEPTGHWLCGDC
Rung_GATAL2 QCVKCGTSGPSNAQWM-VEGSGHWLCVLC
Menn_GATAL2 QCVKCGVATGVQWQ---KNHMGHFLCHSC
Onav_GATAL2 QCVKCGSSSNRGHWF--PDMMGHFLCYHC
Aeme_GATAL2 QCVKCGGVFVMNETTWYRDGTGHYLCSSC
Cper_GATAL2 QCVKCGVSGSDIWW---RDGTGHFLCASC
Cper_GATAL3 ECANCGAAHTPLWR--RDGVTGHLLCNAC
D
ZF1
BR1
ZF2
GATA1 CXXC[ X17 ]CXXC[ X29 ]CXXC[ X17 ]CXXC
GATA2 CXXC[ X17 ]CXXC[ X29 ]CXXC[ X17 ]CXXC
GATA3 CXXC[ X17 ]CXXC[ X29 ]CXXC[ X17 ]CXXC
GATA4 CXXC[ X17 ]CXXC[ X29 ]CXXC[ X17 ]CXXC
GATA5 CXXC[ X17 ]CXXC[X27-30]CXXC[ X17 ]CXXC
GATAL1 CXXC[ X17 ]CXXC[ X34 ]CXXC[ X17 ]CXXC
GATAL2 CXXC[X17-20]CXXC[X23-44]CXXC[X16-17]CXXC
GATAL3 CXXC[ X18 ]CXXC[X31-34]CXXC[ X17 ]CXXC
E
Xiphosura
Scorpiones
Mesothelae
Araneae
Mygalomorphae
Synspermiata
Araneomorphae
Leptonetidae
Austrochiloidea
Palpimanoidea
Eresidae
Araneoidea
Oecobiidae
Deinopidae
RTA Clade
Theridiidae
Mysmenidae
Malkaridae
Mimetidae
Cyatholipidae
Nesticidae
Physoglenidae
Araneidae
F
Cnidaria
Echinodermata
Mammalia
Hexapoda
Crustacea
GATAL1
GATA1
GATA3
GATA2
GATA4
GATA5
GATAL2
0.20
61
8
4
15
17
42
93
60
69
76

1283

1284 **Fig. 4. Phylogenetic characterization of Fuchi and other spider GATA family**
1285 **proteins. (A)** Schematic of GATA-type DNA-binding domain. This domain comprises
1286 two zinc finger motifs (ZF1 and ZF2) and two basic regions (BR1 and BR2). **(B)**
1287 Amino acid sequence alignment of the ZF1-BR1-ZF2 region (plus one adjacent residue
1288 on the N-terminal side) of three canonical GATA family members from cnidarian,
1289 mouse and *Drosophila* (Nvec_GATA, *Nematostella vectensis* GATA; Mmus_GATA6,
1290 *Mus musculus* GATA-6; Dmel_Srp, *Drosophila melanogaster* Serpent). **(C)** Amino acid
1291 sequence alignment of the ZF1 (plus one flanking residue on the N-terminal side) of
1292 various GATA family members from *P. tepidariorum* and other spider species. There
1293 are five canonical (Ptep_GATA1 to GATA5) and three noncanonical (Ptep_GATAL1,
1294 Fuchi, GATAL2b) GATA family members in *P. tepidariorum*. The GATAL2 group,
1295 including Fuchi, is characterized by the signature sequence highlighted in yellow, and
1296 the GATA5 group by the signature sequence highlighted in green. There are two types
1297 of GATAL2 proteins; the ZF1 of one type is aligned with that of the canonical GATA
1298 without gap but the ZF1 of the other type is not. **(D)** Classification of spider GATAs
1299 based on cysteine residue spacing patterns in the ZF1-BR1-ZF2 region. GATA5s in
1300 some spider species had a BR1 of slightly varied length. **(E)** Diagram showing
1301 phylogeny of spider taxa, based on a recent study [99]. **(F)** Maximum likelihood tree
1302 of canonical and non-canonical GATA family members from Araneae, non-Araneae
1303 chelicerates, non-chelicerate arthropods, non-arthropod bilaterians, and cnidarian,
1304 using 75 amino acid sites unambiguously alignable. GATAL2 proteins with imperfectly
1305 aligned ZF1, including Fuchi, were excluded from this analysis. Bootstrap values (100
1306 replicates) are presented at selected nodes. Color geometric codes indicate taxonomic
1307 groups of the sequence source. Details of species names and sequences are available in
1308 Tables S7 and S8 (Additional file 7).
1309

1310

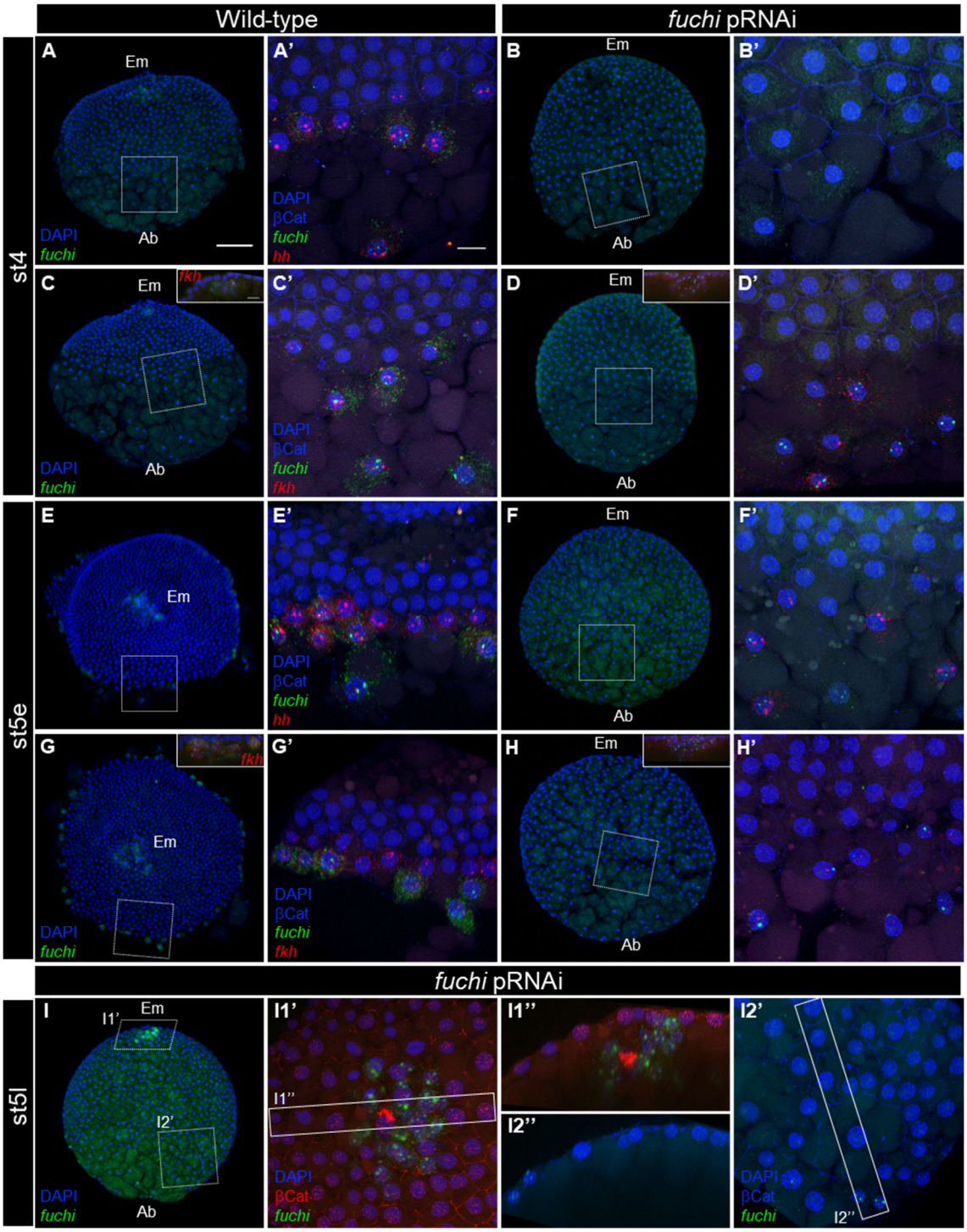



1311

1312 **Fig. 5. Expression of *fuchi* transcript in specific cells on the embryonic and**
1313 **abembryonic sides of early embryos. (A–H)** Chromogenic WISH of wild-type
1314 embryos at indicated stages using probes for *fuchi* transcript (left), counter-stained for
1315 DNA (right). (A, B) Opposite sides of the same embryo. Arrowhead indicates the signal.
1316 (C–F) Lateral view. The forming/formed germ disc (gd) is up. Arrowheads denote cells
1317 expressing *fuchi* at or near the germ-disc rim. (G, H) Forming germ band (gb) with
1318 expanding extraembryonic region (ex). Asterisks mark the embryonic pole, closed

1319 blastopore, or posterior pole. **(I–O)** Reconstruction of confocal stacks showing wild-
1320 type embryos fluorescently stained for *fuchi* transcript (green) and DNA (blue) at 15 (I,
1321 J), 21 (K), 25 (L, M), 32 (N), and 38 (O) h AEL. Asterisks mark the embryonic pole or
1322 closed blastopore. I and J show the same embryo from different directions. (I'–M', N1',
1323 N2', O2') High magnification of the boxed areas in I–O. Cell-cell junctions visualized
1324 by staining for β-catenin are traced in I'–M', and N2'. (I"–M", N2", O2") Optical
1325 sections corresponding to the rectangular regions in I'–M', N2', and O2'. Arrowheads
1326 indicate *fuchi*-expressing cells that appear to be internalizing or internalized. (N1",
1327 O1") Optical sections corresponding to the lines in N and O. Signals for β-catenin are
1328 overlaid in red in I", L", N1", and O1". β-Catenin concentrations mark the cumulus
1329 (cm). Fat arrows indicate *fuchi* expression in the future extraembryonic region. (P, Q)
1330 Simultaneous detection of *fuchi* (green) and *Pt-hh* (red) transcript and DNA (blue) in
1331 cells at or near the germ-disc rim at early (P, P') and late (Q, Q') stage 5. Arrowheads
1332 indicate cells expressing both *fuchi* and *Pt-hh* before internalization and cells
1333 expressing only *fuchi* after internalization. Em, embryonic side; Ab, abembryonic side.
1334 Scale bars, 100 μm in A, I–O; 20 μm in other panels.
1335

1336

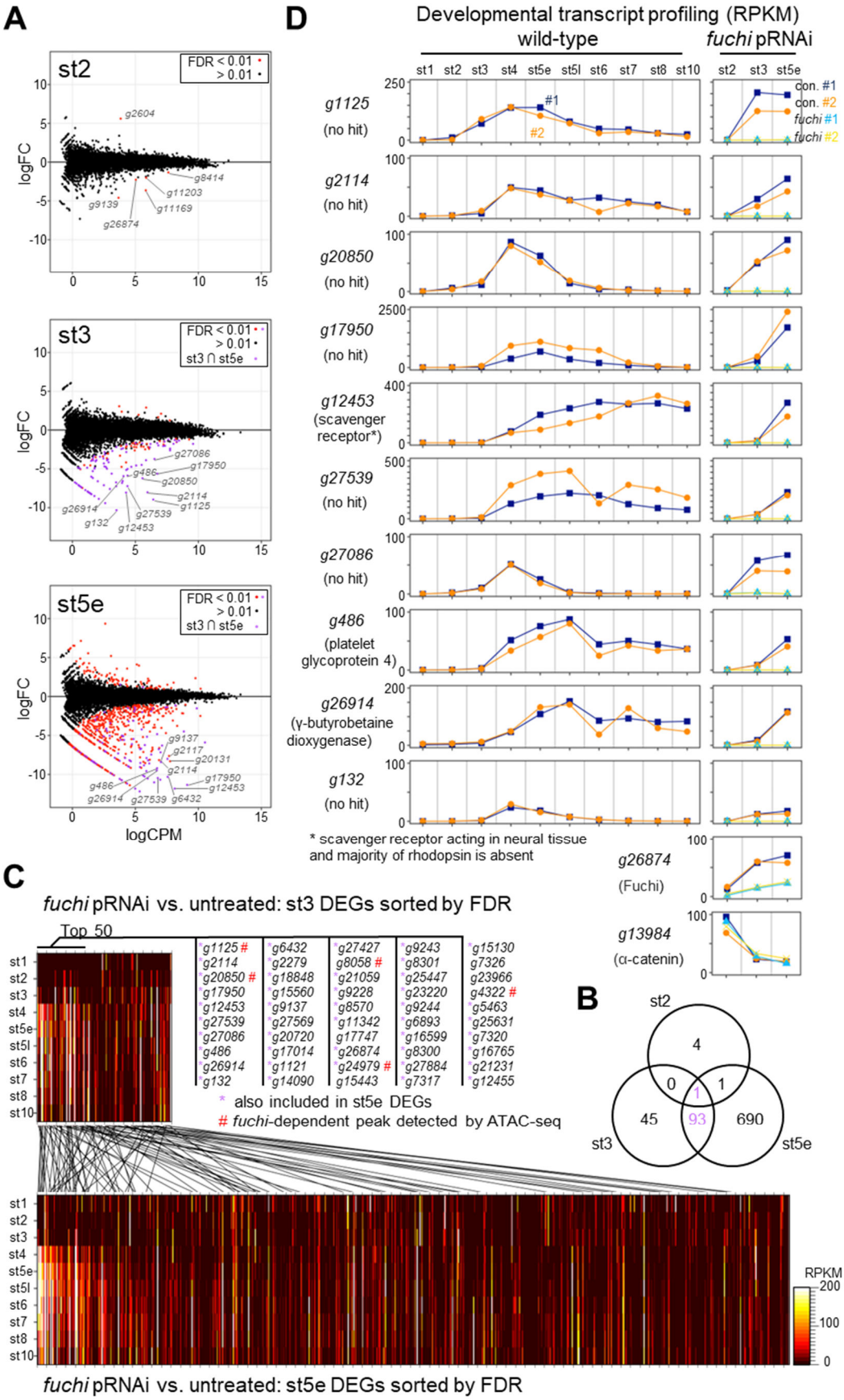



1337

1338 **Fig. 6. Defects of *fuchi* pRNAi embryos in demarcating a forming germ disc. (A–**
1339 **H)** Reconstructions of confocal stacks showing wild-type (A, C, E, G) and *fuchi* pRNAi
1340 (B, D, F, H) embryos stained for *fuchi* (green, A–H) and *Pt-hh* (red, A, B, E, F) or *Pt-*
1341 *fkh* (red, C, D, G, H) transcript, DNA (blue, A–H), and β-catenin (blue, A'–H'; not
1342 shown in A-H) at stage 4 (A–D) and early stage 5 (E–H). In A–D, F, H, and I, the

1343 embryonic side (Em) is to top, with the abembryonic (Ab) to bottom. In E and G, the
1344 embryonic side is viewed from the top. Optical sections show the presence of specific
1345 cells expressing *fuchi* at the embryonic pole (insets, C, D, G, H). The boxed areas in
1346 A-H are magnified in A'-H', respectively. *fuchi* transcript was detectable even in cells
1347 on the abembryonic side of *fuchi* pRNAi embryos, although the signals were restricted
1348 to within the nuclei. Cells expressing *Pt-hh* and *Pt-fkh* transcripts at reduced levels
1349 were observed at regions corresponding to the the germ-disc rim and nearby
1350 abembryonic region in *fuchi* pRNAi embryos at early stage 5, but these cells did not
1351 serve as rim cells demarcating a forming germ disc. (I) Reconstructions of confocal
1352 stacks showing a *fuchi* pRNAi embryo stained for *fuchi* transcript (green, I, I1', I2'),
1353 DNA (blue, I, I1', I2'), and β-catenin (red, I1'; blue, I2') at late stage 5. (I1', I2') High
1354 magnification of the boxed areas in I as indicated. (I1", I2") Cross-sections of the
1355 rectangular regions in I1' and I2'. The cluster of CM cells were eventually internalized,
1356 but with few cells internalized in other regions. Note that the cells spread over the
1357 abembryonic side with epithelial integrity retained to a certain extent. Scale bars, 100
1358 μm in A; 20 μm in A', inset of C.
1359

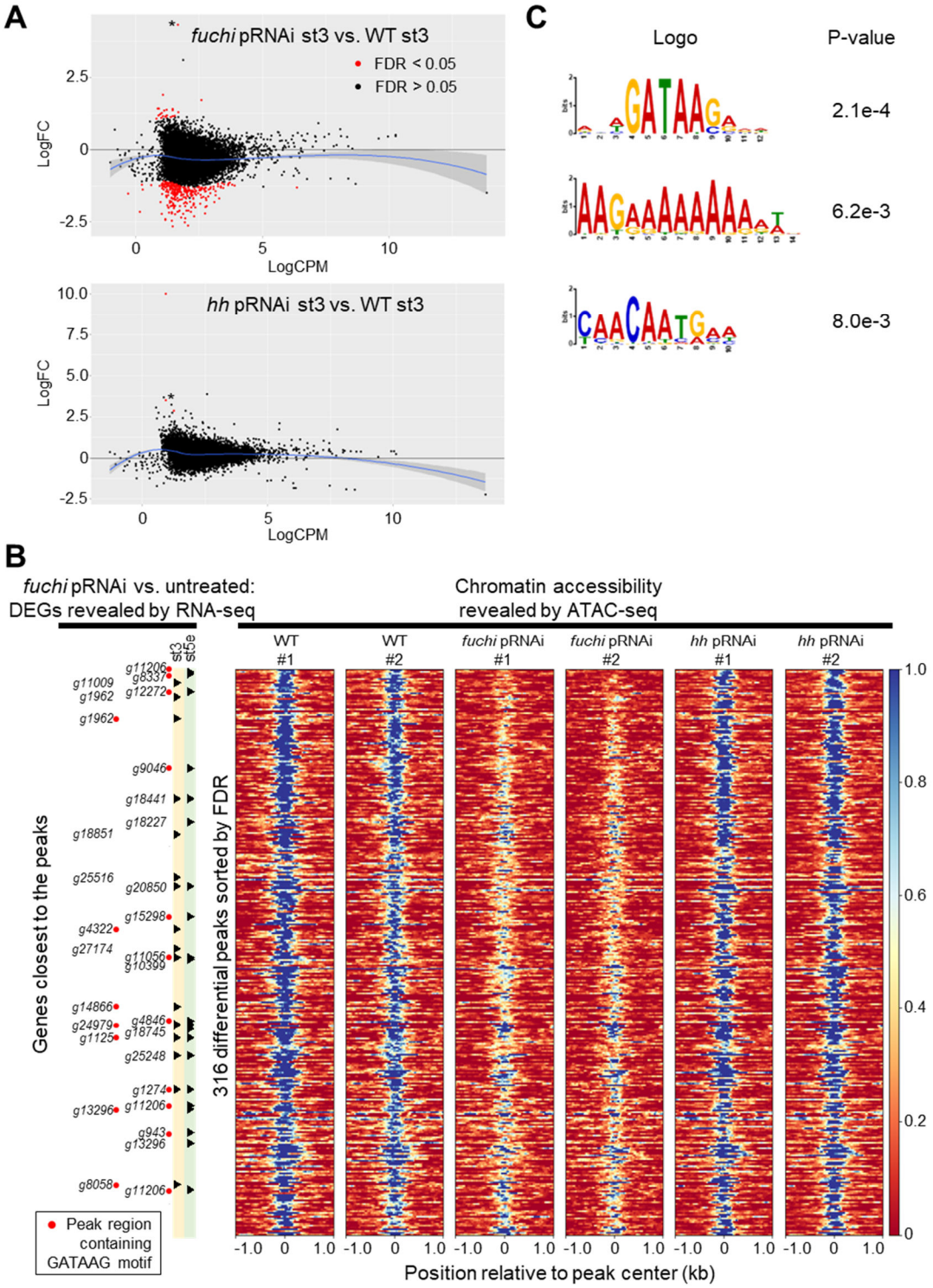
1360


A
st2
FDR < 0.01
> 0.01
logFC
g2604
g8414
g11203
g9139
g11169
g26874
st3
st3 ∩ st5e
g27086
g17950
g486
g20850
g2114
g1125
g26914
g132
g27539
g12453
st5e
g9137
g2117
g20131
g17950
g12453
g6432
logCPM
D
Developmental transcript profiling (RPKM)
wild-type
fuchi pRNAi
st1 st2 st3 st4 st5e st5l st6 st7 st8 st10
st2 st3 st5e
con. #1
con. #2
fuchi #1
fuchi #2
g1125
(no hit)
g2114
(no hit)
g20850
(no hit)
g17950
(no hit)
g12453
(scavenger receptor*)
g27539
(no hit)
g27086
(no hit)
g486
(platelet glycoprotein 4)
g26914
(γ-butyrobetaine dioxygenase)
g132
(no hit)
* scavenger receptor acting in neural tissue and majority of rhodopsin is absent
g26874
(Fuchi)
g13984
(α-catenin)
C
fuchi pRNAi vs. untreated: st3 DEGs sorted by FDR
Top 50
g1125 #
g2114
g20850 #
g17950
g12453
g27539
g27086
g486
g26914
g132
g6432
g2279
g18848
g15560
g9137
g27569
g20720
g17014
g1121
g14090
g27427
g8058 #
g21059
g9228
g8570
g11342
g17747
g26874
g24979 #
g15443
g9243
g8301
g25447
g23220
g9244
g6893
g16599
g8300
g27884
g7317
g15130
g7326
g23966
g4322 #
g5463
g25631
g7320
g16765
g21231
g12455
* also included in st5e DEGs
# fuchi-dependent peak detected by ATAC-seq
B
st2
st3
st5e
4
0
1
1
45
93
690
RPKM
200
100
0
fuchi pRNAi vs. untreated: st5e DEGs sorted by FDR

1361

1362 **Fig. 7. Genome-wide identification of genes whose expression levels are affected**
1363 **following *fuchi* knockdown. (A)** MA-plots of $\log_2$ fold-change (FC) versus $\log_2$
1364 average expression level (CPM) showing DEGs between *fuchi* pRNAi and control
1365 (untreated) embryos at stage 2 (st2), stage 3 (st3), and early stage 5 (st5e). Genes with
1366 FDR < 0.01 are highlighted in red or purple, with the purple genes included in both
1367 stage-3 and early stage-5 DEG lists. **(B)** Venn diagram showing the numbers of the
1368 DEGs from the three sample types and their overlaps. Note that *fuchi* (*g26874*) itself is
1369 included in all the DEG lists. **(C)** Graphs showing developmental profiling of the
1370 transcript levels for the top 10 DEGs from the stage-3 samples in wild-type embryos
1371 and the effects of *fuchi* pRNAi on the transcript levels. Top hit proteins that resulted
1372 from blastx searches against mouse and *Drosophila* RefSeq protein databases are
1373 shown, although with no hits obtained with seven of the ten genes (E-value < 1e-5).
1374 Data on *fuchi* (*g26874*) and α-catenin (*g13984*) validated the experiments. **(D)** Heat
1375 maps showing developmental profiling of the transcript levels for all the DEGs from
1376 the stage-3 (upper) and early stage-5 (lower) samples in wild-type embryos (sorted by
1377 FDR values). The top 50 DEGs from stage-3 are listed, where the genes included in
1378 stage-5 DEGs are indicated by purple asterisks. Genomic regions differentially
1379 accessible between wild-type and *fuchi* pRNAi embryos at stage 3 were found to be
1380 located within or close to some of the DEGs marked by "#"(related to Fig. 8 and
1381 Additional file 10: Table S13).
1382

1383

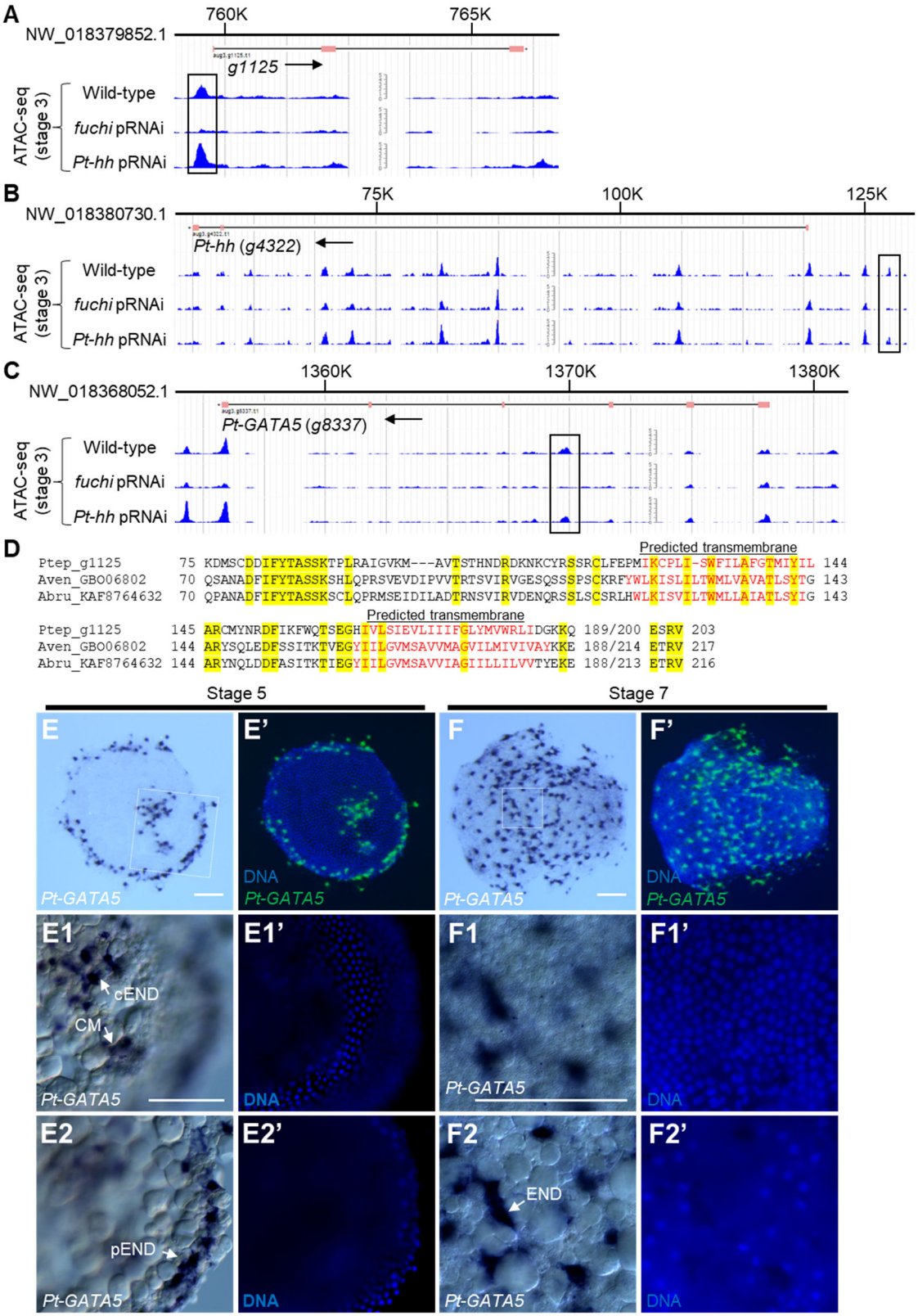



1384

1385 **Fig. 8. Effect of *fuchi* knockdown on chromatin accessibility in specific genomic**
1386 **regions in stage-3 embryos. (A)** Distributions of differentially accessible genomic
1387 regions identified by comparisons between ATAC-seq datasets from *fuchi* (upper) or
1388 *Pt-hh* (lower) pRNAi and wild-type embryos. MA-plots represent $\log_2$ fold-change

1389 (FC) versus $\log_2$ average ATAC signal abundance (CPM). Two biological replicates
1390 were obtained from each sample type and analyzed. Significant differentially accessible
1391 regions (FDR < 0.05) are highlighted in red. Blue lines are loess fits to each distribution
1392 with 95% confidence intervals shaded in gray. Asterisks indicate positive peaks at the
1393 same genomic region. Detection of this region might not be specific to *fuchi* activity.
1394 **(B)** Heat maps showing chromatin accessibilities in the 316 genomic regions identified
1395 as *fuchi*-dependent (sorted by FDR values). Data from the biological replicates are
1396 individually shown. Some of the identified regions are located close to or within the
1397 loci of DEGs identified by the RNA-seq analyses, as shown on the left side. **(C)** Motif
1398 sequences enriched in the *fuchi*-dependent peaks, as revealed by a motif discovery tool
1399 STREME [98]. Peak regions containing detected motif sequences,
1400 RNWGATAAGAVW, are indicated in B (red dots).
1401

1402




1403

1404

1405

1406 **Fig. 9. Three notable examples of genes revealed by combined analyses of the**
1407 **RNA-seq and ATAC-seq datasets. (A–C)** Normalized ATAC-seq signal of stage-3
1408 wild-type, *fuchi* pRNAi, and *Pt-hh* pRNAi embryo at the *g1125* (A), *Pt-hh* (*g4322*) (B),
1409 and *Pt-GATA5* (*g8337*) (C) loci. Line with ticks at the top represents the scaffold
1410 sequence, below which annotated exons (pink boxes) and introns (black lines) are
1411 shown. *fuchi*-dependent peak regions are indicated by open rectangles. Arrows indicate
1412 the orientation of the gene. (D) Amino acid sequence alignment of the *g1125* protein
1413 and its homologs detected in the genomes of two other spider species *Araneus*
1414 *ventricosus* (GBO06802.1) and *Argiope bruennichi* (KAF8764632.1). Residues
1415 identical across all proteins are presented with a yellow background. Residues shown
1416 in red indicate regions predicted as transmembrane domains by InterProScan version
1417 5.53-87.0 [100]. (E, F) Detection of *Pt-GATA5* transcript and DNA (E', F') in stage-5
1418 (E) and stage-7 (F) embryos using chromogenic WISH. (E1, E2, F1, F2) High
1419 magnification of the boxed areas in E and F, along with counter staining for DNA (E1',
1420 E2', F1', F2'). The focal plane is at two different depths; the plane shown in E2 and F2
1421 is deeper than that in E1 and F1, respectively. Signals are observed in cEND, CM, and
1422 pEND cells, which are located below the germ-disc epithelial layer, and in evenly
1423 distributed endoderm (END) cells, which are located below the germ-band epithelial
1424 layer. Scale bars, 100 μm.
1425